

**CARACTERIZACIÓN DE LA DESHONESTIDAD
ACADÉMICA EN CURSOS EN LÍNEA PRIVADOS Y
MASIVOS (MPOC) CON RECONOCIMIENTO
ACADÉMICO A TRAVÉS DE ANALÍTICAS DE
APRENDIZAJE**



DANIEL ALBERTO JARAMILLO MORILLO

Tesis de Doctorado en Ingeniería Telemática

Director:

Mario Fernando Solarte Sarasty
Doctor en Ingeniería Telemática

Co-Director:

José Antonio Ruipérez Valiente
Doctor en Ingeniería Telemática

Universidad del Cauca

Facultad de Ingeniería Electrónica y Telecomunicaciones

Departamento de Telemática

Línea de Investigación Aplicaciones y Servicios sobre Internet

Popayán, 2022

DANIEL ALBERTO JARAMILLO MORILLO

**CARACTERIZACIÓN DE LA DESHONESTIDAD
ACADÉMICA EN CURSOS EN LÍNEA PRIVADOS Y
MASIVOS (MPOC) CON RECONOCIMIENTO
ACADÉMICO A TRAVÉS DE ANALÍTICAS DE
APRENDIZAJE**

**Tesis presentada a la Facultad de Ingeniería Electrónica y
Telecomunicaciones de la Universidad del Cauca para la obtención
del Título de**

**Doctor en:
Ingeniería Telemática**

**Director:
PhD. Mario Fernando Solarte Sarasty
Co-Director:
PhD. José Antonio Ruipérez Valiente**

**Popayán
2022**

Dedicatoria

Dedicado a mi amada esposa, Diana, por el apoyo, ánimo y motivación que me brinda día a día para alcanzar nuevas metas, tanto profesionales como personales.

A mis hijos, Samantha y Santiago, quienes son el mejor regalo que haya recibido en la vida. Son mi mayor tesoro y la fuente más grande de motivación.

A mis sobrinos Leonardo, Alejandro, Isabella y Paula, que este logro lo sientan como propio y les sirva de estímulo en su vida, para continuar con fuerza progresando y luchando por cumplir sus metas.

Agradecimientos

En primer lugar le agradezco a Dios, a quien le debo todo.

A mi esposa, Diana Erazo, porque durante todo este tiempo fue mi apoyo y compañía incondicional. Por estar ahí en los momentos difíciles.

A mis padres, Bayardo y Celina, que siempre me han brindado su apoyo para poder cumplir todos mis objetivos personales y académicos.

A mis hermanos Maritza, Monica, Giovanni y demás familiares, por su amor, confianza y compañía.

A Alba Taimal, por todo el cariño y cuidado que me ha brindado.

A Mario Solarte y Gustavo Ramírez, por todo el apoyo, orientación y guía durante todo este proceso.

A José Ruipérez por todo lo que aprendí a su lado, por su tutoría y contribuciones a la presente investigación.

A Sonia Guerrero, Claudia Burbano, Fabinton Sotelo y Luis Alejandro Cruz por su colaboración durante el tiempo de estudio.

A los profesores y estudiantes de la Maestría y del Doctorado en Ingeniería Telemática de la Universidad del Cauca por sus enseñanzas.

A mis amigos y a todos aquellos a quienes no alcanzo a mencionar. ¡Gracias y mil bendiciones!

Abstract

Thanks to the evolution of communication technologies and their increasing use, in a very short time many institutions and universities around the world have joined the online education leap. This is where new learning technologies such as Massive Open Online Courses (MOOCs) are gradually being incorporated into universities and their professional programmes through their private variants, MPOCs (Massive Private Online Courses) and SPOCs (Small Private Online Courses), which have been successfully applied in various educational environments.

Due to this success and the advantages of having a more controlled private environment, it has become easier for MPOCs and SPOCs to be validated for academic credit within the educational programmes of higher education institutions and since these courses have academic recognition, it is necessary to monitor possible academic dishonesty behaviours, as numerous researchers have highlighted that this issue is greatly facilitated by the online context.

Although numerous algorithms have been proposed for detecting dishonest behaviour, there has been no research on how to transfer this information to instructors, so that they can intervene and decrease the prevalence of this problem. In this paper we present algorithms and tools for the identification and characterisation of suspicious behaviours of academic dishonesty of students in SPOC offered in Selene, an Open edX instance belonging to the Universidad del Cauca.

Keywords: Learning analytics; educational data mining; information visualization; MOOCs; SPOCs; behavioral modeling; machine learning.

Resumen

Gracias a la evolución de las tecnologías de la comunicación y el creciente uso de las mismas, en muy poco tiempo muchas instituciones y universidades del mundo se han sumado al salto de la educación en línea. Es aquí, donde nuevas tecnologías de aprendizaje como los Cursos en Línea Masivos y Abiertos (MOOCs) se están incorporando paulatinamente a las universidades y a sus programas profesionales a través de sus variantes privadas, los MPOCs (Massive Private Online Courses) y los SPOCs (Small Private Online Courses), que se han aplicado con éxito en diversos entornos educativos.

Gracias a este éxito y a las ventajas de tener un entorno privado más controlado, se ha facilitado que los MPOCs y SPOCs sean convalidados por créditos académicos dentro de los programas educativos de las instituciones de educación superior. Sin embargo, dado que estos cursos tienen un reconocimiento académico, es necesario vigilar los posibles comportamientos de deshonestidad académica, ya que numerosos investigadores han sacado a relucir que estos comportamientos se ven muy facilitados por el contexto en línea.

Aunque se han propuesto numerosos algoritmos para la detección de comportamientos deshonestos, no se ha investigado cómo transferir esta información a los instructores, para que puedan intervenir y disminuir la prevalencia de este problema. En este documento presentamos algoritmos y herramientas para la identificación y caracterización de comportamientos sospechosos de deshonestidad académica de estudiantes en SPOC ofrecidos en Selene, una instancia de Open edX que pertenece a la Universidad del Cauca.

Palabras clave: Analítica de aprendizaje; minería de datos educacionales; visualización de información; MOOCs; SPOCs; modelado de comportamiento; aprendizaje máquina.

Tabla de Contenido

Abstract	7
Resumen	9
Índice	11
Lista de tablas	17
Lista de figuras	17
1. Introducción	1
1.1. Planteamiento del Problema	1
1.2. Objetivos	6
1.3. Actividades desarrolladas	6
1.3.1. WP1. Generación de la base inicial de conocimiento	6
1.3.2. WP2. Descripción general de indicadores para la deshonestidad académica en Cursos en Línea Privados y Masivos	7
1.3.3. WP3. Construcción de un prototipo para el análisis del comportamiento de la deshonestidad académica con base en los indicadores seleccionados	7
1.3.4. WP4. Validación del prototipo mediante su aplicación en MPOC/SPOCs con reconocimiento académico ofrecidos en la Universidad del Cauca	7
1.3.5. WP5. Publicación	8
1.4. Resultados Alcanzados	8
1.5. Contenido de la Monografía	9

I	Trabajos relacionados, herramientas y métodos	11
2.	Trabajos relacionados	13
2.1.	Certificación y reconocimiento en MOOC/MPOC/SPOC	14
2.2.	Deshonestidad académica en MOOCs	16
2.2.1.	Exámenes sin control de identidad o sin supervisión	16
2.2.2.	Exámenes supervisados personalmente	17
2.2.3.	Exámenes supervisados automáticamente	18
2.3.	Colaboraciones en MOOCs	22
2.3.1.	Análisis del comportamiento de los estudiantes en los MOOCs	22
2.3.2.	Colaboración en los MOOCs	24
2.4.	Aplicaciones y herramientas de visualización	26
3.	Herramientas y Métodos	29
3.1.	Plataforma MOOC	29
3.1.1.	Plataforma Selene	29
3.1.2.	Características de Open edX	31
3.2.	Casos de estudio	35
3.2.1.	Curso Virtual de Astronomía Cotidiana	36
3.2.2.	Introducción al emprendimiento con Lean Startup	37
3.2.3.	Comprensión de Textos Argumentativos	38
3.2.4.	La danza folclórica como patrimonio cultural	38
3.2.5.	Introducción a la edición de texto científicos y literarios con LaTeX	39
3.3.	Indicadores seleccionados	40
3.3.1.	Interacciones con la plataforma	40
3.3.2.	Extracción de datos para medición de indicadores de interacción	43
3.3.3.	Identificación de estudiantes sospechosos de deshonestidad académica	46
3.3.4.	K-means algorithm	48
II	Identificación y caracterización de comportamientos deshonestos	51
4.	Identificación de estudiantes sospechosos de deshonestidad académica	53
4.1.	Descripción del contexto y del caso de estudio	54
4.2.	Estudiantes identificados como sospechosos de deshonestidad académica	54

4.3.	Análisis del comportamiento y el compromiso de los estudiantes	57
5.	Caracterización de colaboraciones en MPOC/SPOC	63
5.1.	Contexto del estudio	64
5.1.1.	SPOCs utilizados para el estudio	64
5.1.2.	Variables utilizadas	65
5.1.3.	Métricas y método de agrupamiento	66
5.1.4.	Definición de colaboración en este estudio	67
5.2.	Aplicación del método K-means	68
5.2.1.	Caracterización de los estudiantes.	68
5.3.	Detección de colaboraciones	72
5.4.	Caracterización de las Colaboraciones	77
5.4.1.	Asociaciones de parejas	77
5.4.2.	Asociaciones de comunidades	78
III	Propuestas desarrolladas	81
6.	Herramienta de seguimiento a las actividades de aprendizaje	83
6.1.	Descripción de la Herramienta de seguimiento	85
6.1.1.	Indicadores de seguimiento	85
6.1.2.	Diseño de la herramienta	86
6.1.3.	Implementación de la herramienta	87
6.2.	Desarrollo de caso de estudio	89
6.2.1.	Resultados de seguimiento	89
6.3.	Evaluación de usabilidad y utilidad	96
6.3.1.	Resultados en cuanto a Usabilidad	97
6.3.2.	Resultados en cuanto a Utilidad	98
6.3.3.	Resultados de Opinión	99
7.	Dashboard: Herramienta para la identificación y seguimiento	101
7.1.	Analítica de aprendizaje para detectar comportamientos deshonestos . . .	102
7.1.1.	Arquitectura general	102
7.1.2.	Características del Dashboard	104
7.2.	Estudio de caso de evaluación	106
7.2.1.	Cursos	106
7.2.2.	Instructores	107

7.2.3. Diseño de la evaluación	108
7.3. Análisis cualitativo	109
7.4. Resultados de las entrevistas	110
7.4.1. RQ1. ¿Qué percepción tienen los instructores sobre las métricas utilizadas para detectar comportamientos deshonestos son adecuadas?	112
7.4.2. RQ2. ¿Cuál es el grado de fiabilidad que los instructores suponen del algoritmo utilizado?	113
7.4.3. RQ3. ¿Qué estrategias utilizaron o propusieron los instructores para disuadir la deshonestidad académica?	114
7.4.4. RQ4. ¿Creen los instructores que es posible reducir la deshonestidad académica mediante el uso del tablero presentado?	116
IV Conclusiones, limitaciones y trabajos futuros	117
8. Discusión de resultados	119
8.1. Algoritmo para la detección de comportamientos con sospecha de deshonestidad académica	119
8.2. Caracterización de las colaboraciones encontradas	122
8.3. Herramienta de seguimiento	124
8.4. Evaluación de la herramienta de detección de conductas con sospecha de deshonestidad académica	125
9. Conclusiones, Limitaciones y Trabajo futuro	129
9.1. Conclusiones	129
9.1.1. Observaciones finales	129
9.1.2. Limitaciones	131
9.2. Estancias de investigación	134
9.3. Trabajos futuros	135
Apéndice A. Prueba usabilidad, utilidad y opinión	137
Apéndice B. Diseño del mecanismo para el seguimiento de las actividades de aprendizaje de los estudiantes en MOOC	139
B.1. Vista de escenarios	140
B.2. Vista lógica	141
B.3. Vista de procesos	142

B.4. Vista de implementación	143
B.5. Vista de despliegue	145
Bibliografía	147

Lista de Tablas

3.1. Cantidad de usuarios registrados en la plataforma Selene	30
4.1. Similitud en respuestas y cercanía en el tiempo de envío.	57
5.1. Estadísticas de registro y finalización de los cursos	65
5.2. Clusterización de los estudiantes de los cursos	73
5.3. Asociaciones y comunidades	79
6.1. Indicadores de seguimiento propuestos	86
6.2. Resultados sobre la usabilidad de la herramienta	98
6.3. Resultados sobre la utilidad de la herramienta	99
6.4. Opinión sobre la herramienta	99
7.1. Variables, categories, and examples	111

Lista de Figuras

2.1.	Diagrama de relaciones de los tipos de reconocimiento y control de identidad en MOOC.	15
2.2.	Esquema de división de trabajos relacionados a la supervisión automática de las evaluaciones en línea.	19
3.1.	Vista del contenido del curso en Open edX.	32
3.2.	Vista de los contenidos en Open edX.	32
3.3.	Vista de pestaña “Instructor” en Open edX.	33
3.4.	Vista de la estructura de los cursos en Studio	34
3.5.	Vista de la estructura de los cursos en Studio	34
3.6.	Arquitectura general mecanismo propuesto	44
3.7.	Script de extracción (Captura de pantalla)	45
3.8.	Script de procesamiento de registros (Captura de pantalla)	45
3.9.	Evento del archivo tracking.log	46
4.1.	Dendrograms for each one of the metrics clustering.	56
4.2.	Boxplot distribution of grades by cohort.	58
4.3.	Interaction of students with the contents of the course.	59
4.4.	Average number of interactions over time separated by cohort.	61
5.1.	Variables de interacción en los 8 cursos para las cortes 2020 y 2021	66
5.2.	Clústers obtenidos en cada uno de los cursos y cohortes	69
5.3.	Visualización de Boxplot de las variables para los cursos de <i>Lean Startup</i> y <i>Textos Argumentativos</i>	70
5.4.	Visualización de Boxplot de las variables para los cursos de <i>Danza Folclórica</i> y <i>LaTeX</i>	71
5.5.	Dendrograma para el curso de <i>Lean Startup - 2020</i>	73
5.6.	Dendrograma para el curso de <i>Lean Startup - 2021</i>	74

5.7.	Dendrograma para el curso de <i>Textos Argumentativos - 2020</i>	74
5.8.	Dendrograma para el curso de <i>Textos Argumentativos - 2021</i>	75
5.9.	Dendrograma para el curso de <i>Danza Folclórica - 2020</i>	75
5.10.	Dendrograma para el curso de <i>Danza Folclórica - 2021</i>	76
5.11.	Dendrograma para el curso de <i>LaTeX - 2020</i>	76
5.12.	Dendrograma para el curso de <i>LaTeX - 2021</i>	77
6.1.	Diagrama de despliegue - Herramienta de seguimiento	87
6.2.	Captura de pantalla de la herramienta de seguimiento. Historial de ingresos en el curso de <i>Astronomía cotidiana 2016</i>	88
6.3.	Consulta de la constancia en el curso de <i>Astronomía cotidiana 2016</i>	89
6.4.	Menú de navegación herramienta de seguimiento	90
6.5.	Ingresos al curso piloto mes de febrero	90
6.6.	Historial de ingresos al curso en el mes de febrero	91
6.7.	Consulta interacciones de los estudiantes con los vídeos del curso	92
6.8.	Consulta de las interacciones de un estudiante específico con vídeos	93
6.9.	Consulta de las interacciones en el foro del curso	93
6.10.	Consulta de las interacciones en el examen final del curso	94
6.11.	Menú navegación Indicadores compuestos	94
6.12.	Consulta del rendimiento de los estudiantes del curso	95
6.13.	Consulta del progreso de los estudiantes del curso	96
6.14.	Consulta de la constancia de los estudiantes del curso	97
7.1.	Dashboard architecture for the detection of academically dishonesty behaviours.	103
7.2.	Screenshot of the view of students suspected of academic dishonesty in the tool	105
7.3.	Screenshot of the tool's statistics view.	105
7.4.	Screenshot of the tool's behaviours view.	106
7.5.	Fase de análisis cualitativo	109
B.1.	Arquitectura general de la herramienta de seguimiento	139
B.2.	Diagrama de casos de uso	141
B.3.	Diagrama de clases	142
B.4.	Diagrama de actividades (Monitorización)	143
B.5.	Diagrama de actividades (Visualización)	144
B.6.	Diagrama de componentes Módulo de monitorización	144

B.7. Diagrama de componentes Módulo de visualización	145
B.8. Diagrama de despliegue	145

Capítulo 1

Introducción

Índice

2.1. Certificación y reconocimiento en MOOC/MPOC/SPOC	14
2.2. Dishonestidad académica en MOOCs	16
2.2.1. Exámenes sin control de identidad o sin supervisión	16
2.2.2. Exámenes supervisados personalmente	17
2.2.3. Exámenes supervisados automáticamente	18
2.3. Colaboraciones en MOOCs	22
2.3.1. Análisis del comportamiento de los estudiantes en los MOOCs	22
2.3.2. Colaboración en los MOOCs	24
2.4. Aplicaciones y herramientas de visualización	26

Este capítulo de introducción establece el marco de investigación de esta tesis. En primer lugar, en la sección 1.1 presentamos las motivaciones iniciales para llevar a cabo nuestra investigación. En la sección 1.2 se exponen los objetivos que pretendemos alcanzar. En la sección 1.3 se presentan las actividades realizadas. En la sección 1.4 se presentan los resultados alcanzados como publicaciones y desarrollos llevados a cabo. Por último, en la sección 1.5 describimos el esquema del resto del documento.

1.1. Planteamiento del Problema

Gracias a la evolución de las tecnologías de la comunicación y el creciente uso de las mismas, en muy poco tiempo muchas instituciones y universidades del mundo se han sumado al salto de la educación en línea, forzando y consiguiendo que la calidad de los

contenidos deba equiparse a la enseñanza presencial [129]. Esto, ha motivado el avance de nuevas modalidades de aprendizaje que hoy en día se han vuelto muy populares, tal es el caso de los Cursos en Línea Abiertos y Masivos (MOOC, Massive Open Online Course) [26, 118]. Cursos con una gran expansión en el e-learning, que han experimentado un rápido desarrollo y han recibido una atención considerable por muchas instituciones y universidades, presentándolos como una nueva tecnología disruptiva en el ámbito educativo.

Los MOOC son exactamente lo que su nombre indica, cursos en línea que heredan las ventajas del e-learning tradicional, en donde se destacan la versatilidad y flexibilidad. Se describen como versátiles porque se consideran apropiados para todo tipo de contenidos, y flexibles porque permiten el acceso a la educación en cualquier momento y lugar [26, 49, 97]. A diferencia de los cursos en línea convencionales, los MOOC son de carácter abierto y masivo, es decir, no requieren requisitos de inscripción y no cuentan con limitaciones en la participación. Cualquier persona a través de una conexión a internet puede inscribirse y tener acceso a los recursos y al conocimiento de profesores de prestigiosas universidades. El número de participantes puede ir desde cientos a miles de estudiantes, por lo tanto, la cantidad de personas que pueden beneficiarse del aprendizaje es mucho mayor que en los cursos en línea convencionales [26, 77, 123].

El interés generado a nivel mundial por los MOOC se debe a su gran potencial para ofrecer acceso a educación de calidad a cantidades masivas de personas y además disponible a cualquiera independientemente de la ubicación geográfica y de los conocimientos previos adquiridos [23, 78, 123]. Con esto han diversificado las opciones de educación para las personas que por una u otra razón no pueden asistir de manera presencial a las universidades. Además que han generado un gran número de ventajas tanto para los estudiantes como para las instituciones que los imparten [11, 93]. Se presentan como una alternativa a la gran demanda de educación superior, pues con el adecuado diseño del curso manteniendo la calidad académica, permiten el acceso a procesos formativos desarrollados en instituciones de educación superior a grandes cantidades de estudiantes [22, 137]. No sólo ayudan a cubrir dicha demanda, también generan ampliación de cobertura de la educación, mayor oferta académica, ampliación de acceso, comercialización y la posibilidad de desarrollar fuentes de ingreso [137].

Desde un inició los MOOC surgieron como una estrategia gratuita que permitía abarcar una gran cantidad de estudiantes sin un fin académico. Su propósito en principio es facilitar el acceso al conocimiento y por lo tanto no fueron pensados para ser certificados, no fueron creados para un ambiente universitario, ni para ser reconocidos académicamente. Pero gracias a las grandes ventajas y beneficios que conllevan los MOOC, se espera

que pronto puedan ser incorporados en las universidades y que cuenten con un reconocimiento académico [64, 132].

Para las instituciones de educación superior la adopción del modelo MOOC y su incorporación en sus programas formativos representa un reto tecnológico y pedagógico, el cual se está resolviendo gracias a estrategias como los Cursos en Línea Privados y Pequeños (SPOC, Small Private Online Courses) [43, 70] y Cursos en Línea Privados y Masivos (MPOC, Massive Private Online Courses) [53, 86], variantes de los MOOC que se caracterizan por ser limitados en acceso (cursos privados) y por lo tanto también en tamaño, pero siguen teniendo un mayor alcance de participación que cualquier curso en línea convencional [22, 70], permitiendo a las universidades la oportunidad de ampliar su cobertura y llegar a más estudiantes, se presentan como el nuevo camino para la expansión del conocimiento, la innovación universitaria, la empleabilidad y el desarrollo sostenible de escenarios de aprendizajes masivos, por lo que muchas universidades apuestan por su incorporación en la educación superior [77, 78, 123].

Sin embargo, así como surgen ventajas debido a su masividad, también se generan una serie de inconvenientes o desafíos, como: la escalabilidad de las plataformas de enseñanza, la evaluación de las cantidades masivas de estudiantes, el seguimiento a los participantes, la realimentación, entre otras, que de alguna manera complican la puesta en marcha de estos cursos y aún más cuando nos referimos a cursos reconocidos académicamente por instituciones de educación superior.

Aunque se encuentran varios casos en los cuales las universidades han adoptado el modelo MOOC, son pocas las experiencias encontradas en cuanto a cursos ofrecidos en instituciones de educación superior y que estos cuenten como créditos o sean reconocidos académicamente. Algunas de las dificultades identificadas a partir de las experiencias encontradas son:

- La falta de políticas o directrices que permitan el reconocimiento académico y su homologación en los programas de formación [51, 109, 140].
- Pocas herramientas que apoyen el proceso de aprendizaje de los estudiantes, entre ellas herramientas necesarias para el seguimiento del estudiante y la realimentación [62, 130].
- Preocupación por la calidad del diseño de la enseñanza, incluida la motivación y el apoyo a los estudiantes, que es una de las principales razones de la baja tasa de finalización de los MOOC [51, 109].

- Constante incertidumbre sobre la honestidad académica de los estudiantes relacionada a la verificación de la identidad (suplantación), verificación de la propiedad de su trabajo (fraude), supervisión de los exámenes (trampas durante los exámenes), etc.; No se está logrando garantizar que los estudiantes sean honestos. En la mayoría de casos están usando estrategias deshonestas para obtener los créditos académicos ofrecidos, lo cual genera una constante preocupación para permitir que los cursos sean reconocidos académicamente [64, 100, 109].
- Las plataformas de aprendizaje actuales no cuentan con las herramientas tecnológicas necesarias que permitan la identificación de las conductas deshonestas de los estudiantes a fin de lograr mitigar la deshonestidad académica en este tipo de cursos [101, 127].

El reconocimiento académico de los cursos genera un impacto importante en el comportamiento de los estudiantes. En principio, permite reducir las altas tasas de abandono con la que cuentan los MOOC [54], les impulsa a que estén pendientes del curso, hace que los estudiantes sean más autorregulados, participativos, etc., pues se tiene una mayor motivación para terminar el curso [60, 62].

Sin embargo, el reconocimiento entregado a los estudiantes hace que se comporten en función de las evaluaciones a presentar. Se ha identificado en cursos con reconocimiento académico que los estudiantes presentan una mayor actividad en plataforma exactamente en los horarios en los que se ha programado las evaluaciones sumativas [60]. Los estudiantes en este caso se inscriben al curso teniendo como objetivo principal el alcanzar la nota que les permitirá conseguir dicho reconocimiento [60, 62].

El reconocimiento y que los exámenes deban ser en línea conlleva a que se presenten conductas poco honestas con el fin de obtener mejores calificaciones [16, 82]. Entre las diferentes estrategias o comportamientos deshonestos en cursos MOOC se tienen: búsqueda de respuestas compartidas por otros participantes, suplantación de identidad, realizar exámenes en compañía de expertos en el tema, obtener respuestas correctas registrando varias cuentas, preguntar o buscar en foros de preguntas y respuestas relacionadas con el tema, etc. [16]. La deshonestidad académica en cursos en línea, hace que el reconocimiento de los MOOC por parte de instituciones de educación superior sea complicado [2, 64]. Lo cual ha limitado o demorado la incorporación de la estrategia MOOC al ámbito universitario, además que las conductas engañosas conducen a un aprendizaje deficiente [127].

Por otra parte, son pocas las plataformas que han incorporado mecanismos que tienen por objetivo identificar las conductas deshonestas de los estudiantes. Entre las más sobre-

salientes se encuentran los casos de Coursera y Edx, que han implementado mecanismos de vigilancia virtual del estudiante a través de cámaras web [35–37]. Sin embargo, son procesos complejos para el estudiante y requieren unos requisitos mínimos en cuanto a hardware y ancho de banda de conexión a internet, que limita su utilización en todos los contextos, sobre todo en el ámbito colombiano. Además, dichos procesos aún les falta lograr que sean completamente automáticos y así, ser escalables.

Ya se han encontrado propuestas que apuntan a entender el comportamiento del estudiante a través de su interacción con la plataforma de aprendizaje (Analíticas del aprendizaje, Learning Analytics) con el fin de lograr identificar e incluso predecir la deshonestidad antes de identificar dichas conductas [80, 100, 101, 127]. Sin embargo, estos estudios y los algoritmos propuestos han sido validados mediante análisis retrospectivos, en lugar de aplicar los algoritmos en tiempo real. Es decir, no hemos encontrado en las plataformas de aprendizaje actuales herramientas tecnológicas necesarias que permitan identificar los comportamientos deshonestos de los estudiantes y sirvan de apoyo a los instructores encargados de los cursos.

Considerando lo anterior y de acuerdo con trabajos analizados, surge la siguiente pregunta:

¿Cómo identificar conductas con sospecha de fraude en Cursos en Línea Privados y Pequeños con reconocimiento académico para brindar apoyo a los instructores encargados de los cursos?

Con la respuesta a esta pregunta se espera aportar a la identificación y caracterización de la deshonestidad académica en SPOC para contribuir en los procesos de reconocimiento académico de los cursos. Se busca que a través de la analítica del aprendizaje (estudio del comportamiento del estudiante mediante el análisis de los datos de interacción registrada en plataforma), técnicas que ayudan a comprender el proceso de aprendizaje, se pueda contribuir en la caracterización de las conductas deshonestas de los estudiantes y permita su identificación. Sería valioso proporcionar pautas instructivas que puedan ser utilizadas para identificar el engaño en SPOC y así facilitar su acreditación e incorporación al ámbito universitario. Con esto, también proponemos el diseño de herramientas que permitan identificar y hacer seguimiento a los estudiantes que presenten comportamientos con sospecha de deshonestidad académica, con el objetivo de que los instructores puedan conocer los comportamientos de sus estudiantes, permitiendo una mejor comprensión e intervenciones durante el desarrollo del curso.

1.2. Objetivos

Como mencionamos en la Sección 1.1, esta tesis se centra en contribuir a la identificación y caracterización de conductas deshonestas en SPOC mediante el análisis de datos. Analizando la actividad de los estudiantes como interacción con contenidos, publicación en foros de discusión y resultados alcanzados, que nos permita entender los comportamientos de deshonestidad académica en línea y las colaboraciones ilícitas.

Los resultados potenciales de esta investigación van hacia el desarrollo de aplicaciones basadas en el análisis del aprendizaje y el análisis de algoritmos de comportamiento. Las lecciones aprendidas en términos de comprensión del comportamiento de los estudiantes nos ayudan a inferir la actividad y motivaciones de los estudiantes, obtener recomendaciones de diseño instructivo, conocer perfiles de los estudiantes con fines de apoyo a los instructores encargados de los cursos. Más concretamente, los objetivos específicos son los siguientes:

- Proponer métricas a partir de un conjunto de indicadores que contribuyan a la descripción de los comportamientos deshonestos en MPOC.
- Diseñar estrategias que a través de Analíticas del Aprendizaje facilite la identificación de comportamientos potencialmente deshonestos de los estudiantes sobre una plataforma MOOC.
- Verificar las métricas y estrategias propuestas para la caracterización de los comportamientos deshonestos mediante el desarrollo de una prueba de concepto en un MPOC en la Universidad del Cauca.

1.3. Actividades desarrolladas

Para la elaboración del cronograma de actividades y con el fin de dar cumplimiento a los objetivos planteados, se usó como referencia la descomposición jerárquica WBS (Works Breakdown Structure) sugerida en la metodología PMBOK (Project Management Base of Knowledge) por el PMI (Project Management Institute), específicamente en el área de gestión del alcance (Scope Management) [59].

1.3.1. WP1. Generación de la base inicial de conocimiento

- Definición del objetivo de la vigilancia (planeación).

- Construcción del corpus (búsqueda y captación).
- Análisis del corpus (análisis)

1.3.2. WP2. Descripción general de indicadores para la deshonestidad académica en Cursos en Línea Privados y Masivos

- Definición de los criterios para la elección de indicadores.
- Búsqueda y análisis de indicadores para la deshonestidad académica en cursos en línea.
- Especificación de los indicadores seleccionados.

1.3.3. WP3. Construcción de un prototipo para el análisis del comportamiento de la deshonestidad académica con base en los indicadores seleccionados

- Especificación de casos de estudio.
- Análisis de las técnicas que permitan la construcción de un prototipo para el estudio del comportamiento de los estudiantes.
- Análisis de las técnicas que permitan la construcción de un prototipo para la identificación y caracterización de comportamientos deshonestos en SPOC con reconocimiento académico.
- Implementación del prototipo sobre una instancia de Open edX.

1.3.4. WP4. Validación del prototipo mediante su aplicación en MPOC/SPOCs con reconocimiento académico ofrecidos en la Universidad del Cauca

- Desarrollo de los casos de estudio.
- Obtención de resultados.
- Análisis de los resultados obtenidos.

1.3.5. WP5. Publicación

- Generación de artículos sobre los avances de la investigación.
- Generación del documento final (monografía).

1.4. Resultados Alcanzados

En el proyecto concluido se lograron los siguientes resultados:

- Síntesis del estado del conocimiento, donde se muestra el avance que se a realizado hasta el momento con relación a la deshonestidad académica en MOOCs y SPOCs.
- Definición de un marco teórico y metodológico para la identificación de comportamientos con sospecha de deshonestidad académica.
- Documentación sobre las métricas y algoritmos propuestos para la identificación de estudiantes con sospecha de deshonestidad académica en SPOCs con reconocimiento académico.
- Resultados sobre el uso de las métricas y algoritmos propuestos sobre SPOCs de la Universidad del Cauca.
- Diseño de mecanismos para la identificación y seguimiento de conductas de estudiantes con sospecha de deshonestidad académica según las métricas y algoritmos propuestos.
- Implementación de una herramienta para la identificación y seguimiento de conductas con sospecha de deshonestidad académica.
- Resultados de la evaluación de la herramienta.
- Publicaciones y/o ponencias de los siguientes artículos:
 - Publicación de artículo y conferencia: “*Follow-Up of Learning Activities in Open edX: A Case Study at the University of Cauca*” en Digital Education: Out to the World and Back to the Campus EMOOCs (2017). Lecture Notes in Computer Science. DOI: https://doi.org/10.1007/978-3-319-59044-8_26
 - Publicación de artículo: “*Hábitos de ingreso y resultados en las evaluaciones en cursos en línea masivos con reconocimiento académico*” en Revista Ingeniería e Innovación (2017). DOI: <https://doi.org/10.21897/23460466.1105>

- Conferencia: “*Estrategia de seguimiento a las actividades de aprendizaje de los estudiantes en cursos en línea masivos y privados (MPOC) con reconocimiento académico en la Universidad del Cauca*” en Séptima Conferencia de Directores de Tecnología de Información, TICAL 2017 Gestión de las TICs para la Investigación y la Colaboración (2017). Disponible en: <https://n9.cl/ba872>
- Conferencia: “*Patrones de acceso en cursos en línea masivos con reconocimiento académico*” en VIII Congreso Internacional sobre Aplicación de Tecnologías de la Información y Comunicaciones Avanzadas (ATICA2017) (2017). Disponible en: <https://n9.cl/jaramillo>
- Publicación de artículo: “*Hacia la creación del MOOC para el Mejoramiento en el Tránsito a la Educación Superior*” en Revista Ingeniería e Innovación (2019). DOI: <https://doi.org/10.21897/23460466.1711>
- Publicación de artículo: “*Identifying and Characterizing Students Suspected of Academic Dishonesty in SPOCs for Credit through Learning Analytics*” en International Journal of Educational Technology in Higher Education (2020). DOI: <https://doi.org/10.1186/s41239-020-00221-2>
- Publicación de artículo: “*Data-driven detection and characterization of communities of accounts collaborating in MOOCs*” en Future Generation Computer Systems (2021). DOI: <https://doi.org/10.1016/j.future.2021.07.003>

1.5. Contenido de la Monografía

El resto del documento está organizado de la siguiente manera:

El Capítulo 2 presenta el análisis de diferentes estudios relacionados con las áreas en relación con este trabajo. Nosotros describimos diferentes trabajos que están relacionados a la certificación y reconocimiento, deshonestidad académica y su identificación, y herramientas o aplicaciones destinadas a el seguimiento y visualización de los comportamientos de los estudiantes. Todo esto en el contexto de los MOOCs, MPOCs y SPOCs.

El Capítulo 3 presenta los métodos y herramientas utilizadas. Nosotros describimos diferentes herramientas que usamos, los casos de estudio analizados, la selección de indicadores que usamos durante todo el desarrollo del trabajo, el algoritmo para la identificación de estudiantes con comportamientos sospechosos de deshonestidad académica y el método utilizado para el análisis de colaboraciones deshonestas entre estudiantes.

El Capítulo 4 es uno de los primeros capítulos de resultados y presenta la validación del algoritmo para la identificación de comportamientos sospechosos de deshonestidad académica en SPOCs. Este capítulo presenta cómo el algoritmo fue aplicado sobre los datos de dos SPOCs de la universidad del Cauca.

El Capítulo 5 presenta una caracterización de las colaboraciones deshonestas identificadas con ayuda del algoritmo propuesto. En este capítulo se hace un análisis de los comportamientos de los estudiantes identificados como sospechosos de deshonestidad académica y la relación que hay entre ellos (colaboraciones) mediante el uso de una técnica de agrupamiento (clustering).

El Capítulo 6 presenta la primera herramienta construida para el seguimiento a las actividades de aprendizaje de los estudiantes y su validación en SPOCs ofrecidos en la Universidad del Cauca. En esta herramienta no se tuvo en cuenta la identificación de los estudiantes con sospecha de deshonestidad académica, pero presenta una propuesta de seguimiento a los estudiantes como apoyo a los instructores para el desarrollo de los cursos.

El Capítulo 7 presenta un estudio sobre la evaluación de una aplicación web desarrollada que incorpora identificación de estudiantes con comportamientos sospechosos de deshonestidad académica y seguimiento a sus actividades. Aquí se realiza un análisis de la percepción de los instructores sobre las métricas para la identificación de estudiantes y la utilidad de la herramienta desarrollada.

En el capítulo 8 se discute sobre los resultados obtenidos, presenta recomendaciones para el uso de los hallazgos encontrados para mejorar los procesos de aprendizaje.

Finalmente en el Capítulo 9 se presentan algunas observaciones finales, limitaciones a nuestro trabajo e ideas para trabajos futuros.

Parte I

Trabajos relacionados, herramientas y métodos

Capítulo 2

Trabajos relacionados

Índice

3.1. Plataforma MOOC	29
3.1.1. Plataforma Selene	29
3.1.2. Características de Open edX	31
3.2. Casos de estudio	35
3.2.1. Curso Virtual de Astronomía Cotidiana	36
3.2.2. Introducción al emprendimiento con Lean Startup	37
3.2.3. Comprensión de Textos Argumentativos	38
3.2.4. La danza folclórica como patrimonio cultural	38
3.2.5. Introducción a la edición de texto científicos y literarios con LaTeX	39
3.3. Indicadores seleccionados	40
3.3.1. Interacciones con la plataforma	40
3.3.2. Extracción de datos para medición de indicadores de interacción	43
3.3.3. Identificación de estudiantes sospechosos de deshonestidad académica	46
3.3.4. K-means algorithm	48

Este capítulo presenta los trabajos relacionados con nuestro trabajo y que abarcan las distintas áreas que se han investigado y que están en relación con los objetivos. Este capítulo se divide en cuatro secciones. La sección 2.1 se centra en los diferentes tipos de certificación y reconocimiento que existen en los MOOC y SPOC. La sección 2.2 se

relaciona con trabajos previos sobre deshonestidad académica. La sección 2.3 revisa la analítica del aprendizaje aplicada en aplicaciones y herramientas de visualización. Por último, la sección 2.4 está relacionada con los posibles enfoques para evaluar herramientas de analíticas del aprendizaje.

2.1. Certificación y reconocimiento en MOOC/MPOC/S-POC

Aunque los MOOCs nacieron con un enfoque informal, cuyo objetivo principal era proporcionar conocimiento abierto a cualquier persona, con el tiempo se hicieron muy populares y empezaron a surgir diferentes tipos de certificación. El primer y más popular tipo de reconocimiento que se encuentra en la literatura es la certificación de participación. Se trata de la certificación más sencilla que se otorga, sin efectos académicos, y suele ser emitida por el proveedor de la plataforma MOOC a aquellos alumnos que completan un determinado porcentaje de las actividades de aprendizaje propuestas, normalmente de forma gratuita y descargable en línea desde la plataforma de gestión del aprendizaje [3]. Sin embargo, este tipo de certificaciones están desapareciendo de la mayoría de las plataformas y se está produciendo un cambio hacia las certificaciones de finalización y reconocimientos de créditos, todo ello de pago para un modelo más sostenible.

Los certificados de finalización o credenciales de aprovechamiento exigen el cumplimiento de unos requisitos más estrictos que la certificación de participación y, en algunos casos, un pago previo. Suelen ser emitidos por la institución responsable del diseño, impartición y evaluación del MOOC. Aunque tampoco suelen tener efectos académicos, su principal objetivo es certificar que el alumno alcanzó unos objetivos mínimos en el curso [3]. En algunas de las experiencias para entregar un certificado de finalización se realiza un control de identidad supervisado, es decir, el alumno debe realizar un examen en línea que es monitorizado durante todo el tiempo por un evaluador a través de una cámara web o desplazarse a la institución que ofrece el curso para realizar un examen presencial. Algunas instituciones contratan a instituciones externas presentes en las ciudades de los estudiantes para realizar un examen presencial supervisado.

Por último, existe un certificado de convalidación, que tiene reconocimiento académico o proporciona créditos que son válidos en las carreras profesionales. Lo emite la institución responsable del diseño, oferta y evaluación del MOOC e indica el número de créditos equivalentes de convalidación u homologación. como en el caso de los Micro-

Masters de edX la mayoría de los cursos que han proporcionado créditos válidos han sido desarrollados como SPOCs y han sido ofrecidos por instituciones de educación superior que han querido incorporar los MOOCs a sus programas profesionales, pero han ofrecido los cursos a un grupo privado y reducido de participantes [17, 60]. Además, hay muchos MOOCs que aunque ofrecen créditos para ser aprobados por instituciones de educación superior, los créditos sólo se conceden a los participantes que han pagado por ellos. Esto significa que algunos contenidos y evaluaciones no están disponibles para todo el mundo, sino sólo para el grupo de participantes de pago.

Por otro lado, en la mayoría de las experiencias de cursos que convalidaron créditos académicos, se realizaron sin controles de identidad y los instructores expresaron su preocupación por validar la identidad de los alumnos, ya sea mediante exámenes proctorados o proponiendo controles de identidad a través de herramientas en línea. En algunos casos, para la obtención de créditos se exige una prueba de evaluación no presencial o en línea con supervisor personal, en la que se verifica la identidad del alumno y se supervisa la presentación del examen para garantizar que no haya conductas deshonestas.

Cabe destacar que en la mayoría de las experiencias no existe un control del comportamiento del alumno en la realización de las actividades de evaluación (control de identidad, verificación de la propiedad del trabajo, proctorado durante los exámenes, etc.). Son muy pocos los casos en los que las instituciones realizan un control de la honestidad académica. Basándonos en la literatura, hemos generado la siguiente figura (Figura 2.1), que muestra las relaciones entre los diferentes tipos de reconocimiento académico y la estrategia de control de identidad utilizada.



Figura 2.1: Diagrama de relaciones de los tipos de reconocimiento y control de identidad en MOOC.

2.2. Deshonestidad académica en MOOCs

Esta sección se divide, según la figura 2.1, en las tres categorías siguientes: Exámenes sin control de deshonestidad o exámenes no supervisados, exámenes supervisados por los instructores y exámenes supervisados automáticamente.

2.2.1. Exámenes sin control de identidad o sin supervisión

Los MOOCs surgieron como una estrategia gratuita para llegar a un gran número de estudiantes sin una finalidad académica. Su objetivo era facilitar el acceso al conocimiento y, por lo tanto, no pretendían ser reconocidos académicamente ni ser válidos para obtener créditos académicos. Por este motivo, la mayoría de las experiencias MOOC no cuentan con un control de identidad ni con un seguimiento del comportamiento de los alumnos durante sus actividades de evaluación. Por ejemplo, en Kenia se desarrolló un MOOC sobre derechos humanos, cuyo objetivo era llegar al mayor número posible de alumnos y para ello ofrecían un certificado digital de participación de la Universidad de Estocolmo (Suecia) [65]. sin embargo, no tenían ninguna comprobación de la honestidad de los alumnos. Esta fue una de las primeras experiencias de un curso certificado sin control de la honestidad de los alumnos.

Desde entonces, empezaron a surgir diferentes cursos que otorgan certificados de finalización de manera no formal, e incluso insignias digitales [64]. Esta tendencia se inició porque los estudiantes tenían un claro interés en obtener un reconocimiento académico de la finalización del curso. A partir de estas experiencias, se está empezando a explorar la posibilidad de combinar validaciones no formales, como insignias y certificados, con créditos formales [24, 90]. Por ejemplo, en el caso de los MicroMasters, algunas universidades (MIT, Galileo University, Monterey Tech), de acuerdo con edX, convalidan certificados obtenidos por asignaturas que forman parte de un programa de Máster. Una situación similar encontramos con los certificados profesionales que ofrece Coursera [24, 58, 90].

Otra modalidad que ha surgido es el reconocimiento de los cursos basados en MOOC como créditos dentro de los programas profesionales de las instituciones de educación superior [18, 60]. Un ejemplo se presentó en la Universidad de X, donde se pueden reconocer créditos válidos para cualquier programa de grado de la institución a cualquier estudiante que obtenga una calificación de aprobado en uno de sus SPOCs ofrecidos internamente. La situación actual, en la que cada vez hay más cursos en línea que conceden créditos, hace más necesario disponer de medios que garanticen la honestidad de los estu-

diantes a la hora de realizar sus trabajos y exámenes. Sin embargo, no hemos encontrado en la literatura ninguna herramienta que pueda ayudar a los instructores a mejorar esta cuestión.

2.2.2. Exámenes supervisados personalmente

La falta de supervisión del entorno de los estudiantes al realizar los exámenes ha sido un gran reto para la educación superior a distancia. Existe una clara correlación entre el aumento de la deshonestidad académica y la ausencia de políticas claras de honestidad institucional dentro de los cursos, y más aún cuando los exámenes se realizan en línea [82]. Desde que los MOOC se hicieron muy populares, ha habido un debate sobre la evaluación en el entorno de aprendizaje en línea. Se ha argumentado que hay problemas significativos en la evaluación para otorgar el crédito académico en los MOOC, debido a la autenticación de los estudiantes y la supervisión de los exámenes en línea [108], ya que la institución necesita garantizar que la persona que completó las evaluaciones es la misma que se inscribió en el curso.

Existen diferentes técnicas de autenticación y verificación. Entre ellas, la opción de que el estudiante se desplace al centro que ofrece el curso para realizar el examen en persona. Por ejemplo, la Universidad de Cincinnati desarrolló un MOOC que ofrecía créditos académicos a los participantes y en el que los estudiantes debían superar un examen presencial exhaustivo. Este trabajo pone de manifiesto la necesidad de exámenes supervisados para la concesión de créditos educativos.

Existen diferentes técnicas de autenticación y verificación. Entre ellas, la opción de que el estudiante se desplace al centro que ofrece el curso para realizar el examen en persona. Por ejemplo, la Universidad de Cincinnati desarrolló un MOOC gratuito que ofrecía créditos académicos a los participantes que completaran un MOOC y se inscribieran posteriormente en un programa de certificado o título. Los estudiantes que se matriculaban en los cursos con créditos debían superar un examen presencial exhaustivo supervisado. Este trabajo pone de manifiesto la necesidad de exámenes supervisados para la concesión de créditos educativos.

Otra opción es la monitorización en línea a través de cámaras. Por ejemplo, Deider en 2020 presentó un software de grabación a través de cámaras web llamado Respondus Monitor (RM) [36], en este trabajo, los alumnos eran monitorizados a través de cámaras web en dos instantes concretos de tiempo. El estudio descubrió que había una disminución en el rendimiento de los estudiantes cuando sabían que estaban siendo monitorizados. Por lo tanto, los autores concluyeron que el monitoreo en línea es una herramienta eficaz para

mitigar la deshonestidad académica en los cursos en línea. El enfoque aquí es permitir a los estudiantes hacer el examen en cualquier lugar y en cualquier momento. El audio, el vídeo y la pantalla compartida del ordenador del estudiante (incluyendo la cámara web y el micrófono) se transmiten de forma encriptada al servidor del servicio de supervisión y se graban. A continuación, un supervisor capacitado reproduce estas grabaciones a velocidad acelerada y señala cualquier actividad sospechosa mediante anotaciones para que el profesorado la tenga en cuenta. Otra plataforma de supervisión es RPNOW, que utiliza este enfoque y permite a los instructores recibir y verificar rápidamente los resultados de los exámenes de los estudiantes una vez que los supervisores han completado su proceso de revisión [35].

Un estudio similar se llevó a cabo en 2020 durante la pandemia COVID-19. En este caso, se probaron diferentes plataformas de conferencias (Microsoft Teams, Zoom, etc.) para generar grupos y supervisar a los estudiantes a través de cámaras web. Inicialmente, se comenzó con grupos de 30 estudiantes y posteriormente se aumentó a 100. Esta situación puso de manifiesto la necesidad de supervisar automáticamente a los alumnos durante el examen, ya que la supervisión humana se hace difícil con el aumento del número de alumnos [55]. En la misma línea, se presentó otro caso de estudio de proctorado humano utilizando una cámara de seguridad de 360 grados, donde se menciona que la mejora de las características de la cámara utilizada mejora la seguridad del proctorado. Sin embargo, el hecho de que en el año 2020 todavía existan estudios sobre el proctorado humano, implica que todavía existen lagunas en el proctorado automático y por tanto es necesario profundizar en este tema, incluso explorando otras técnicas que permitan identificar comportamientos deshonestos como el análisis de los datos de interacción de los alumnos con la plataforma de aprendizaje y no sólo utilizando la webcam [126].

2.2.3. Exámenes supervisados automáticamente

Esta sección se refiere al trabajo relacionado con la supervisión automática. Esto significa que no es necesario tener personal para supervisar lo que hace el alumno durante un examen en línea. Clasificamos los trabajos relacionados en dos grandes categorías, la identificación y autenticación del alumno, y la detección de comportamientos inadecuados, que se muestran en la Figura 2.2.

2.2.3.1. Métodos de identificación y autenticación

Una de las preocupaciones a la hora de certificar cursos online es comprobar la identidad de los alumnos. Es necesario saber que la persona que va a recibir la certificación es la

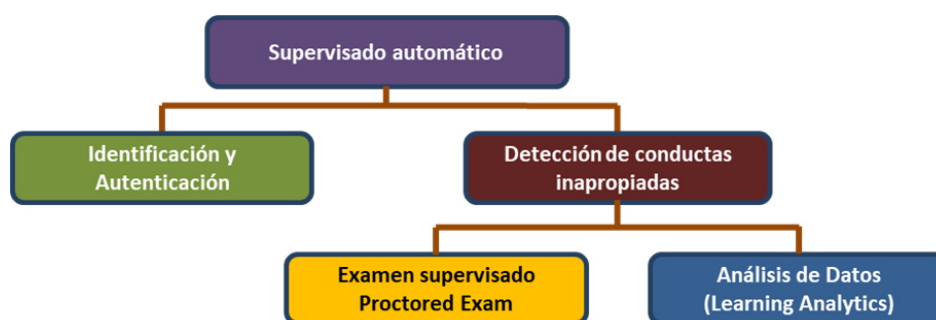


Figura 2.2: Esquema de división de trabajos relacionados a la supervisión automática de las evaluaciones en línea.

misma que realizó las actividades del curso. En 2016, se presentó un experimento con una propuesta de técnica para la identificación y autenticación en línea que en lugar de depender de la verificación manual de un examinador, utilizaba una comparación automática utilizando un modelo matemático de identificación facial de los estudiantes [117]. Destacaron la importancia de la identificación de los estudiantes para la entrega de certificados y el reconocimiento de los cursos realizados. Por ejemplo, informaron de un caso en el que un usuario engañó al sistema sosteniendo una foto suya delante de la cámara. Por ello, se recomienda que un examinador humano realice una segunda revisión de los datos para mejorar la fiabilidad e integridad de los exámenes online con reconocimiento académico.

Además de estas técnicas de reconocimiento facial para la autenticación de estudiantes, existen otras técnicas que mezclan la identificación del estudiante y el análisis de los datos con métricas como el movimiento del ratón y las pulsaciones del teclado (es decir, la interacción del estudiante con el sistema) [41, 124]. Se trata de sistemas multi-biométricos para la autenticación continua de los alumnos en plataformas de e-learning, que en función del dispositivo utilizado y de la interacción realizada con el mismo, el sistema es capaz de atestiguar la identidad del alumno a lo largo de la experiencia de aprendizaje.

Estos sistemas se presentan como herramientas más transparentes desde el punto de vista del alumno, ya que no requieren dispositivos adicionales ni acciones explícitas, abriendo un nuevo enfoque hacia la identificación de comportamientos deshonestos no sólo a través de la webcam, sino también mediante el seguimiento de los datos de las interacciones del alumno.

2.2.3.2. Detección de comportamientos deshonestos

La vigilancia de los exámenes es una tarea compleja. Es decir, el control de las actividades de los estudiantes se hace difícil para los vigilantes en las salas de examen, más aún cuando los cursos son en línea. Se trata de un enfoque costoso y que requiere mucho trabajo. Además, es una tarea difícil controlar a todos los estudiantes al mismo tiempo. Por lo tanto, el reconocimiento automático de las actividades de los estudiantes en los exámenes en línea es un tema de investigación necesario.

Existen diferentes trabajos destinados a controlar automáticamente ciertos comportamientos inadecuados de los alumnos durante los exámenes programados en los cursos [12, 13, 84]. De acuerdo con lo que hemos encontrado en la literatura, hemos clasificado estos trabajos como proctored o learning analytics-driven, como se muestra en la Figura 2.2. El primer grupo se refiere a los enfoques que utilizan elementos multimedia, como la cámara web o el micrófono del alumno, para supervisar y analizar su comportamiento de forma automática. Los estudios del segundo grupo de trabajos, utilizan la analítica de datos de la interacción con la plataforma para detectar la deshonestidad académica de los estudiantes.

En cuanto a los sistemas de supervisión de exámenes basados en cámaras web, hay muchos trabajos que lo hacen. Por ejemplo, algunos trabajos tienen en cuenta comportamientos anómalos, como enfocar los ojos del alumno fuera de la pantalla durante mucho tiempo, personas distintas al alumno en el vídeo, las voces de otras personas, entre otros comportamientos que podrían ser catalogados como deshonestos [33, 92]. Además, existen técnicas que no solo implican que el alumno sea observado a través de la webcam, sino que las cámaras pueden ser utilizadas para averiguar lo que el alumno está viendo. Atoum en 2017 presentó un sistema totalmente automatizado que utiliza dos cámaras y un micrófono para monitorizar las acciones del alumno durante la presentación de los exámenes [13]. Su planteamiento se basa en la monitorización automática continua del alumno a través de una webcam de escritorio y una cámara wearable, que proporciona un grado muy alto de visibilidad, y la monitorización de audio. Con los datos de estos dispositivos, utilizaron un clasificador para detectar comportamientos fraudulentos. Validaron el sistema con 24 estudiantes que realizaron diferentes tipos de trampas mientras hacían un examen online. La evaluación una precisión del 87 % en la detección de trampas.

Estos estudios previos revisados plantean diferentes puntos de vista. Mientras que algunos trabajos mencionan la identificación del comportamiento tramposo, otros mencionan los comportamientos anómalos, o la sospecha de comportamiento tramposo. Esto se debe, por un lado, a que todavía no existe una caracterización completa de lo que es un

comportamiento tramposo y, como factor cultural, podría cambiar de un país a otro o en diferentes cursos. Por otro lado, la identificación de comportamientos anómalos puede no implicar necesariamente la presencia de un comportamiento deshonesto, por lo que también tenemos la necesidad de que un examinador humano revise los casos.

Ahora bien, en cuanto a las técnicas de análisis de datos, Coursera ha sido una de las primeras instituciones en desarrollar mecanismos complejos basados en la biometría de las pulsaciones del teclado (patrones de tecleo) para verificar la identidad de los participantes durante las sesiones de los MOOC [37]. Sin embargo, se menciona que este método es poco fiable, complejo y no detecta el fraude, sólo verifica la identidad del estudiante.

Centrándonos en la detección de comportamientos tramposos, encontramos diferentes trabajos que tienen en cuenta las interacciones de los alumnos con la plataforma. Los trabajos que destacan en el ámbito de los MOOC están relacionados con algoritmos para detectar CAMEO (Copying Answers using Multiple Existences Online), uno de los métodos de engaño reportados en el que se utilizan múltiples cuentas para obtener las respuestas correctas a los exámenes, que luego son enviadas por una cuenta maestra; todas estas cuentas son utilizadas por el mismo alumno para obtener un certificado. Algunos de estos estudios detectaron estos comportamientos basándose en la IP del alumno, asumiendo que todos los envíos desde la misma IP podían pertenecer al mismo alumno, y así identificando y etiquetando los envíos realizados con el método CAMEO [9, 16, 87, 105]. Los algoritmos de segunda generación eliminaron la fuerte dependencia de la PI, utilizando bien técnicas de aprendizaje automático [105], bien combinando psicometría y detección de anomalías [8]. Sin embargo, aunque estos trabajos mencionan que los métodos de engaño mencionados y las técnicas de detección son relevantes para otros entornos online distintos a los MOOCs, la generalización no está clara ya que, por ejemplo, CAMEO sí requiere que los usuarios puedan crear múltiples cuentas en la plataforma, algo que puede no ser factible en entornos académicos o controlados por K-12.

Sobre la base de estos trabajos anteriores, han surgido otros estudios que no utilizan las direcciones IP para detectar comportamientos deshonestos, ya que esto es fácil de ocultar por los estudiantes. Por ejemplo, se demostró que los métodos basados en anomalías desarrollados en [8] también pueden detectar otros tipos de trampas en MOOCs [10]. El trabajo actual se basa de forma importante en [104], que se centra en la detección de enlaces colaborativos “invisibles” entre estudiantes. Con base en su algoritmo, desarrollamos un sistema para detectar la colaboración no autorizada (compartir respuestas) teniendo en cuenta dos métricas, la similitud de las respuestas enviadas por los estudiantes y la cercanía en las marcas de tiempo de envío [61], esta propuesta fue validada con

dos cursos de SPOC en la Universidad del Cauca y estos fueron los algoritmos que se tuvieron en cuenta para la creación del tablero de control presentado en este trabajo.

2.3. Colaboraciones en MOOCs

En esta sección, nos centramos en presentar una visión general de las investigación que están más estrechamente relacionadas con nuestro trabajo. En primer lugar, en la subsección 2.3.1 revisamos los estudios que han aplicado técnicas de minería de datos educativos y análisis de aprendizaje para modelar el comportamiento de los estudiantes. A continuación, en la subsección 2.3.2 nos centramos en los estudios que han analizado el comportamiento de colaboración en los MOOC/MPOC/SPOC.

2.3.1. Análisis del comportamiento de los estudiantes en los MOOCs

Existe una gran diversidad en el tipo de trabajos publicados en el contexto del modelado de los estudiantes en los MOOCs. Gran parte de ellos se han centrado en el modelado de las motivaciones de los estudiantes para participar en estos cursos y sus preferencias [5,96,139]. Varios estudios se han centrado específicamente en la motivación de los estudiantes con las características de gamificación, por ejemplo, para analizar sus percepciones hacia la obtención de insignias en un MOOC gamificado [89] o para proponer métricas que permitan inferir qué estudiantes están ganando insignias intencionadamente [102]. Estos estudios pretenden comprender mejor las motivaciones de los alumnos de los MOOC para adaptar los materiales y atender mejor las necesidades e intereses de los alumnos.

Otro propósito predominante de modelar el comportamiento de los estudiantes ha sido predecir la deserción de los estudiantes en los MOOC. Por ejemplo, [54] presentó un predictor de abandono basado en la actividad de interacción de los estudiantes con la plataforma MOOC que puede proporcionar un factor de riesgo de abandono fiable. [96] también presentó un marco para modelar y comprender el compromiso de los estudiantes en los cursos en línea basado en datos de rastreo, utilizando un modelo probabilístico para conectar el comportamiento de los estudiantes con la finalización del curso. Estos estudios han buscado la posibilidad de implementar sistemas que puedan ayudar a mejorar la finalización de los MOOC.

Además, otra línea de investigación clave en los MOOCs ha sido la investigación de qué comportamientos afectan a los resultados del aprendizaje. Por ejemplo, [4] realizó dos experimentos para analizar qué características de comportamiento estaban relaciona-

das con los niveles de compromiso y los resultados positivos del aprendizaje. Además, [106] realizó un estudio sobre una instancia de Khan Academy construyendo un modelo de predicción de ganancias de aprendizaje que incluía diferentes indicadores de actividad y datos de comportamiento. Encontraron una serie de comportamientos correlacionados positivamente con las ganancias de aprendizaje (por ejemplo, los estudiantes que siguen las recomendaciones hechas por los instructores del curso), mientras que otros estaban correlacionados negativamente (por ejemplo, los comportamientos irreflexivos). Los resultados de este tipo de estudios pueden ayudar a entender a los instructores e investigadores qué comportamientos pueden tener un impacto positivo o negativo en los resultados del aprendizaje, y así permitir la posibilidad de promover o desalentar ciertos comportamientos.

Un gran número de estudios de clustering en MOOCs han aplicado estas técnicas para encontrar diferentes perfiles de comportamiento de los estudiantes en función de cómo interactuaban con las actividades [25, 42, 68, 102]; hay matices entre estos estudios, por ejemplo [102] tenía como objetivo inferir perfiles de compromiso con respecto a las características de gamificación de Khan Academy, [25] se centró en la extracción de patrones de estrategias de aprendizaje autorregulado, y tanto [42] como [68] se centraron en la extracción de diferentes subpoblaciones de estudiantes en función de cómo se involucraron con las actividades. Otros estudios han aplicado el clustering con fines alternativos; por ejemplo, [76] utilizó enfoques de clustering para proporcionar recomendaciones personalizadas de MOOCs a los usuarios en función de sus características o [121] lo aplicó para estudiar diferentes perfiles de participación en foros de discusión de MOOCs. Además, el clustering también se ha utilizado dentro de los estudios de los MOOC para la formación de grupos. Por ejemplo, [79] lo utilizó para agrupar a alumnos con perfiles similares para el proceso de revisión por pares y [110] lo utilizó para agrupar a alumnos similares para actividades de aprendizaje colaborativo. Como vemos, la mayoría de los estudios han utilizado el clustering bien para encontrar perfiles de alumnos en los MOOCs, con fines de recomendación, o para la formación de grupos con el fin de desarrollar algún tipo de actividad entre compañeros. Sin embargo, hasta donde sabemos, el clustering no se ha aplicado dentro de los MOOCs con el propósito de caracterizar las colaboraciones.

Los estudios mencionados en esta subsección han demostrado diversos propósitos para realizar el modelado del comportamiento de los alumnos de los MOOC. Sin embargo, a pesar de que la colaboración de los estudiantes es una de las oportunidades más destacadas en los MOOCs, pocos trabajos han reportado resultados sobre el modelado de comportamiento que se realiza para detectar o caracterizar la colaboración en los

MOOCs.

2.3.2. Colaboración en los MOOCs

Aunque numerosos estudios se han centrado en analizar cómo se comportan los estudiantes en los entornos MOOC, sólo unos pocos han profundizado en las colaboraciones de los estudiantes. En esta dirección, [29] presentó varias reflexiones sobre los procesos de seguimiento y evaluación desde dos sistemas de aprendizaje colaborativo: El primero se definió con el objetivo de involucrar a los estudiantes en un proceso social en torno a la composición de objetos de aprendizaje multimedia interactivos, mientras que el segundo buscaba ayudar a los instructores en el diseño de escenarios de aprendizaje colaborativo con un conjunto de servicios embebidos en Moodle. Al experimentar con estos dos enfoques de aprendizaje colaborativo, los autores proporcionaron recomendaciones sobre cómo aplicar estos enfoques a los MOOC para reducir la carga de trabajo de los instructores. Sin embargo, no analizaron las colaboraciones e interacciones entre los estudiantes que tuvieron lugar en los cursos.

Por otra parte, la mayoría de las plataformas MOOC ofrecen una funcionalidad técnica limitada para el trabajo colaborativo. Después de examinar el apoyo a la colaboración en las plataformas MOOC de Coursera, edX, Udacity y MiriadaX, [116] alentó el trabajo futuro para mejorar las funciones de apoyo al aprendizaje colaborativo en los MOOC. Basándose en el análisis, los autores implementaron un conjunto de herramientas que pueden apoyar la colaboración en la plataforma MOOC OpenHPI. Este conjunto de herramientas consistía en un espacio virtual general para el aprendizaje colaborativo en línea, que admite grupos de estudio, aprendizaje centrado en temas y equipos en grupos de trabajo tanto públicos como privados. Para la comunicación en línea, se añadió una combinación de herramientas síncronas y asíncronas, como un espacio de colaboración de laboratorio que ofrece a los grupos de aprendizaje la oportunidad de compartir artefactos. [115] continuó esta línea de trabajo examinando las implicaciones prácticas de algunas formas de aprendizaje colaborativo que se implementaron en la plataforma OpenHPI. La conclusión más importante de su estudio fue que el número de participantes que contribuían al foro aumentaba considerablemente cuando los instructores participaban en el proceso colaborativo. Sus resultados también confirmaron que la participación en los foros de los MOOCs funciona mejor con un gran número de participantes, ya que tanto los estudiantes como los instructores son más activos porque hay más interacciones en los foros.

Varios estudios han analizado los efectos que la colaboración puede tener en los dife-

rentes resultados del aprendizaje en los MOOC. En este sentido, [20] investigó si participar en un MOOC con amigos o compañeros puede mejorar tanto la finalización del curso como la interacción social de los estudiantes durante el mismo. En este estudio, enviaron encuestas a los estudiantes para analizar a aquellos que se inscribieron con amigos, y los resultados sugirieron que inscribirse en un MOOC con compañeros se correlacionaba positivamente con la tasa de finalización del curso, el nivel de rendimiento y el uso del foro de discusión. Demostraron que había un efecto positivo en el rendimiento académico de los estudiantes y una mayor interacción en línea cuando los estudiantes se inscribían con amigos o compañeros. [75] investigó los beneficios de las colaboraciones en los MOOC a través de un estudio de caso de aula invertida. Sus resultados sugirieron que los estudiantes en los MOOCs prefieren estudiar en grupos, y que la facilitación social dentro de los grupos de estudio puede hacer que el aprendizaje de conceptos difíciles sea una experiencia más agradable. Los estudiantes informaron de una alta satisfacción general con este enfoque de aprendizaje en grupo de estudio y la investigación reveló que a los estudiantes les gustaba estar en sintonía con el grupo mientras veían los vídeos del MOOC y completaban las evaluaciones. Sin embargo, ninguno de estos dos estudios analizó los comportamientos reales que estos estudiantes realizaron en las plataformas MOOC mientras colaboraban juntos.

La colaboración en los foros de discusión de los MOOC también ha sido un tema común en la literatura [31, 40]. Por ejemplo, [31] utilizó métodos de análisis de aprendizaje para recuperar y analizar los datos de la interacción de los estudiantes con los foros del curso. Los autores mostraron que el 20 % de los estudiantes colaboraban en los foros a lo largo del curso y eran responsables del 50 % del total de los mensajes. Del mismo modo, [40] presentó un estudio de los foros de discusión de los MOOC con el objetivo de extraer automáticamente la estructura de los posts de las discusiones para entender cómo los estudiantes colaboran entre sí.

La mayoría de los estudios sobre la colaboración en los MOOCs exploran cómo los estudiantes interactúan a través de una herramienta de colaboración o qué beneficios se obtienen de estas colaboraciones. Sin embargo, nuestro enfoque es muy diferente al de estos estudios, ya que utilizamos un algoritmo basado en datos para detectar y caracterizar las cuentas de los estudiantes que colaboran cuando no hay un estímulo específico para colaborar o herramientas adicionales para hacerlo. Pretendemos saber cómo colaboran los estudiantes y si estas colaboraciones están orientadas al aprendizaje o a la obtención de un certificado sin esfuerzo; hasta ahora no se ha informado de enfoques como éste en la literatura.

2.4. Aplicaciones y herramientas de visualización

Hemos revisado tanto los MOOC como los entornos tradicionales en línea que utilizan cuadros de mando de análisis del aprendizaje. Sin embargo, no encontramos ninguno que detectara el comportamiento deshonesto de los estudiantes en los cursos en línea. La mayoría de los cuadros de mando tienen como objetivo supervisar la actividad de los estudiantes con la plataforma.

En este contexto se enmarca MoocViz [95], una herramienta que permite una visión de la información para el análisis del comportamiento de los estudiantes en el aprendizaje MOOC. Se basa en datos brutos obtenidos de cursos ya desarrollados en plataformas como Coursera y edX. La herramienta permite el análisis del flujo de clics en los vídeos del curso, el análisis de los foros, entre otros. Los resultados se clasifican utilizando diferentes colores, gráficos circulares de diferentes tamaños y líneas de transición. Sin embargo, trabaja con datos de clickstream de foros y vídeos como elementos aislados, y no permite visualizar el comportamiento de los estudiantes en el ámbito evaluativo y la deshonestidad académica.

En cuanto a la visualización de datos para caracterizar el comportamiento de aprendizaje de los estudiantes en un MOOC, hemos encontrado trabajos que se basan en el foro del curso, ya que es uno de los principales componentes de la interacción social. Message Lens es uno de los trabajos que permite un análisis visual de los datos del foro, como conocer el nivel de similitud y prevalencia de un tema, cuáles son las palabras más utilizadas, tener una representación gráfica de la red de interacción entre los alumnos que se va formando a medida que interactúan con el foro. Propusieron un modelo con vistas grupales, globales e individuales a partir del uso de métricas de comportamiento de los temas de discusión, la actitud de los estudiantes y la comunicación entre ellos [133]. Con un enfoque similar, VisMooc estudia el número de hilos generados por una pregunta o respuesta en un foro [113]. Iforum, por su parte, desarrolla un sistema de análisis visual que permite explorar los datos del foro desde tres perspectivas de interacción: los mensajes, la interacción de los usuarios y los hilos de discusión. También promueve una herramienta llamada “Thread River” que puede ilustrar la información de largos hilos de conversación [45].

Por otro lado, en cuanto al análisis del abandono, en [71] mostraron una forma de representar gráficamente la clasificación de los alumnos en dos categorías diferenciadas por colores, en las que los alumnos con mayor probabilidad de éxito en el curso se muestran en naranja y los de menor probabilidad en azul, basándose en una estrategia de machine learning y tomando como referencia los datos recogidos en el foro de un curso MOOC

de Iversity.

En el caso de T-Cal, se muestra la hora en la que un alumno se conectó, la hora de entrada y la hora de salida, los gráficos de los alumnos que se conectaron al mismo tiempo y el tiempo que permanecieron en la plataforma. Toda esta información para la clasificación se basa en los foros de discusión [44].

Existen varios cuadros de mando que sirven de apoyo al profesor y a los alumnos en cuanto al seguimiento de las actividades que los alumnos realizan a través de la plataforma de aprendizaje. Sin embargo, ninguno de ellos está orientado a la deshonestidad académica, creemos que esta es la primera herramienta que detecta el comportamiento inadecuado de los estudiantes en los SPOC y se pone a disposición de los instructores. Aunque la herramienta ha sido desarrollada para Open edX, también podría extenderse a otras plataformas de aprendizaje en línea fácilmente.

Capítulo 3

Herramientas y Métodos

Índice

4.1. Descripción del contexto y del caso de estudio	54
4.2. Estudiantes identificados como sospechosos de deshonestidad académica	54
4.3. Análisis del comportamiento y el compromiso de los estudiantes .	57

3.1. Plataforma MOOC

Esta sección presenta una visión general de la plataforma de aprendizaje que se utilizó para el desarrollo de los casos de estudio. En la Subsección 3.1.1 presentamos un resumen del surgimiento de la plataforma y el tipo de cursos que se imparten en ella, y en la Subsección 3.1.2 se presentan las características de la plataforma por las cuales se definen las interacciones de los estudiantes y por tanto los indicadores seleccionados para la identificación de estudiantes con sospecha de fraude.

3.1.1. Plataforma Selene

El entorno virtual de aprendizaje utilizado es Selene, una instancia de Open edX que fue implementada en 2016 por la Universidad del Cauca. Esta iniciativa fue apoyada por la universidad con el objetivo de promover la creación y despliegue de cursos MOOC para la adquisición y desarrollo de habilidades profesionales, conocimientos y nuevas competencias en la universidad.

Tabla 3.1: Cantidad de usuarios registrados en la plataforma Selene

Course	2016-I	2016-II	2017-I	2017-II	2018-I	2018-II	2019-I	2019-II	2020-0	2020-I	Accounts
Introduction to entrepreneurship with Lean Startup				265	317	270	100	190	50	136	1328
Comprehension of Argumentative Texts	105	110	109	103	97	102	98	101		83	825
Introductory Drone Course			133	101	106						340
Editing scientific and literary texts with LaTeX		102	99	104	100	120	92	99	46	82	844
Folk dance as cultural heritage						172	112	108		104	496
Nursing Programme Educational Project						413		335			748
Fundamentals of Technology for Businesses in Tourism							34			46	80
ICT for Mobile Health Elective							24				24
Emphasis IV - Telematics						13	12	18	6	20	69
Introduction to the Ophidic Accident						6		1240			1246
Moving towards community-based rural tourism.						2				859	861
Introduction to the Internet of Things IoT.								42			42
Daily Astronomy	433	428	517								1378
Total Enrolled	538	640	858	573	620	1098	472	2133	102	1330	8281

Los primeros cursos ofrecidos en la plataforma fueron SPOCs y MPOCs con reconocimiento académico. El objetivo era incorporar los MOOCs a los programas de formación de la universidad y para ello se empezaron a ofrecer cursos privados, que permitían controlar el acceso de los estudiantes a los cursos y tener un menor número de participantes; aun así, el número de estudiantes por curso es mayor que en los cursos universitarios convencionales. La principal característica de estos cursos es que están reconocidos académicamente y por lo tanto, tienen créditos que son válidos en los programas de formación profesional de la universidad. La mayoría de los cursos que se han ofrecido en Selene son cursos de Formación Social y Humana Integral (FISH). Son cursos electivos y transversales a toda la universidad; cualquier estudiante de la universidad puede participar en ellos. También se han desarrollado varios MOOCs que han sido apoyados y financiados por proyectos internacionales y nacionales como MOOC-Maker, MOOC-MenTES y MOOC-Unicauca. A continuación se presentan algunas estadísticas de los cursos ofrecidos (Tabla 3.1). Los cursos están reconocidos académicamente y son válidos para dos créditos dentro de los programas de formación de grado. Además, el acceso a los cursos está controlado por el administrador de la plataforma, por lo que la posibilidad de crear cuentas adicionales y realizar CAMEO no es factible.

Desde su inicio, estos cursos han sido considerados internamente como SPOCs ya que son cursos privados sólo disponibles para los estudiantes de Selene Unicauca. Todos estos cursos han tenido no menos de 100 alumnos inscritos, pero en algunos periodos algunos de los cursos han tenido 200 y hasta 400 participantes. En cualquier caso, el número de alumnos es siempre más de cuatro veces superior al de los cursos presenciales de la Universidad de Cauca. Los hemos clasificado como SPOC porque utilizan tecnologías MOOC (incluyendo Open edX), y son privados pero no masivos, ya que los instructores han conseguido hasta ahora dar soporte a cada uno de los alumnos del curso.

3.1.2. Características de Open edX

Como se mencionó anteriormente, la plataforma Selene es una instancia de Open edX y por tanto tienen las mismas características.

Open edX es una plataforma de código abierto que está disponible gratuitamente para la comunidad en general. Cualquier institución puede descargarla y ejecutar su propia instancia, permitiendo a los educadores ampliar la plataforma para construir herramientas de aprendizaje que satisfagan precisamente sus necesidades y los desarrolladores aportar nuevas funciones a la plataforma.

Open edX cuenta con los siguientes componentes:

- LMS
- CMS
- Módulo XBlock
- Open edX Insights

En el presente trabajo se utilizaron el LMS y el CMS, los cuales se describen a continuación.

3.1.2.1. Sistema de Gestión de Aprendizaje (LMS, Learning Management System)

El LMS de Open edX es la herramienta que los estudiantes y docentes usan para visualizar el contenido de los cursos, permite acceder a los videos, libros de texto y problemas organizados bajo una estructura que los presenta en secciones, subsecciones y unidades. En la Figura 3.1 se muestra como el contenido del curso se despliega en la pestaña curso del LMS. Es así como los estudiantes pueden acceder a los contenidos y su interacción con este menú es una de las opciones a través de las cuales los estudiantes navegan en los cursos. Además, la Figura 3.2 muestra la visualización de los contenidos en plataforma. En la figura se muestra una barra de navegación con la que los estudiantes pueden avanzar, retroceder o buscar unidades específicas. Esta es otra manera por la cual los estudiantes pueden navegar en el curso.

Por otra parte, el LMS incluye un espacio dirigido específicamente para los docentes, con opciones para producir informes y administrar el curso a medida que se ejecuta. En la Figura 3.3 se muestra la interfaz que es visualizada por el docente.

En la vista "Instructor" se presenta información de los inscritos en el curso, inscribir o dar de baja a estudiantes, generar grupos de trabajo para mostrar contenidos o administrar

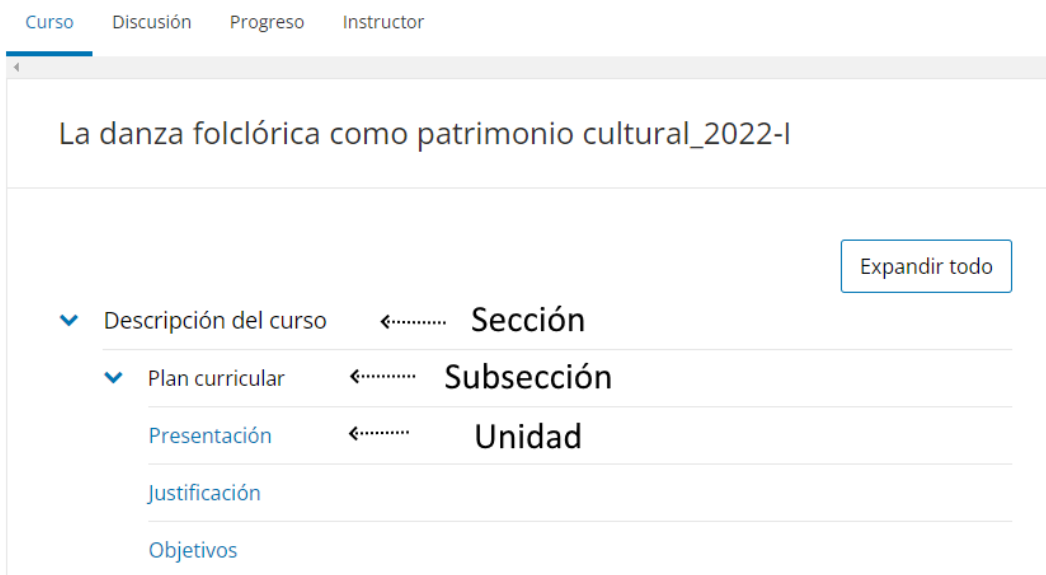


Figura 3.1: Vista del contenido del curso en Open edX.



Figura 3.2: Vista de los contenidos en Open edX.

los foros, generar y descargar reporte sobre las evaluaciones del curso, etc. Además, también es posible dar credenciales especiales a los usuarios que ayudan con el desarrollo del curso, como moderadores de foros, usuarios de prueba, profesores asistentes, entre otros.

3.1.2.2. Sistema de Gestión de Contenidos (CMS, Content Management System)

El CMS de Open edX es conocido como Studio y es la herramienta que se utiliza para crear cursos. En Studio se puede gestionar la estructura del curso y agregar el contenido, donde son incluidos los exámenes, vídeos, lecturas y otros recursos para los estudiantes. Para que un usuario pueda crear cursos es necesario que la cuenta cuente con permisos especiales, esto solo lo puede hacer el administrador de la instancia Open edX instalada.

FISH: Danza_Folclorica
La danza folclórica como patrimonio cultural_2022-I

Curso Discusión Progreso **Instructor**

PANEL DE CONTROL DEL INSTRUCTOR

Información del curso Inscripciones y roles Cohortes Administrador de Estudiantes Descarga de Datos Analíticas

Información del curso

Información de inscripción

Verificado	0
Tomar como asistente	99
Honor	0
Profesional	0
Total	99

Número total de inscritos (incluye administradores, personal de apoyo y estudiantes) por ruta

Figura 3.3: Vista de pestaña “Instructor” en Open edX.

3.1.2.2.1. Estructura del curso

Studio permite la construcción de la estructura de un curso de una manera jerárquica, teniendo en cuenta las Secciones que corresponden a la categoría de mayor nivel, luego están las subsecciones que puede representar un tema o cualquier otra organización de contenidos que se desee y éstas a su vez, se dividen en unidades. En particular, todos los contenidos se crean dentro de las unidades y son llamados componentes. En la Figura 3.4 se muestra la construcción de la estructura de un curso, con las secciones, subsecciones y unidades vista en Studio edX.

3.1.2.2.2. Contenidos

Son cuatro tipos de componentes que se pueden crear como contenido en una unidad, foros, ejercicios, videos y componentes Html¹. En la Figura 3.5 se muestra la interfaz para adicionar componentes a una unidad.

Componentes de Discusión o Foro: Se puede añadir un componente de Discusión a la Unidad con el fin de publicar preguntas relacionadas con la unidad y dar a los estudiantes la oportunidad de responder e interactuar entre ellos y el profesor. Estos estarán disponibles en la estructura del curso, pero además se pueden acceder a ellos en la pestaña

¹HTML, sigla en inglés de HyperText Markup Language (lenguaje de marcas de hipertexto), hace referencia al lenguaje de marcado para la elaboración de páginas web.

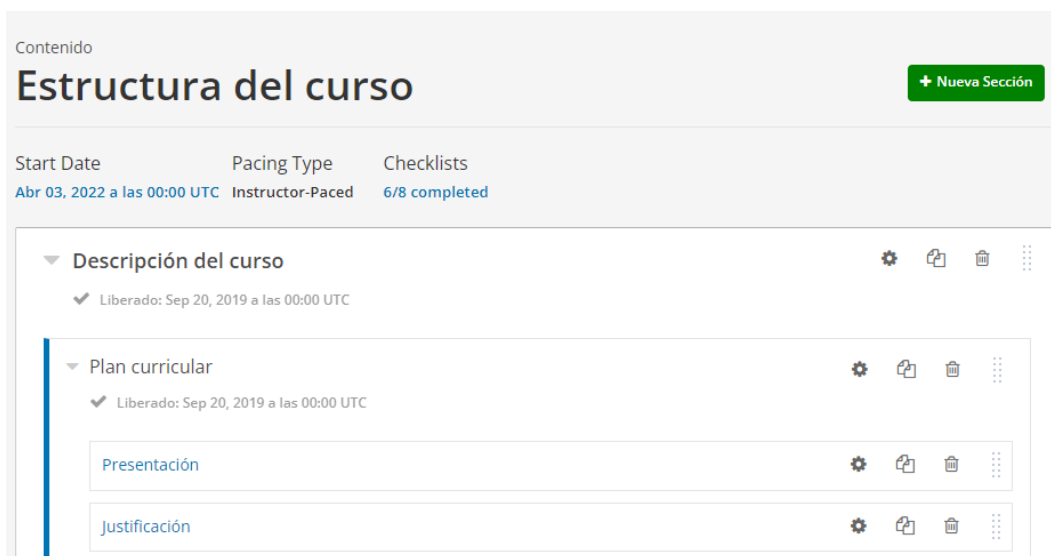


Figura 3.4: Vista de la estructura de los cursos en Studio

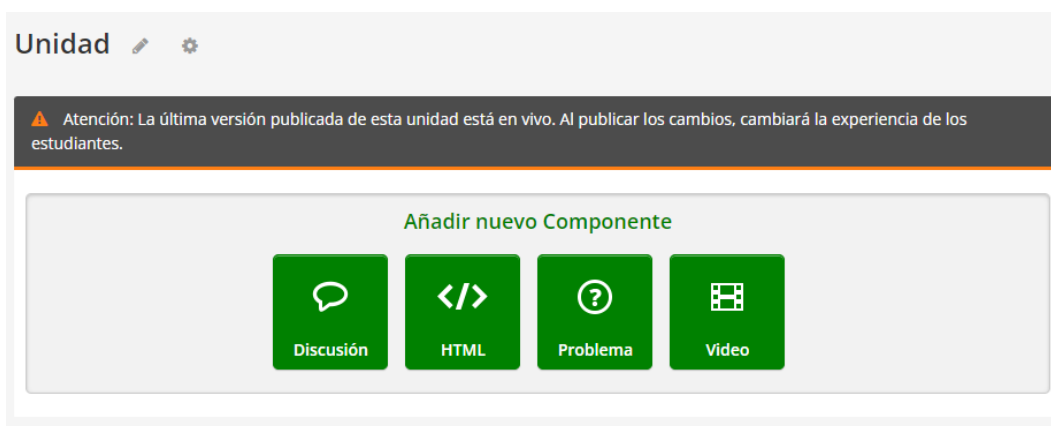


Figura 3.5: Vista de la estructura de los cursos en Studio

discusión en el LMS.

Componentes HTML: En un componente HTML se puede añadir texto, listas, enlaces, imágenes, presentaciones, pdf, etc., en código html.

Componentes de Ejercicios: El componente de Ejercicio permite añadir ejercicios interactivos, de calificación automática al contenido del curso. Se puede crear distintos tipos de ejercicios con Studio. Todos los ejercicios reciben una puntuación, pero, por defecto, los ejercicios no cuentan en la nota del estudiante si esto no es configurado en la subsección que contiene el ejercicio.

Los ejercicios más comunes que están disponibles en Studio son:

- Checkbox.
- Lista desplegable.
- Opción múltiple.
- Entrada numérica.
- Entrada de texto corta.

Componentes de Video: La plataforma permite agregar videos como componentes de una unidad a través de Youtube. Todos los vídeos del curso deberán ser publicados allí, pero dado que YouTube no está disponible en todos los países del mundo, es recomendable además publicar copias de los vídeos en terceras empresas como Amazon S3. Cuando un estudiante visualiza un vídeo en el curso, se buscará primero el mismo en YouTube. Si no estuviera disponible, se mostrará automáticamente el vídeo desde la localización alternativa. El estudiante también puede pulsar sobre un enlace para descargar el vídeo desde dicha localización alternativa. Es así como los estudiantes también tienen interacciones con contenidos como reproducir un video, pausar o finalizar de ver.

3.2. Casos de estudio

Los cursos utilizados en este estudio se denominan: “Introducción al emprendimiento Lean Startup” (Curso Lean Startup), “Comprensión de textos argumentativos” (Curso de textos), “La danza folclórica como patrimonio cultural” (Curso de danza), e “Introducción a la edición de textos científicos y literarios con LaTeX” (LaTeX) de la Universidad del Cauca ofrecidos a través de la plataforma Selene para obtener créditos. Además, utilizan contenidos, programación y horarios similares a los que se encuentran frecuentemente en los MOOCs.

Estos cursos se han organizado a través de unidades temáticas divididas en módulos. La evaluación se lleva a cabo según las directrices del reglamento institucional, con un examen por módulo del curso. Las evaluaciones se realizan a través de exámenes en línea con preguntas de selección múltiple con única respuesta correcta. Cada evaluación consta de 12 a 40 preguntas con la posibilidad de realizar el examen en dos fechas diferentes. Cada examen está disponible para los estudiantes sólo durante una ventana de tiempo de 60 minutos para minimizar las posibilidades de comportamientos deshonestos o de compartir las respuestas. Además, hay cuestionarios de prueba que no cuentan para la

calificación del curso. Los instructores organizan un examen de práctica inicial para que los estudiantes se familiarizaran con la mecánica de las evaluaciones del curso.

3.2.1. Curso Virtual de Astronomía Cotidiana

El curso virtual de Astronomía cotidiana, se propuso por el grupo AIDA (Agrupación para el Impulso y Desarrollo de la Astronomía) de la Universidad del Cauca para ser desarrollado como una materia electiva FISH (Componente de Formación Social Integral y Humana) de la universidad y que tiene como propósito introducir a los estudiantes de la universidad en los principios, métodos, y herramientas de la educación en línea para el estudio de conceptos básicos de la Astronomía y su impacto con la sociedad contemporánea.

Debido a los antecedentes existentes en la Universidad del Cauca del ofrecimiento exitoso de cursos no presenciales (virtuales) en Astronomía, se plantea el desarrollo de este curso de manera virtual y para su implementación se opta por utilizar la plataforma Selene, este curso se constituye en una experiencia académica propicia para la utilización de los principios, los métodos y las herramientas de la educación en línea aplicadas al desarrollo de actividades presenciales así como de la utilización de otros servicios de las Tecnologías de Información y Comunicación (vídeo conferencia, web 2.0, simulación de software), presentándose como un candidato idóneo para ser el piloto del presente trabajo.

El curso fue diseñado y ejecutado por Mario Solarte Sarasty, profesor titular de planta de la Universidad del Cauca e integrante de grupo AIDA. Tuvo una duración de 18 semanas, comenzando el 15 de febrero y terminando el 10 de junio de 2016, contando en un comienzo con la inscripción de 403 estudiantes de diferentes programas de la Universidad del Cauca.

La temática del curso se dividió en tres temas principales:

- Unidad temática I: Fundamentos de Astronomía
- Unidad temática II: Desarrollo histórico de la Astronomía
- Unidad temática III: Astronomía práctica

Y los contenidos utilizados en su mayoría fueron vídeos realizados por el docente a cargo, pero también se utilizaron diapositivas, animaciones, simulaciones y páginas web. La nota de los estudiantes se obtuvo teniendo en cuenta los temas principales y se distribuyó de siguiente manera:

- Evaluación de la Unidad temática I: un primer parcial correspondiente al 35 % de la nota final.
- Evaluación de la Unidad temática II: un primer parcial correspondiente al 35 % de la nota final.
- Evaluación de la Unidad temática III: un primer parcial correspondiente al 30 % de la nota final.

La calificación de cada una de esas notas corresponde a la sumativa de las actividades semanales de aprendizaje definidas y que consistieron de: la participación en foros de discusión, elaboración de mapas conceptuales, la presentación de pruebas en línea y el desarrollo de talleres.

La plataforma tecnológica permitió el uso de servicios como preguntas y respuestas en los foros, para facilitar la interacción entre los distintos actores del proceso. Además, ocasionalmente se realizaron actividades presenciales por miembros del grupo AIDA tales como: conferencias, vídeo proyecciones, salidas de observación, y participación en proyectos relacionados con los objetivos del curso. Este fue el primer curso que se ofreció a través de la plataforma de aprendizaje Selene.

3.2.2. Introducción al emprendimiento con Lean Startup

El curso Introducción al emprendimiento con Lean Startup tiene como propósito fundamental, introducir a los estudiantes en el conocimiento y uso de una de las metodologías para el desarrollo de emprendimientos que mayor éxito ha tenido en los últimos tiempos: el Lean Startup.

La mayoría de sus contenidos son vídeos y lecturas, y se dividen en dos unidades:

- Unidad Temática I: Emprendimiento de alto impacto innovador
- Unidad Temática II: Emprendimiento con Lean Startup
 - Metodología desarrollo de clientes
 - Modelo de Negocio
 - Validación de Problemas
 - Producto Mínimo Viable

La evaluación se realiza mediante cuestionarios de selección múltiple aplicados por cada subsección del contenido. Teniendo en total 6 evaluaciones. Una por cada temática y un examen final. La nota final es el promedio del resultado de todas las evaluaciones.

3.2.3. Comprensión de Textos Argumentativos

el curso virtual “Comprensión de Textos Argumentativos” se enmarca en la línea temática Sociedad, arte, lenguaje y cultura, ya que esta línea temática busca desarrollar, en los estudiantes, conocimientos y estimular aptitudes relacionadas con diferentes expresiones artísticas, el lenguaje y la cultura, en sus complejas implicaciones sociales. Ofrece espacios para mejorar en los estudiantes las competencias comunicativas y textuales, el conocimiento y respeto de la diversidad lingüística y cultural y un acercamiento crítico a los procesos y prácticas de producción y recreación cultural.

Este curso tiene dos notas divididas en un primer 70 % y un 30 % final para obtener la aprobación del curso. Los contenidos son los siguientes:

- Unidad 1: Inferencias Enunciativas.
- Unidad 2: Inferencias Macroestructurales y Superestructurales.
- Unidad 3: Inferencias Argumentativas.

Para la nota del 70 % del curso se evalúa la Unidad 1 y Unidad 2, programando 3 exámenes por unidad para un total de 6 notas. Para la nota del 30 % final realiza un solo examen. En total 7 evaluaciones para todo el curso mediante exámenes de selección múltiple.

3.2.4. La danza folclórica como patrimonio cultural

El curso FISH electivo en modalidad virtual “La Danza folclórica como patrimonio cultural” se enmarca en la línea temática de “sociedad, arte, lenguaje y cultura” y pretende contribuir a la formación integral del estudiante de la Universidad del Cauca desde el reconocimiento y apropiación de las raíces culturales representadas en las danzas folclóricas del Departamento del Cauca.

Para cumplir tal propósito, este curso está dividido en tres unidades temáticas. En la primera se abordarán temas relacionados con la conceptualización, los niveles, los tipos y la valoración del patrimonio. Esta primera unidad busca fortalecer concepciones locales en torno a la importancia del patrimonio inmaterial de las comunidades locales a las que pertenecen cada uno de los estudiantes del curso.

En la segunda unidad se considerarán aspectos generales sobre la danza y el folclor, así como también haciendo énfasis en las manifestaciones culturales y corpóreas, elementos fundamentales en el tema de la danza. Conocer y aprender acerca del Departamento del Cauca y de las manifestaciones culturales como la danza ayuda a conservar

las raíces culturales y contribuye significativamente al desarrollo personal integrando la comunicación de sentimientos y emociones que se dan en la danza en un espacio virtual. Con los temas abordados en esta unidad se pretende sensibilizar sobre la importancia de la danza como manifestación cultural y elemento fundamental patrimonio inmaterial de las comunidades locales.

La tercera y última unidad, puntualiza sobre las tres danzas más representativas del Departamento del Cauca, agrupadas en tres zonas, norte, centro y sur; siendo estas la Juga, el Sotareño y el Bambuco Patiano respectivamente. De cada una de ellas se tratará la historia, el vestuario, la coreografía, la música y la vivencia. Con esta última unidad se pretende que el estudiante pueda reconocer los bailes típicos del Departamento de Cauca como una manifestación cultural de sus territorios.

La evaluación del curso se hace a través de cuestionarios de selección múltiple y los porcentajes para la nota final son un primer 70 % y un 30 % final.

La nota del 70 % esta compuesto por la suma de la unidad 1 y la unidad 2. La nota del 30 % esta compuesta por una evaluación final. En total se realizan 3 exámenes en todo el curso, un examen por cada una de las unidades.

3.2.5. Introducción a la edición de texto científicos y literarios con LaTeX

El curso de Introducción a la edición de texto científicos y literarios con LaTeX está diseñado para aprender conceptos generales sobre el uso de LaTeX. Un sistema de composición de textos, orientado a la creación de documentos escritos que presenten una alta calidad tipográfica.

El curso se estructura en 15 semanas y los contenidos son:

- Unidad I: Introducción a LaTeX
 - Semana 1: Introducción al curso en línea
 - Semana 2: Introducción a LaTeX
 - Semana 3: Tipografía y formatos de documento
 - Semana 4: Instalación de LaTeX
 - Semana 5: Motor y Recursos de LaTeX
- Unidad II: Estructura y formato de un documento
 - Semana 6: Estructura de un documento

- Semana 7: Encabezado y cuerpo de un documento
- Semana 8: Listas, figuras y tablas
- Semana 9: Fórmulas y expresiones numéricas
- Semana 10: Formatos de referencias bibliográficas
- Unidad III: Referencias bibliográficas y plantillas
 - Semana 11: Gestión de referencias bibliográficas
 - Semana 12: Plantillas
 - Semanas 13 y 14: Proyecto final
 - Semana 15: Valoración del proyecto final

Se realiza una evaluación por semana mediante cuestionarios de selección múltiple. Obteniendo el primer 35 % a partir de las evaluaciones de la Unidad 1, el segundo 35 % de la evaluaciones de la Unidad 3 y el 30 % final de las evaluaciones de la Unidad 3.

3.3. Indicadores seleccionados

3.3.1. Interacciones con la plataforma

En esta sección se describen indicadores asociados a las interacciones de los estudiantes con la plataforma de aprendizaje. La mayoría de estos indicadores, se centran en la cantidad de interacciones que realiza cada estudiante con diferentes elementos de la plataforma y también cantidad de interacciones con los recursos educativos en los cursos. A continuación, se presenta una clasificación para las actividades de aprendizaje encontradas en el ámbito MOOC y que es tomada en cuenta para la presentación de los indicadores encontrados y que más adelante algunos de ellos son utilizados para la identificación de estudiantes con sospecha de deshonestidad académica:

- Interacción estudiante – plataforma de aprendizaje, en este contexto se tienen en cuenta actividades como el ingreso del estudiante a la plataforma, ingreso al curso, ingreso al contenido del curso, etc. [39, 54].
- Interacción estudiante - evaluaciones, es decir, tareas como: realizar exámenes, tareas, talleres, etc. [7].
- Interacción Estudiante – Contenidos; actividades de revisión de presentaciones, vídeos, archivos de audio, textos y contenidos complementarios [27, 52, 69].

- Interacción Estudiante - Herramientas de Comunicación, incluyen la participación en los foros, chats y redes sociales [29].

3.3.1.1. Indicadores según la Interacción Estudiante - Plataforma de aprendizaje

Los datos de las interacciones estudiante - plataforma de aprendizaje, en este contexto se presentan como el número de ingresos al curso o el tiempo que el estudiante permanece en la misma e incluso los inconvenientes presentados con contenidos, tareas, herramientas, entre otros. Dichos datos pueden permitir al tutor actuar a tiempo y brindar al estudiante un acompañamiento para evitar su deserción. Además, son de gran utilidad en el diseño y ejecución de los cursos [39, 54, 136], ayudando en gran medida a reducir las altas tasas de abandono que se presentan en los MOOC.

- **Ingresos a un Curso por estudiante:** Número de veces que un estudiante ha ingresado en un curso, mensual, semanal o diario.
- **Total de ingresos de un Curso:** Número total de veces que los estudiantes han ingresado en un curso, mensual, semanal o diario.
- **Ingresos en Secciones y Subsecciones por estudiante:** Número de veces que un estudiante ha ingresado en una sección y subsecciones.
- **Ingresos en Secciones y Subsecciones por curso:** Número de veces las sección y subsecciones han sido visitadas por los estudiantes.
- **Historial de Ingresos:** Registro de hora y fecha del ingreso de cada uno de los estudiantes en un curso, sección y subsección.
- **Tiempo de permanencia:** El tiempo que el estudiante ha tomado para cada ingreso.

3.3.1.2. Indicadores según la Interacción Estudiante - Evaluación

La información que se puede obtener de las interacciones de los estudiantes con las evaluaciones es muy importante. Además de ver el rendimiento del estudiante en términos de los resultados de las evaluaciones, es posible obtener cuál es el comportamiento del estudiante, saber el interés en el curso, conocer la forma de abordar las tareas a realizar, etc. Desde la bibliografía, se han identificado indicadores que se proponen como parte del seguimiento a los estudiantes y que en algunos casos ya se monitorizan [7, 99].

- **Resultados de Evaluaciones:** Resultados de evaluaciones como tareas, talleres, laboratorios o exámenes, obtenidos por los estudiantes. En la mayoría de los MOOC, al trabajar con evaluaciones automáticas, se tiene la única posibilidad de una evaluación sumativa. Sin embargo en un proceso educativo, se debería tener en cuenta la calidad, la efectividad, la forma, la cantidad, etc., del trabajo realizado por el estudiante.
- **Número de Evaluaciones:** El número de tareas, talleres, laboratorios o exámenes realizadas por los estudiantes. Este indicador ayuda a entender cuál es el avance o progreso del estudiante en cuanto a actividades de evaluación trabajadas.
- **Intentos por estudiante:** En algunos MOOC dependiendo del diseño y la metodología empleada por el docente o tutor, se permite al estudiante realizar una misma actividad evaluativa varias veces. Este indicador hace referencia al número de veces que un estudiante repite una evaluación.
- **Historial de presentación:** Registro de hora y fecha en que un estudiante contestó a una evaluación.
- **Frecuencia de presentación:** Frecuencia con la que los estudiantes realizan las evaluaciones o exámenes, se puede medir por estudiante o por curso.
- **Tiempo de desarrollo:** El tiempo que cada estudiante toma en desarrollar una actividad de evaluación.

3.3.1.3. Indicadores según la Indicadores Interacción Estudiante – Contenidos

Con miles de estudiantes viendo vídeos de MOOC se producen enormes cantidades de datos de navegación. Tales datos a gran escala ofrecen una gran oportunidad para los instructores y analistas educativos para profundizar en los comportamientos de aprendizaje en línea en una escala sin precedentes [112]. A continuación se describen los indicadores encontrado:

- **Vistos de un vídeo:** Número total de estudiantes que han visto un vídeo específico.
- **Repeticiones de un vídeo:** Número de veces que los estudiantes han visto un vídeo específico.
- **Repeticiones de un vídeo por estudiante:** Número de veces que cada estudiante ha visto un vídeo específico.

- **Vídeos vistos por estudiante:** Número de vídeos que un estudiante ha visto.
- **Vídeos vistos por curso:** Número de vídeos que los estudiantes han visto.
- **Progreso:** Número de vídeos vistos contrastados con el total de vídeos existentes, por cada estudiante y por el curso en general.

3.3.1.4. Indicadores según la Interacción Estudiante - Herramientas de Comunicación

Las herramientas sociales, tales como foros o redes sociales, se emplean normalmente para conectar a los participantes MOOC [6], los registros de las interacciones de los estudiantes a través de estas herramientas de comunicación podrían ofrecer datos importantes que ayuden al diseño y el ofrecimiento de las actividades de los cursos y pueden ser de gran importancia en el seguimiento de los estudiantes. A continuación se presentan los indicadores asociados a las interacciones con foros:

- **Foros creados:** Número de foros que un estudiante ha creado.
- **Comentarios creados:** Número de comentarios hechos por un estudiante en los foros de otros estudiantes.
- **Respuestas creadas por estudiante:** Número de respuestas que un estudiante ha dado a las preguntas hechas en el foro del curso.

3.3.2. Extracción de datos para medición de indicadores de interacción

En esta subsección se presenta el diseño e implementación de un mecanismo para la medición de indicadores de interacción de los estudiantes con la plataforma de aprendizaje. Este diseño se realizó teniendo en cuenta la estructura y funcionamiento de Open edX. En la Figura 3.6 se presenta la arquitectura general del mecanismo construido.

A continuación se describe cada uno de los componentes de la arquitectura y su relación con los demás elementos.

3.3.2.1. Tracking.log

La plataforma Selene se despliega sobre un equipo servidor con una versión de Ubuntu server 14.04 y es una instancia de Open edX. En la plataforma se genera un archivo de

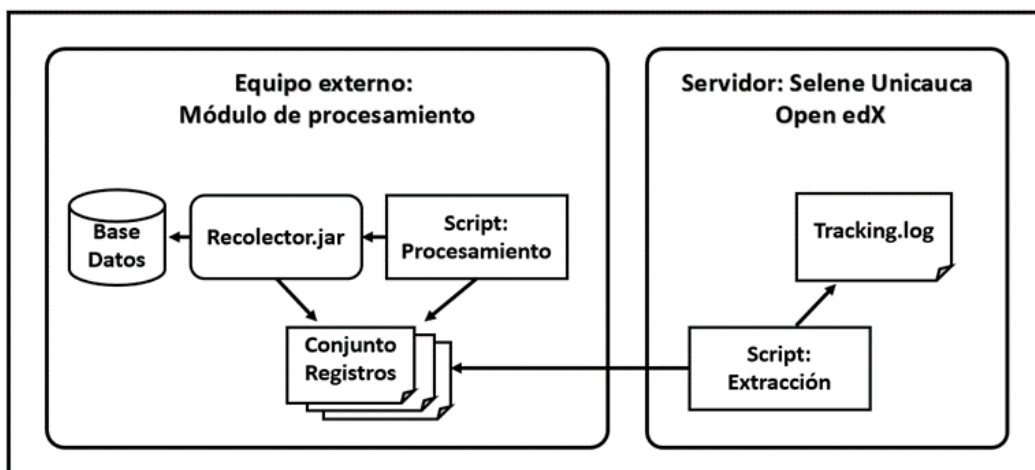


Figura 3.6: Arquitectura general mecanismo propuesto

registro llamado tracking.log. En este archivo se registran todas las interacciones de los estudiantes con el entorno de aprendizaje Selene, tales como: inicio de sesión, ingreso en los cursos, navegación en contenidos, presentación de foros e información de la interacción con las actividades evaluativas. La ubicación del archivo tracking.log dentro de la instancia es: /edx/var/log/tracking/. Aquí periódicamente dependiendo de la cantidad de datos de interacción generados se crea un nuevo archivo de registro y se comprime y almacena el anterior dejando registro de las interacciones en plataforma desde que ésta es dada de alta.

3.3.2.2. Script: Extracción

Los scripts son un conjunto de instrucciones generalmente almacenadas en un archivo de texto que deben ser interpretados línea a línea para su ejecución. El script de extracción lo que contiene son los comandos utilizados para realizar una copia de toda la carpeta que contiene los registros tracking.log de la plataforma y una línea de sincronización de dicha carpeta en un equipo externo. De esta manera todos los registros pueden ser extraídos y puestos a disposición para ser procesados sin afectar el funcionamiento de la plataforma de aprendizaje. Este script es ejecutado cada 5 minutos. La Figura 3.7 muestra el script de extracción.

3.3.2.3. Script: Procesamiento

Una vez con todos los registros disponibles en el equipo externo, el script de procesamiento captura el archivo tracking.log que es el archivo que tiene la información de

```
GNU nano 2.5.3      File: script extraccion.sh      Modified
cp -r /edx/var/log/tracking/ /home/selene/respaldo/

rsync -avzhie "ssh -p 22" usuario@192.168.1.103:/home/recolector/Registros
/home/selene/respaldo
```

Figura 3.7: Script de extracción (Captura de pantalla)

las interacciones de los estudiantes, realiza una copia del mismo, lo ubica en una dirección específica en donde el Recolector.jar puede procesarlo y repite esta operación periódicamente cada 5 minutos a través del programa Cron² presente en Ubuntu. Así, los datos procesados de las interacciones pueden ser utilizados para realizar seguimiento a los estudiantes en todo momento. Para obtener los datos de eventos pasados, el script fue modificado a fin de que descomprimiera uno por uno los archivos de registro y ordenara la ejecución del Recolector.jar para cada registro presente en la carpeta sincronizada, así se logró crear un Data set con datos de interacción de los estudiantes desde el momento en que se creó la plataforma. En la Figura 3.8 se muestra el script de procesamiento el cual es ejecutado cada 5 minutos.

```
GNU nano 2.5.3      File: script procesador.sh      Modified
cd /home/recolector/tracking/
rm -rf tracking.log

cp /home/recolector/Registros/tracking.log /home/recolector/tracking
java -jar Recolector.jar

echo "tracking procesado"
```

Figura 3.8: Script de procesamiento de registros (Captura de pantalla)

3.3.2.4. Recolector.jar

Es un ejecutable creado en lenguaje Java, que al ejecutarse busca el archivo tracking.log en una ubicación específica, lo lee tomando evento por evento presente en el archivo de registro, guarda en un buffer cada evento relacionado a las actividades de interacción con la plataforma para luego procesarlos.

El procesamiento de la consiste en detectar qué actividad se está realizando para luego guardarla en una base de datos con un formato definido. El archivo tracking está escrito en Json, por lo que el recolector debe interpretar dicha información de manera adecuada. En la Figura 3.9, se muestra un ejemplo de los eventos registrados en el archivo tracking.log.

²Cron es un administrador regular de procesos en segundo plano que ejecuta procesos o scripts a intervalos regulares de tiempo (por ejemplo, cada minuto, día, semana o mes).

```
{
  "username": "nombreusuario",
  "event_source": "browser",
  "name": "play_video",
  "accept_language": "en-US,en;q=0.8,es;q=0.6",
  "time": "2017-11-03T02:18:21.315028+00:00",
  "agent": "Mozilla/5.0 (Windows NT 6.2; WOW64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/61.0.3163.100 Safari/537.36",
  "page": "http://openedx.galileo.edu/courses/course-v1:MOOC-Maker+UCN01+2017_02/courseware/5afe24c850eb4d64a1f098664e051b4e/bf20f4fc67004f50a42b44b6e1059ba1/",
  "host": "openedx.galileo.edu",
  "session": "Sa36a65aae02b6182f10cc7ffa9f8b53",
  "referer": "http://openedx.galileo.edu/courses/course-v1:MOOC-Maker+UCN01+2017_02/courseware/5afe24c850eb4d64a1f098664e051b4e/bf20f4fc67004f50a42b44b6e1059ba1/",
  "context": {
    "user_id": 4559,
    "org_id": "MOOC-Maker",
    "course_id": "course-v1:MOOC-Maker+UCN01+2017_02",
    "path": "/event/"
  },
  "ip": "186.86.29.150",
  "event": "{\"code\": \"ZoAbBwtYha4\", \"id\": \"3c7643dccb5d4baa9d4850d2b59e26b9\", \"currentTime\": 0.029999805450439454}",
  "event_type": "play_video"
}
```

Figura 3.9: Evento del archivo tracking.log

Como se observa, dentro de la línea de registro se guarda toda la información necesaria para el seguimiento a las actividades. Para el ejemplo, la pareja “*name:play_video*”, identifica que un estudiante reprodujo un vídeo dentro de la plataforma de aprendizaje, incluyendo información del id del estudiante, fecha, sección y subsección, curso, etc.

A continuación se describen los valores del campo “*name*” que ayudan a identificar las actividades realizadas por los estudiantes.

- *play_video*: Reproducción de un vídeo.
- *Stop_video*: Registra cuando un usuario llega al final del video.
- *Pause_video*: Una pausa en el vídeo.
- *Seek_video*: Búsquedas dentro del vídeo.
- *edx.forum.thread.created*: Creación de un nuevo hilo en el foro.
- *edx.forum.comment.created*: Comentario dentro de un hilo en el foro.
- *edx.forum.response.created*: Respuesta a los comentarios.
- *edx.forum.thread.voted*: Votos realizados en los foros.
- *dashboard*: Ingreso a los cursos.
- *edx.ui.lms.link_clicked*: Ingresos en los contenidos del curso.
- *problem_check*: Envío de respuestas en una evaluación.

3.3.3. Identificación de estudiantes sospechosos de deshonestidad académica

El método que implementamos contiene ideas de trabajos anteriores [99] e incorpora otras nuevas. Hemos utilizado dos criterios para identificar a los alumnos sospechosos

de fraude: los alumnos que tienen respuestas de examen muy similares y los alumnos que presentan exámenes muy cercanos en el tiempo. Para realizar el análisis basado en estos criterios, se utiliza la siguiente nomenclatura matemática: Tenemos N que denota el número total de estudiantes y M que denota el número total de variables. También tenemos N vectores, uno para cada uno de los estudiantes de manera que:

$$\vec{sp}_i = [sp_{i,1} \quad sp_{i,2} \quad \cdots \quad sp_{i,M}], i \in \{1 \cdots N\} \quad (3.1)$$

SP_i contiene todas las variables M del estudiante i en el curso. En caso de que el alumno no haya presentado ningún examen, la variable será $sp_{ij} = N.A.$.

Esto define la matriz $SP \in N^{N \times M}$ como:

$$SP = \begin{pmatrix} \vec{sp}_1 \\ \vec{sp}_2 \\ \vdots \\ \vec{sp}_N \end{pmatrix} = \begin{pmatrix} sp_{1,1} & sp_{1,2} & sp_{1,3} & \cdots & sp_{1,M} \\ sp_{2,1} & sp_{2,2} & sp_{2,3} & \cdots & sp_{2,M} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ sp_{N,1} & sp_{N,2} & sp_{N,3} & \cdots & sp_{N,M} \end{pmatrix} \quad (3.2)$$

Donde $SP_{i,j}$ representa el valor de la variable j para el alumno i . Es decir, para el método de identificación de estudiantes por similitud de respuestas, $SP_{i,j}$ representa la respuesta j presentada por un estudiante i . En el caso de los estudiantes identificados por presentar los exámenes muy cerca en el tiempo, $SP_{i,j}$ representa la marca de tiempo del examen j presentado por el estudiante i .

Entonces, definimos $DS \in R^{N \times N}$ una matriz de disimilitud como sigue:

$$DS = \begin{pmatrix} ds_{1,1} & ds_{1,2} & ds_{1,3} & \cdots & ds_{1,N} \\ ds_{2,1} & ds_{2,2} & ds_{2,3} & \cdots & ds_{2,N} \\ \vdots & \vdots & \vdots & \cdots & \vdots \\ ds_{N,1} & ds_{N,2} & ds_{N,3} & \cdots & ds_{N,N} \end{pmatrix} \quad (3.3)$$

Donde cada entrada $ds_{i,j}$ representa la diferencia entre los valores de los estudiantes i y j basada en una métrica de disimilitud. Implementamos estas matrices para las dos métricas siguientes

- Las respuestas seleccionadas para cada ítem de cada examen del curso.
- Las marcas de tiempo de presentación para cada examen en el curso.

A continuación explicamos con detalle cómo se calcula cada métrica de distancia.

3.3.3.1. Matriz de distancia basada en la comparación de las respuestas de los exámenes

Cada elemento de la matriz DS se calcula mediante una función de disimilitud $dis_{SS}(sp_i, sp_j) \in R$ que opera sobre los vectores de respuestas de los estudiantes. En este caso, la función de disimilitud se basa en el coeficiente de coincidencia simple (SMC, *Simple Matching Coefficient*).

$$dis_{SSMC}(s\vec{p}_i, s\vec{p}_j) = 1 - \frac{\text{Conteo de coincidencias entre } sp_i \text{ y } sp_j}{M} \quad (3.4)$$

De esta forma, cuando sp_i y sp_j tengan exactamente las mismas respuestas, la distancia será 0 y cuando sp_i y sp_j no coincidan en ninguna de las respuestas, la distancia será 1.

3.3.3.2. Matriz de distancia basada en los tiempos de presentación de los exámenes

En este caso, el cálculo de la matriz de distancia DS se realiza mediante la Desviación Media Absoluta (DAM, *Mean Absolute Deviation*). Donde la distancia entre dos vectores se calcula como la media de las diferencias absolutas entre ambos vectores.

$$dis_{SSMAD}(s\vec{p}_i, s\vec{p}_j) = \frac{1}{M} \sum_{k=1}^M |sp_{i,k} - sp_{j,k}| \quad (3.5)$$

De este modo, la distancia $ds_{i,j}$ entre dos estudiantes i y j es la media de las distancias temporales de cada envío. Cuanto más cerca estén los envíos de dos estudiantes en el tiempo, menor será la distancia entre ellos.

3.3.4. K-means algorithm

El algoritmo K-means, creado por MacQueen en 1967 es el algoritmo de clustering más conocido y utilizado por su eficacia y robustez. Sigue un procedimiento simple de clasificación de elementos en un determinado número K de clústeres. El nombre de K-means viene porque representa cada uno de los clusters por la media (o media ponderada) de sus puntos, es decir, por su centroide. Cada cluster por tanto es caracterizado por el

elemento central o centroide que se encuentra en el centro o el medio de los elementos que componen el cluster, así cada elemento es ubicado dependiendo de su distancia al centroide. Puede aplicarse el algoritmo K-means para problemas de agrupación por similitud que puede ayudar al investigador a una comprensión cualitativa y cuantitativa de grandes cantidades de datos de N-dimensiones [120, 134]. La implementación iterativa de K-Means busca maximizar las distancias entre clusters y minimizar la distancia total entre los miembros de un mismo grupo. En otras palabras se espera que los K grupos resultantes tengan una gran similitud dentro de cada grupo, pero poca similitud (disimilitud) entre los grupos [120]. El procedimiento del algoritmo es el siguiente [120, 134]:

- Se comienza seleccionando K objetos al azar del conjunto total y asignándolos como patrones o centroides de las K clases o grupos que se van a buscar.
- A continuación, se calculan todas las distancias euclídeas de todos los objetos restantes a todos los K centroides, y se asigna la pertenencia a cada objeto al cluster que tenga más cercano. Entonces se recalcula el centroide de cada cluster, como la media de todos los objetos que lo componen, buscando minimizar el valor de J, que es un sumatorio de todos los sumatorios de las distancias euclídeas de los objetos de cada clase al centroide de su respectiva clase. Mediante J se logra minimizar la distancia total entre los miembros de un mismo grupo. J define en (5.1). Hacer Ecuación.
- Los dos pasos anteriores se repiten sucesivamente hasta que los centros de todos los grupos permanezcan constantes. Logrando así que cada vez los elementos se ubiquen más cerca del centroide y que las similitudes de los elementos sean lo más cercanos posible.

Parte II

Identificación y caracterización de comportamientos deshonestos

Capítulo 4

Identificación de estudiantes sospechosos de deshonestidad académica

Índice

5.1. Contexto del estudio	64
5.1.1. SPOCs utilizados para el estudio	64
5.1.2. Variables utilizadas	65
5.1.3. Métricas y método de agrupamiento	66
5.1.4. Definición de colaboración en este estudio	67
5.2. Aplicación del método K-means	68
5.2.1. Caracterización de los estudiantes.	68
5.3. Detección de colaboraciones	72
5.4. Caracterización de las Colaboraciones	77
5.4.1. Asociaciones de parejas	77
5.4.2. Asociaciones de comunidades	78

Esta sección presenta los resultados del estudio, con una primera subsección centrada en la identificación de los estudiantes sospechosos de ser deshonestos según nuestro método, y la segunda subsección caracteriza el compromiso y los resultados de aprendizaje de esos estudiantes.

4.1. Descripción del contexto y del caso de estudio

Como estudio de caso hemos utilizado los siguientes dos cursos ofrecidos durante el primer semestre del 2019: “Introducción al emprendimiento con Lean Startup” (*Curso Lean Startup*) y “Comprensión de textos argumentativos” (*Curso Textos*) de la Universidad del Cauca, ofrecidos a través de la plataforma de aprendizaje Selene. En los cursos seleccionados, tuvimos 192 alumnos matriculados (92 *Curso Lean Startup* y 100 *Curso de Textos*), donde se tuvo una tasa global de finalización del 76 %. De los 192 estudiantes matriculados, 147 (76 %) presentaron el examen final y 96 (50 %) estudiantes aprobaron los cursos.

Utilizando el método descrito en la sección 3.3.2, obtuvimos todos los registros de interacción de los estudiantes con los exámenes presentados a través de la plataforma de aprendizaje, incluyendo la identificación de los estudiantes, las respuestas a los exámenes, la fecha y la hora de presentación, etc. Se obtuvieron un total de 297355 interacciones de los estudiantes durante el periodo de abril a julio de 2019. Decidimos incluir a aquellos estudiantes que completaron más de la mitad de los exámenes del curso (es decir, que al menos presentaron tres exámenes), para evitar medidas ruidosas; por lo tanto, el análisis incluye los datos de esos 147 estudiantes (69 *Curso Lean Startup* y 78 *Curso de Textos*).

4.2. Estudiantes identificados como sospechosos de deshonestidad académica

Utilizamos el método descrito en la Sección 3.3.3 para identificar a los estudiantes sospechosos de deshonestidad académica. El resultado del método son dos grupos, uno para cada una de las métricas de disimilitud propuestas. Así, tenemos un grupo para la métrica relativa a la similitud en las respuestas y otra para la métrica de cercanía en el tiempo de presentación de los exámenes.

Con el fin de identificar a los estudiantes con alta similitud en sus respuestas, tomamos todas las respuestas de los 5 módulos de los cursos, tomando un total de 75 respuestas presentadas por estudiante en el *Curso Lean Startup* y 142 respuestas presentadas por estudiante en el *Curso de Textos*. A continuación, utilizamos la matriz con 75 respuestas para el *Curso de Iniciación a la Lean* y utilizamos la matriz con 142 respuestas para el *Curso de Textos* para calcular las matrices de distancia. Por último, tomamos todos los pares de estudiantes con un 90 % de similitud en sus respuestas a lo largo de todo

el curso. En la práctica, esto significa que seleccionamos aquellos pares de estudiantes que tienen una distancia $ds_{i,j}$ inferior a 0,1 en la matriz de distancia. Basándonos en esta métrica, identificamos a 15 estudiantes para el *curso de iniciación a la informática* y a 11 estudiantes para el *curso de textos* que son sospechosos de fraude.

Por otro lado, también identificamos otra cohorte de estudiantes en función de la cercanía temporal de sus envíos. En este caso, para el cálculo de la matriz de distancia utilizamos las cinco marcas de tiempo en que se presentó el examen de cada módulo. Por último, los estudiantes detectados como sospechosos de fraude fueron seleccionados en función de una distancia media absoluta inferior a cinco minutos. Esto significaría que todos los exámenes se presentaron en promedio entre estos estudiantes a menos de cinco minutos de distancia. En este caso, se detectaron como sospechosos de fraude una cohorte de 17 alumnos para el *curso de iniciación a la informática* y 15 alumnos para el *curso de textos*.

Llama la atención cómo se identificaron los mismos alumnos en ambas métricas, a excepción de unos pocos alumnos, a pesar de haber utilizado métricas completamente diferentes y no relacionadas. Para ser conservadores, intersectamos los dos conjuntos de alumnos manteniendo sólo los 15 alumnos que están en ambas cohortes para el *Curso de iniciación a la informática* y 10 alumnos que están en ambas cohortes para el *Curso de textos*, haciendo entonces obligatorias las condiciones de similitud de respuestas y cercanía temporal.

Podemos visualizar cómo se organizan los estudiantes sospechosos de deshonestidad académica en la Figura 4.1. Esta visualización muestra un dendrograma para cada una de las métricas de similitud, que es un diagrama de árbol con las cohortes formadas mediante la creación de clusters de observaciones basadas en sus niveles de similitud. El nivel de similitud se mide en el eje vertical y las observaciones se especifican en el eje horizontal.

En el caso del dendrograma de alumnos identificados por la similitud de sus respuestas (Responses Cluster Dendrogram), el eje vertical muestra la similitud de los alumnos con una medida entre 0 y 1. Donde un 0 indica una coincidencia exacta de las respuestas y un 1 indica una diferencia del 100 % en sus respuestas. Obsérvese cómo los alumnos han sido agrupados en parejas o tríadas según la similitud que presentan en sus respuestas.

En el caso de la cohorte de estudiantes identificada por la cercanía temporal (Dendrograma de grupos temporales) el eje vertical muestra la proximidad de los envíos de los estudiantes medida en horas. Los grupos creados corresponden exactamente a los mismos grupos creados por la agrupación análoga en la Figura 2 basada en la similitud de las respuestas, a pesar de que las métricas son completamente diferentes. La tabla 4.1 muestra las similitudes en las respuestas y la cercanía de los envíos en tiempo para cada uno de

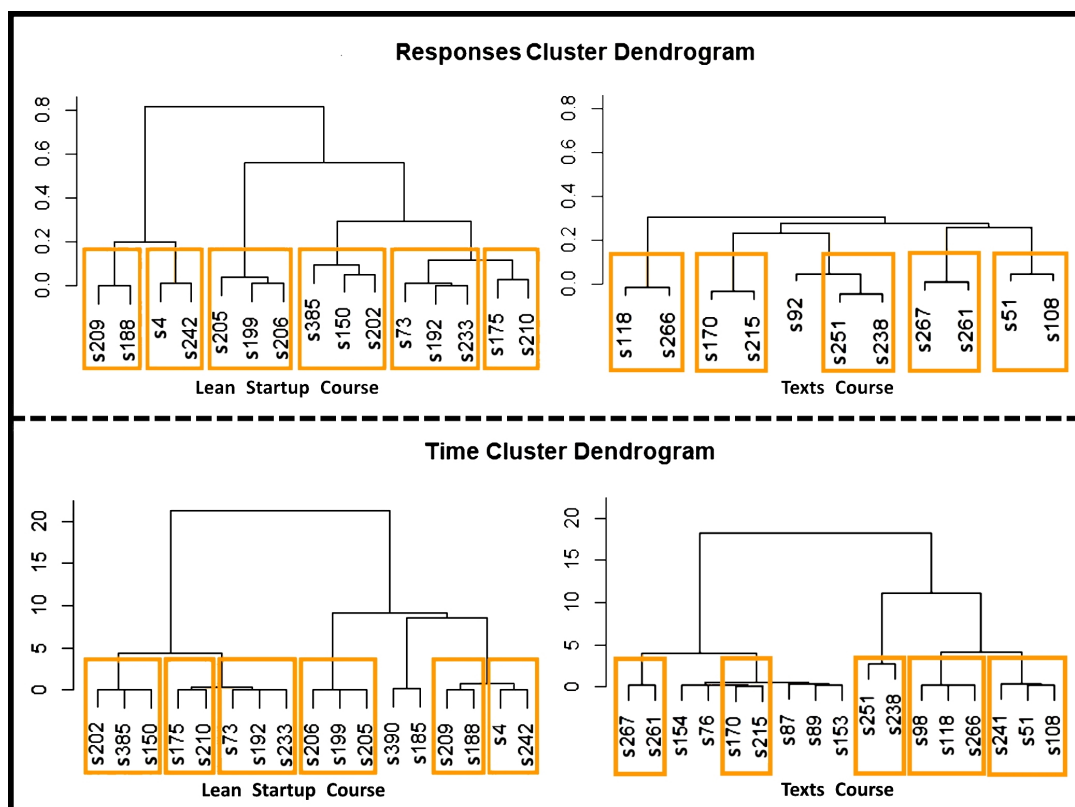


Figura 4.1: Dendrograms for each one of the metrics clustering.

los grupos generados. Como podemos ver todos los grupos tienen similitudes muy altas en sus respuestas y una distancia muy baja en el tiempo de su presentación. Podemos interpretar cada fila de la siguiente manera: La fila con los alumnos s205, s199 y s206, indicaría que este grupo de alumnos tiene el 96 % de sus respuestas en común (del total de 75 preguntas del curso), y que la distancia media entre las marcas de tiempo en que presentaron cada examen (de los cinco exámenes del curso) es de sólo 11 segundos. Por lo tanto, los resultados son evidentes y muestran valores que indican que estos grupos de estudiantes están trabajando juntos. Por otro lado, estos tres estudiantes, además de ser identificados por nuestro algoritmo, también fueron identificados como tramposos por el instructor del curso, ya que estos estudiantes se quejaron individualmente al profesor por la baja calificación recibida en el examen del módulo 4 aludiendo que la plataforma les había calificado mal. Entonces, el profesor revisó sus respuestas y descubrió que sus respuestas erróneas eran idénticas. Esto es una confirmación positiva de que el algoritmo funciona correctamente.

Además, la fila con los estudiantes s251 y s238, indica que este grupo de estudiantes tiene el 100 % de sus respuestas en común (del total de 142 preguntas del curso), y que la

Tabla 4.1: Similitud en respuestas y cercanía en el tiempo de envío.

Curso de Lean Startup			
Estudiantes	Similitud en respuestas	Cercanía en el tiempo	Nota final
s209, s188	100 %	0:02:05	92
s4, s242	100 %	0:02:36	72
s175, s210	97 %	0:00:07	96
s385, s150, s202	91 %	0:02:53	89
s73, s192, s233	99 %	0:03:18	92
s205, s199, s206	96 %	0:00:11	55
Textos Argumentativos			
Estudiantes	Similitud en respuestas	Cercanía en el tiempo	Nota Final
s118, s266	97 %	0:01:59	67
s170, s215	99 %	0:00:05	71
s251, s238	100 %	0:01:35	74
s261, s267	96 %	0:04:56	70
s51, s108	93 %	0:01:40	69

nota media final fue de 74 en una escala de 0 a 100. Así que estos estudiantes cometieron exactamente los mismos errores en 37 preguntas con cuatro respuestas posibles. Estos son sólo algunos ejemplos de análisis detallados, pero podemos establecer conclusiones similares del resto de casos.

4.3. Análisis del comportamiento y el compromiso de los estudiantes

Esta subsección de resultados se centra en analizar el comportamiento del grupo de estudiantes identificados como sospechosos de deshonestidad académica y compararlo con el comportamiento del resto de los alumnos. Para realizar dicha comparación, dividimos a todos los alumnos que fueron incluidos en el análisis considerando dos criterios: si el alumno fue detectado como sospechoso y si el alumno aprobó el curso. Entonces, se crean un total de tres cohortes: Sospechosos identificados, Estudiantes regulares (aprobados) y Estudiantes regulares (reprobados).

Como parte del primer análisis, se analizan las diferencias en las calificaciones por cohorte, la Figura 4.2 muestra una visualización de boxplot con la distribución de calificaciones por cohorte.

Los estudiantes pueden aprobar el curso con una nota superior a 60 en una escala de 0 a 100. La cohorte de sospechosos identificados tuvo una tasa de aprobación del 100 %,

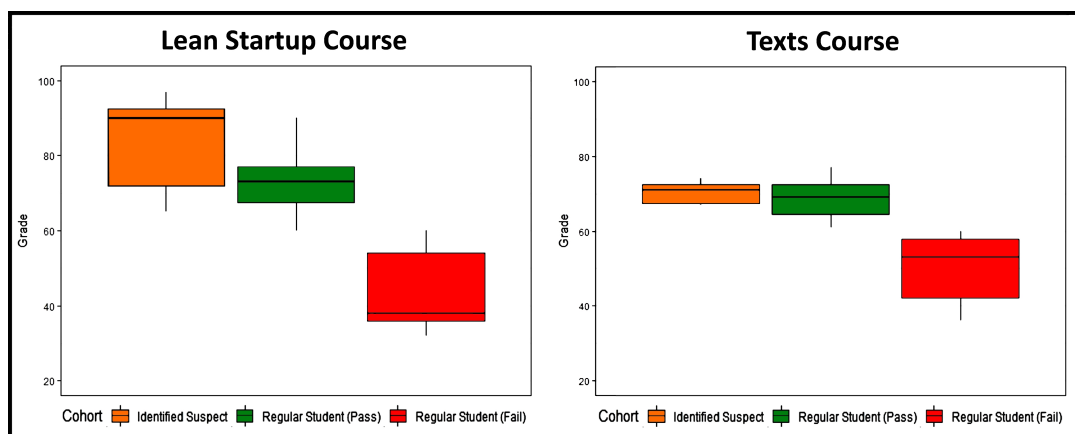


Figura 4.2: Boxplot distribution of grades by cohort.

mientras que el resto de los estudiantes tuvo una tasa de éxito del 72 % para el *curso de iniciación* y del 51 % para el *curso de textos*. Obsérvese que las mejores calificaciones las obtuvo la cohorte de estudiantes que fueron identificados como sospechosos de deshonestidad académica. Con una nota media del curso de 84 frente a los 72 del resto de alumnos que aprobaron y de los 40 de los alumnos que suspendieron para el *Curso de Iniciación a la Lealtad* y con una nota media del curso de 68 frente a los 63 del resto de alumnos que aprobaron y de los 41 de los alumnos que suspendieron para el *Curso de Textos*.

A continuación, analizamos la interacción y el comportamiento de los alumnos con los cursos para encontrar diferencias entre cohortes. La figura 4.3 muestra el número de métricas de compromiso e interacción que los alumnos tuvieron con los contenidos del curso.

En general, podemos observar que en ambos cursos la cohorte de alumnos que no aprobó, tuvo niveles de actividad ligeramente inferiores al resto, pero que las diferencias entre los alumnos que aprobaron el curso legítimamente y los que son sospechosos de deshonestidad académica no son tan notables. Además, encontramos que el número de interacciones con el contenido del *curso de iniciación a la lealtad* es mayor para los alumnos sospechosos de fraude en contraste con el *curso de textos*, donde los alumnos con mayor número de interacciones con el contenido son los alumnos que aprobaron legítimamente el curso. Una explicación puede ser que los alumnos identificados como sospechosos de fraude en el *Curso de Iniciación a la Lejanía*, interactúan mucho más con los contenidos en busca de las respuestas a los exámenes porque es un curso con mucho contenido. Esto no ocurre en el *curso de textos* porque es un curso más orientado a la práctica. Por otro lado, observamos que los diferentes vídeos vistos son mayores para los

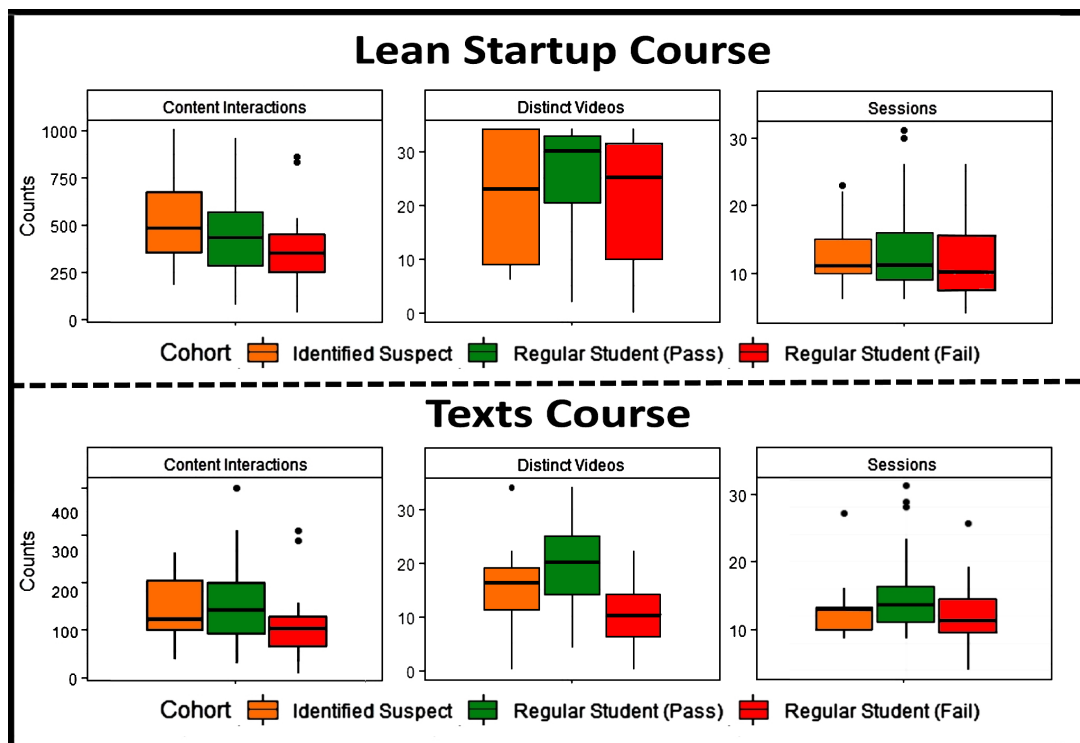


Figura 4.3: Interaction of students with the contents of the course.

alumnos regulares que aprobaron el curso. Creemos que estos alumnos tienen un mayor interés por aprender y no sólo por obtener una buena nota.

Realizamos pruebas t para cada una de las métricas comparando las cohortes de sospechosos identificados con los alumnos regulares que aprobaron, y no encontramos diferencias estadísticamente significativas entre las cohortes. Por tanto, aunque observamos algunas diferencias en las medias de algunos indicadores al comparar ambas cohortes, estas diferencias no son estadísticamente significativas (probablemente debido al tamaño de la muestra).

Por último, en la figura 4.4 se muestra el número de interacciones a lo largo del tiempo con el número de eventos por día durante el desarrollo del curso. Esta visualización puede ayudar a conocer si la actividad de cada cohorte fue más o menos espaciada en el tiempo, y el comportamiento de cada cohorte con respecto a las fechas de los exámenes programados. Los patrones generales parecen indicar que los estudiantes que fueron identificados como sospechosos de deshonestidad académica concentran su actividad más en las fechas de examen que el resto de las cohortes, que tienen una actividad un poco más espaciada en el tiempo en la mayoría de los casos. Podemos hacer coincidir los picos con las fechas de los exámenes (ver módulo debajo del eje x), y como recordatorio nos gustaría aclarar que la mayoría de los exámenes tenían dos fechas posibles. Para el *Curso*

de *Iniciación a la Lean* las dos fechas estaban separadas por una semana y para el *Curso de Textos* la segunda fecha era el día después del examen del módulo.

Para el *Curso Lean Startup*, el número medio de interacciones por día para la cohorte sospechosa de fraude es de 6,05 con una varianza de 108,9, mientras que el número medio de acciones para los estudiantes regulares que aprobaron el curso es de 4,69 con una varianza de 51,9. Por tanto, los alumnos sospechosos de fraude fueron más activos con la plataforma, y su actividad se concentró en menos días, más concretamente en las fechas de los exámenes, mientras que los alumnos que aprobaron el curso sin ser académicamente deshonestos fueron un poco menos activos, y su actividad fue más espaciada durante todo el calendario del curso.

En cambio, para el *Curso de Textos*, el número medio de interacciones por día para la cohorte sospechosa de fraude es de 2,71 con una varianza de 33,76, mientras que el número medio de interacciones para los alumnos regulares que aprobaron el curso es de 3,48 con una varianza de 49,93. Vemos que en este caso los alumnos regulares que aprobaron el curso son más activos que los alumnos sospechosos de fraude y también que su actividad está menos concentrada en las fechas de los exámenes. Esto puede deberse a que el *curso de textos* es más práctico y las respuestas a los exámenes no están explícitas en los contenidos del curso, a diferencia del *curso Lean Startup* que es un curso más teórico. Por otro lado hemos encontrado que los alumnos identificados como sospechosos de fraude, tienen interacciones previas a los exámenes, justo cuando el tutor estaba soltando una lectura necesaria para resolver el examen.

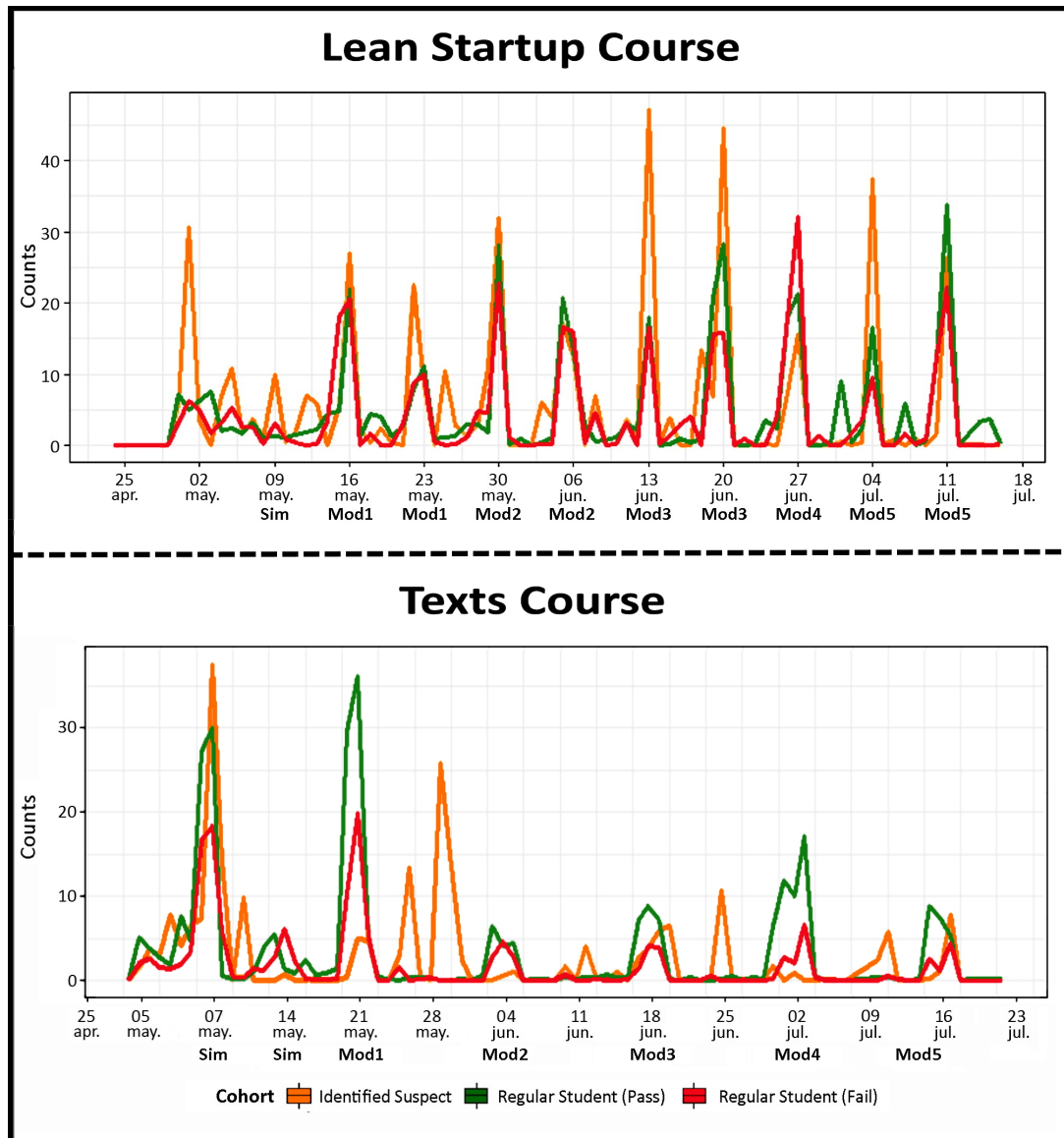


Figura 4.4: Average number of interactions over time separated by cohort.

Capítulo 5

Caracterización de colaboraciones en MPOC/SPOC

Índice

6.1. Descripción de la Herramienta de seguimiento	85
6.1.1. Indicadores de seguimiento	85
6.1.2. Diseño de la herramienta	86
6.1.3. Implementación de la herramienta	87
6.2. Desarrollo de caso de estudio	89
6.2.1. Resultados de seguimiento	89
6.3. Evaluación de usabilidad y utilidad	96
6.3.1. Resultados en cuanto a Usabilidad	97
6.3.2. Resultados en cuanto a Utilidad	98
6.3.3. Resultados de Opinión	99

La colaboración se considera uno de los principales motores del aprendizaje y se ha estudiado ampliamente en numerosos contextos, incluidos los cursos en línea tradicionales. Sin embargo, estudios previos se han restringido al análisis de la colaboración basado en el foro y las interacciones sociales, sin tener en cuenta otras posibilidades como la sincronización en las interacciones con la plataforma. Por ello, en esta sección se presenta un estudio de caso con el objetivo de implementar un enfoque basado en datos para caracterizar las colaboraciones en SPOCs. Aplicamos el algoritmo para la detección de estudiantes con sospecha de deshonestidad académica como indicador de colaboración, y lo aplicamos a los datos de 4 SPOCs de la Universidad del Cauca que se

ejecutaron en el 2020 y 2021. Encontramos tres perfiles diferentes de cuentas de usuario, que se agrupaban en parejas y comunidades más grandes que mostraban diferentes tipos de asociaciones entre cuentas de usuario. La caracterización de estas cuentas de usuario sugirió que algunas de ellas podrían representar auténticas asociaciones colaborativas de aprendizaje en línea, pero que según los lineamientos de los cursos se presentan como comportamientos deshonestos.

5.1. Contexto del estudio

Esta sección se presentan los cursos y las variables que se tuvieron en cuenta para el estudio. En total se contó con datos de 8 cursos y con 5 variables que indican la interacción de los estudiantes con la plataforma de aprendizaje.

5.1.1. SPOCs utilizados para el estudio

Los datos de este estudio provienen de los cuatros SPOCs descritos en la sección de metodología. Para que el estudio sea mas profundo, en esta ocasión se han tenido en cuenta dos cortes de los cursos ofrecidos. La corte ofrecida en el año 2020 y la corte ofrecida en el año 2021. Usualmente se ofrecen dos cortes por año. Sin embargo, debido a la situación presentada por el Covid19, solo se ofreció una corte por año. Es así como para este estudio se tendrán en cuenta datos de 8 cursos ofrecidos.

5.1.1.1. Curso de Emprendimiento con Lean Startup corte 2020 y 2021:

Para el corte del 2020 del curso de Lean Starup se contaron con 138 estudiantes registrados de los cuales sólo 28 finalizaron y aprobaron el curso. Creemos que la temática y el inicio de la pandemia afecto la participación en el curso y por eso se presentó una baja tasa de finalización. Para el corte del 2021 se contaron con 211 estudiantes registrados en el curso y con un porcentaje de aprobación del 40 % finalizaron 84 estudiantes.

5.1.1.2. Curso de Comprensión de Textos Argumentativos corte 2020 y 2021:

En el curso de Comprensión de Textos Argumentativos para el 2020 se obtuvieron 86 estudiantes de los cuales aprobaron el curso 47, siendo el 55 % del total y para el 2021 se logró un porcentaje de aprobación del 69 % de un total de 78 estudiantes registrado para ese corte.

Tabla 5.1: Estadísticas de registro y finalización de los cursos

Curso	Inscritos	Aprobaron	Finalizaron	Porcentaje de Aprobación
Lean Startup -2020	138	28	28	20 %
Lean Startup -2021	211	84	90	40 %
Textos Argumentativos -2020	86	47	45	55 %
Textos Argumentativos -2021	78	54	55	69 %
Danza Folclórica - 2020	109	65	65	60 %
Danza Folclórica - 2021	107	68	71	64 %
LaTeX - 2020	83	59	53	71 %
LaTeX - 2021	78	58	56	74 %

5.1.1.3. Curso de Edición de Textos científicos y literarios con LaTeX corte 2020 y 2021:

Para el corte del 2020 del curso de Edición de Textos científicos y literarios con LaTeX se contaron con 83 estudiantes registrados con una tasa de finalización del 71 % y para el corte del 2021 se contaron con 78 registrados y 56 aprobados. Siendo este uno de los cursos que más alta tiene la tasa de finalización.

5.1.1.4. Curso de Danza Folclórica corte 2020 y 2021:

Para el corte del 2020 del curso de Danza Folclórica se contaron con un total de 109 estudiantes registrados, de los cuales 65 presentaron las tres unidades y aprobaron el curso. Para el corte del 2021 se contaron con 107 registrados y 71 aprobados.

La Tabla 5.1 presenta la cantidad de estudiantes que se inscribieron en el curso, cuantos presentaron examen final y cuantos aprobaron. En el corte del 2020 se presentó la emergencia sanitaria por Covid-19 lo que hizo que hubieran una gran cantidad de abandono, especialmente para el curso de Introducción al emprendimiento con Lean Startup.

5.1.2. Variables utilizadas

Como se mencionó en la sección 3.3, las variables a tenerse en cuenta para el agrupamiento mediante el método de Kmeans están relacionadas con las interacciones de los estudiantes con la plataforma de aprendizaje. Una gráfica de las variables que se han recolectado para los 8 cursos se muestran en la Figura 5.1.

Las variables que se muestran en la Figura 5.1 se resumen en 4 grupos. Interacciones con vídeos, menú de navegación (contenidos), exámenes y foros. Además que se ha agregado las notas finales de los estudiantes, sin embargo, no se tienen en cuenta al momento de realizar las agrupaciones.

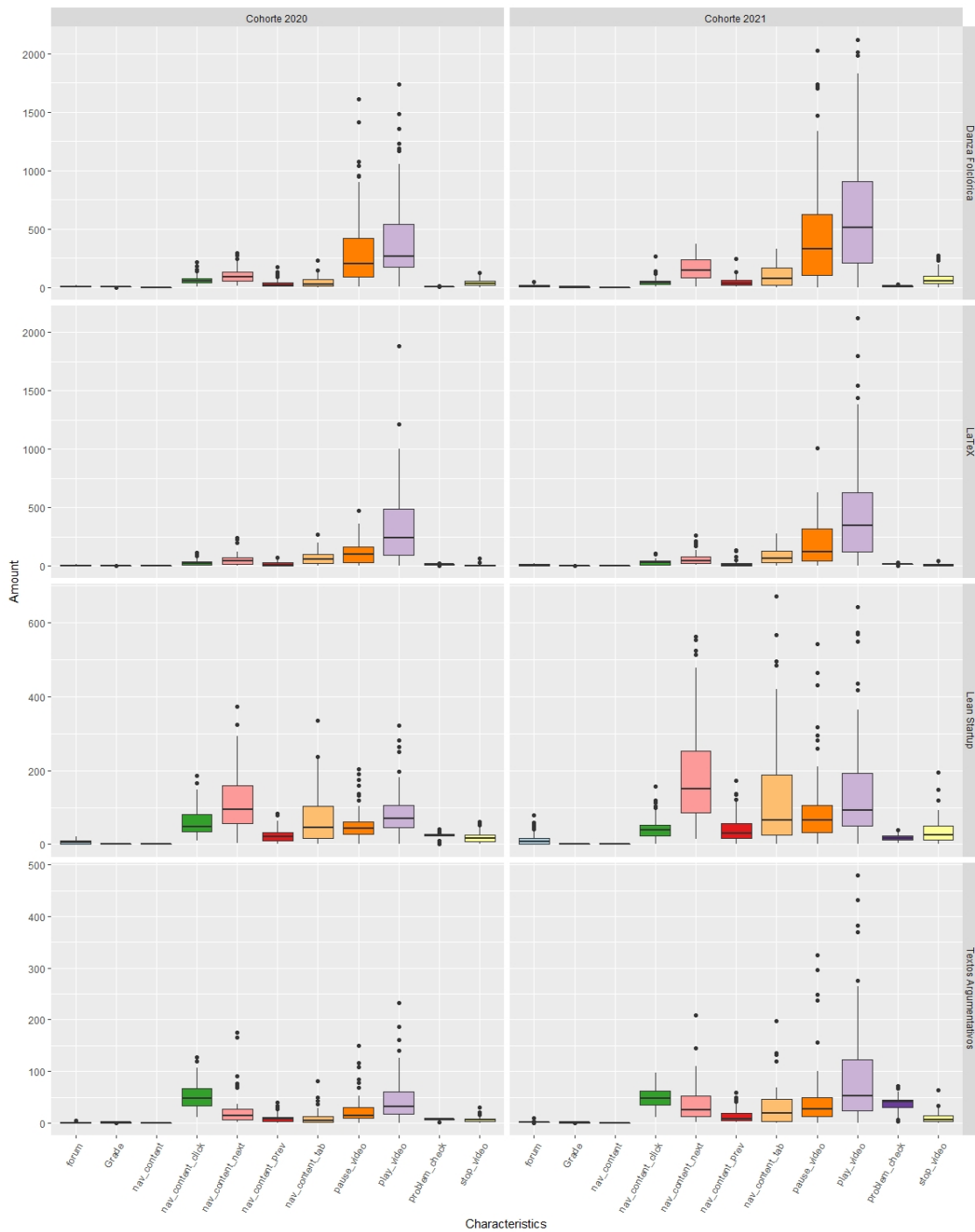


Figura 5.1: Variables de interacción en los 8 cursos para las cortes 2020 y 2021

5.1.3. Métricas y método de agrupamiento

Para la agrupación de los estudiantes se optó por un método de clustering, el cual es una técnica de Machine Learning, en la que el aprendizaje realizado es no supervisado

(unsupervised learning). Desde un punto de vista práctico, el clustering juega un papel muy importante en aplicaciones de data mining, tales como exploración de datos científicos, recuperación de la información y minería de texto, aplicaciones sobre bases de datos espaciales (tales como GIS o datos procedentes de astronomía), aplicaciones Web, marketing, diagnóstico médico, análisis de ADN en biología computacional, y muchas otras. Dentro de los métodos de Clustering, para medir la similaridad entre objetos se suelen utilizar diferentes formas de distancia: distancia euclídea, de Manhattan, de Mahalanobis, etc.

Normalmente se aplican técnicas de clustering cuando no tenemos una idea clara de los grupos subyacentes en una población, y entonces se agrupan los sujetos con base a alguna similitud inherente entre ellos [111]. Por lo tanto, aplicamos la metodología de agrupación en la subsección 3.3.4 para encontrar diferentes tipos de cuentas de estudiantes en función de su compromiso con la plataforma de aprendizaje. Este proceso de agrupación se aplica por separado en los 4 cursos y sus cohortes. Este tipo de enfoques de clustering para encontrar diferentes perfiles de estudiantes en los MOOCs han sido utilizados en estudios anteriores con éxito ferguson2015examinando, khalil2017clustering, chen2017examinando, ruiperez2017detectando.

Para el estudio se consideraron los métodos de Particionado y Recolocación debido a que lo que se busca es la división de los datos para lograr clasificar individuos en K clusters. Dentro de particionado y recolocación, se destacan en la literatura dos métodos: Expectation Maximization (EM) y K-medias. A partir de la experimentación se comprobado que se obtienen mejores resultados con EM. Sin embargo, el proceso para implementación es más complejo y el tiempo requerido para obtener los datos es demorado. Por otra parte, el método K-Medias ha logrado equipararse al método EM cuando se tratan de datos con variables no gaussianas. Además, el proceso de implementación es muy sencillo y el tiempo de procesamiento bajo. Por lo que se ha optado por el método K-means por su fácil implementación y que la variables a utilizar no son gaussianas.

5.1.4. Definición de colaboración en este estudio

Como hemos visto en los trabajos relacionados en la sección 2.3, la colaboración y el aprendizaje colaborativo se han definido y puesto en práctica de muchas maneras diferentes. En este estudio, nos enfocamos en la noción previamente reportada de sincronidad temporal como un estado en el cual las actividades de un grupo colaborador están sincronizadas a lo largo del tiempo, es decir, cuando los miembros del grupo están trabajando en la misma actividad al mismo tiempo, tenemos que está surgiendo una co-

laboración [100] y además, envío de respuestas similares. En nuestro escenario de caso, detectamos estas colaboraciones basados en los mismos criterios para detectar sospechas de fraude. Es decir, se identificaron las colaboración cuando los estudiantes presentan más del 90 % de similitud en sus respuestas y tienen una diferencia de tiempo en el envío de menos del 5 minutos. Lo que en algunos casos puede ser muy inferior. La razón para tener en cuenta estos indicadores es que es la probabilidad estadística de que dos o más cuentas presenten cuestionarios similares y casi al mismo tiempo, es muy baja. Nos referimos a esto como “colaboraciones invisibles” que no se pueden detectar simplemente mirando la interacción social en línea en foros u otras herramientas sociales.

5.2. Aplicación del método K-means

La Figura 5.2 muestra los tres clústers formado a partir de las variables de interacción con vídeos, interacción con contenidos, cantidad de envíos de exámenes y foros publicados. Se obtuvieron tres clústers por cada uno de los cursos y cohortes. Donde cada cluster es representado con un color diferente.

Los clústers obtenidos en cada curso son diferentes y dependen de las interacciones que hayan tenido los estudiantes a través de la plataforma de aprendizaje. Cada cluster se forma con estudiantes que tienen similares características de interacción con los cursos. La siguiente sección describe las características de los clústers formados.

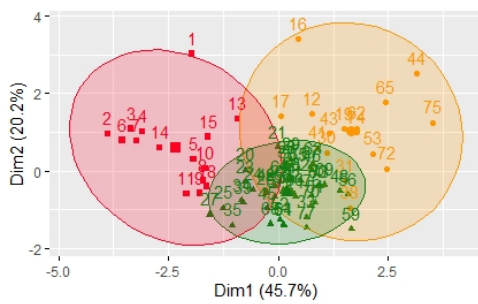
5.2.1. Caracterización de los estudiantes.

Las Figuras 5.3 y 5.4 presentan las variables de interacción relacionadas a cada clúster.

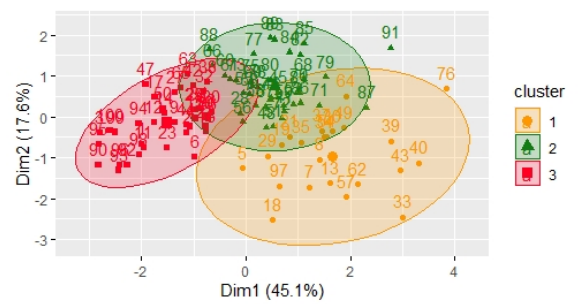
A partir de estas gráficas se obtienen las características de los tres grupos que se forman cuando se aplica el método de K-means. La Figura 5.3 muestra las variables de los cursos de *Lean Startup* y *Textos Argumentativos* para los años 2020 y 2021. Donde se visualiza que cada curso tiene características similares en las dos cohortes, pero que tienen características diferentes entre los cursos. Como todos los cursos trabajan sobre una misma metodología de enseñanza, creemos que la diferencia se debe a los contenidos y la temática del curso.

La Figura 5.4 muestra las variables para los cursos *Danza Folclórica* y *LaTeX* para los años 2020 y 2021.

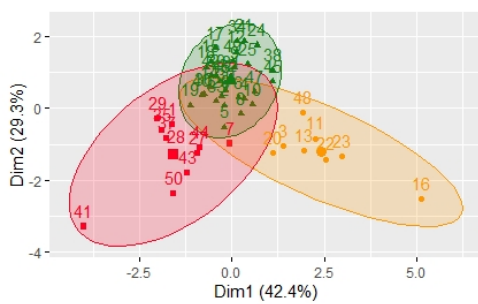
La figura muestra como los cursos de *Danza Folclórica* y *LaTeX* tienen comportamientos similares entre ellos e incluso con el curso de *Lean Startup*. Es decir, el único



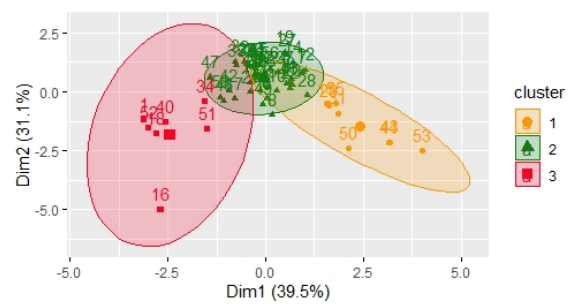
(a) Lean Startup - 2020



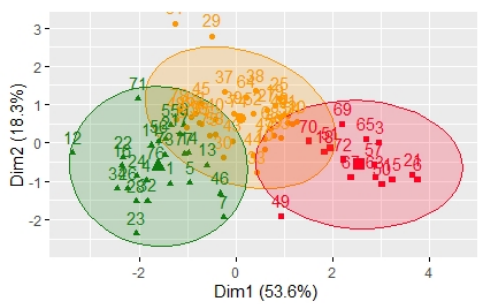
(b) Lean Startup - 2021



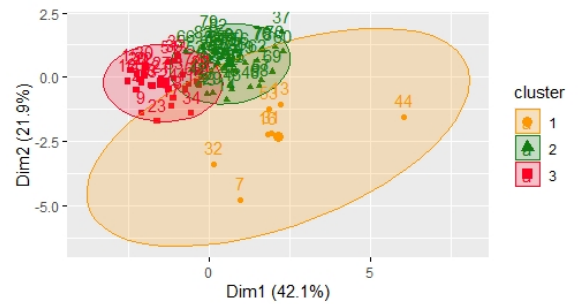
(c) Textos Argumentativos - 2020



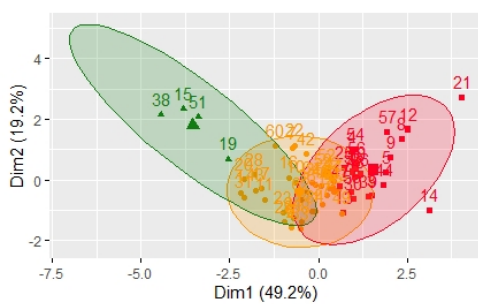
(d) Textos Argumentativos - 2021



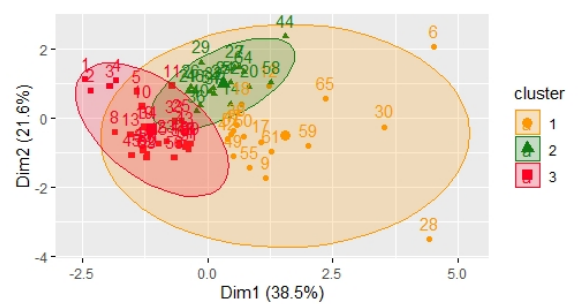
(e) Danza Folclórica - 2020



(f) Danza Folclórica - 2021



(g) LaTeX - 2020



(h) LaTeX - 2021

Figura 5.2: Clústers obtenidos en cada uno de los cursos y cohortes

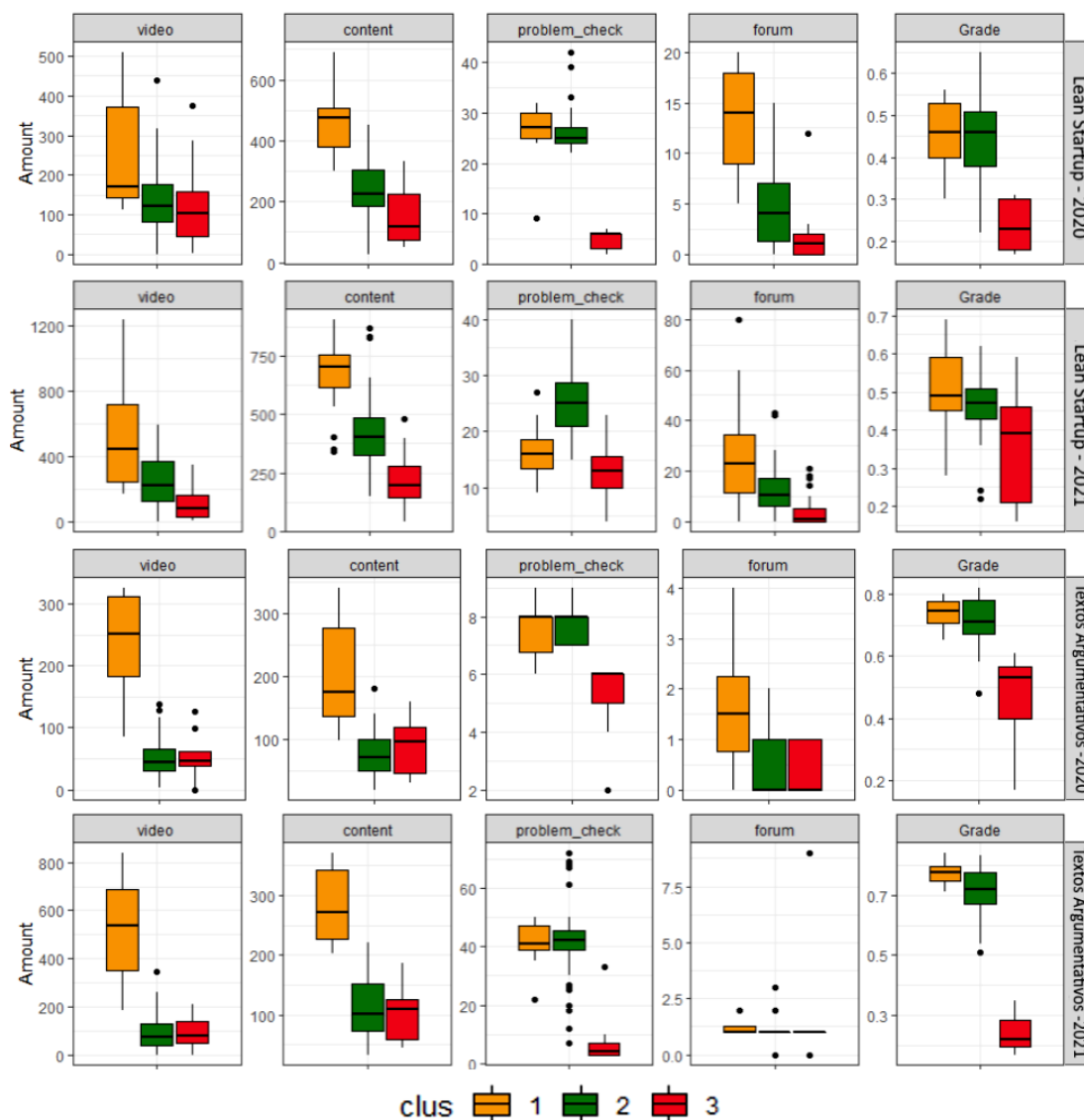


Figura 5.3: Visualización de Boxplot de las variables para los cursos de *Lean Startup* y *Textos Argumentativos*

curso que es un poco diferente es el de *Textos Argumentativos* en el cual la cantidad de interacciones entre el cluster 2 y 3 son comparables, excepto para la cantidad de exámenes presentados. A continuación se describen los tres grupos en general para todos los cursos.

- Cluster 1:** Las cuentas que pertenecen a este grupo tienen una alta nota final y la media más alta para las variables de interacción con vídeos, interacción en contenidos, intentos en exámenes y actividad en los foros. Además, tienen una varianza muy alta para todas las variables, por lo que hay diferentes tipos de cuentas en

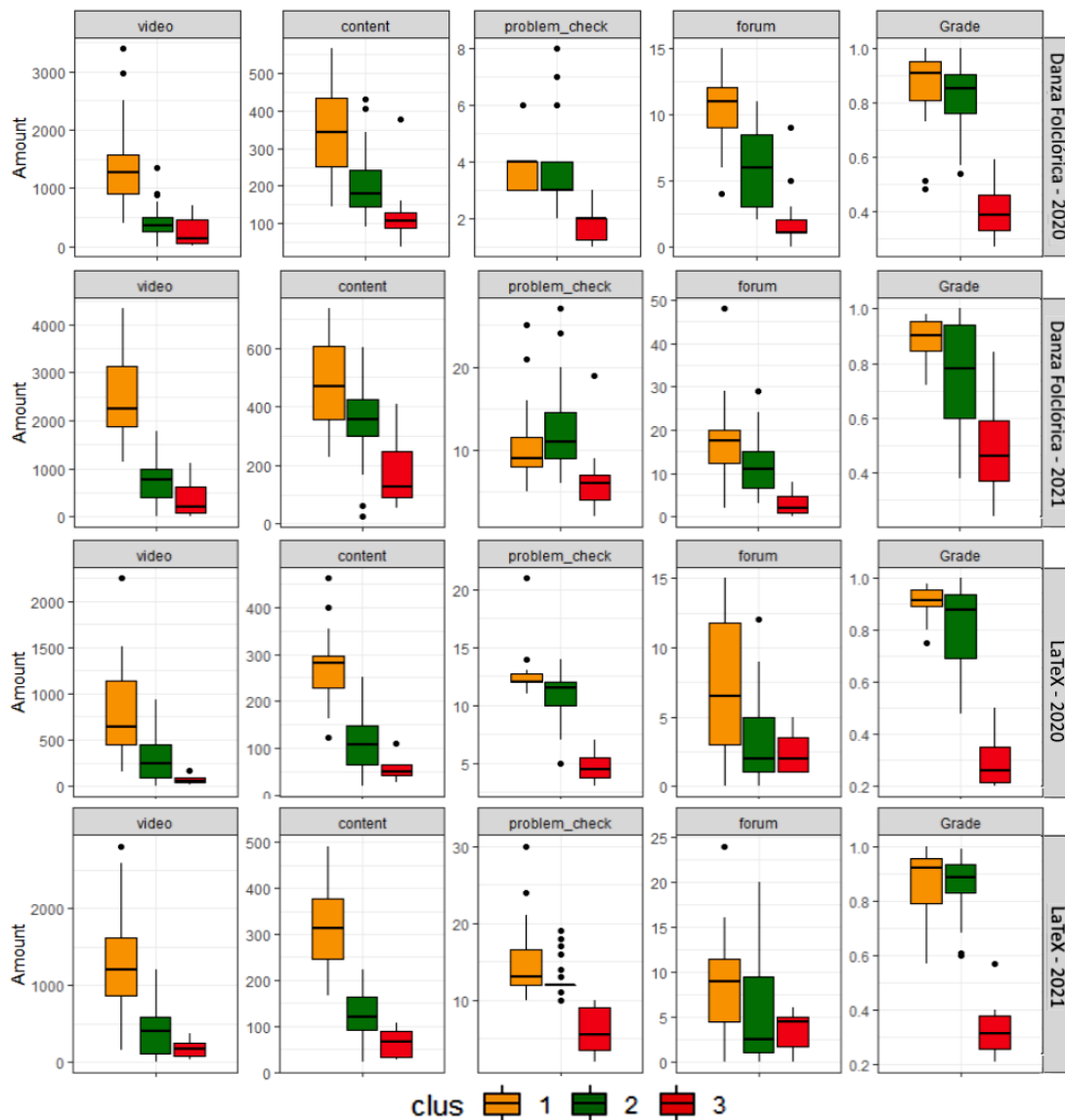


Figura 5.4: Visualización de Boxplot de las variables para los cursos de *Danza Folclórica* y *LaTeX*

cuanto a la cantidad de interacción, algunas con actividad moderada en plataforma y algunas con la mayor actividad del curso. En general, dado que este cluster tenía los valores más altos para las variables de actividad y también un valor alto de nota final, estas cuentas se esforzaron e invirtieron tiempo en el curso logrando altas calificaciones y demostrando su compromiso con el curso.

- Cluster 2:** Este cluster contiene cuentas que también tenían un valor alto en la nota final. Sin embargo, hubo diferencias importantes en comparación con el clúster 1 en lo que respecta al resto de las variables. Lo más importante es que en términos

de interacción con vídeos, las cuentas del cluster 1 tenían un uso muy alto de vídeos, mientras que en el cluster 2 era el caso contrario, donde los usuarios de la mayoría de las cuentas interactuaban muy poco con los vídeos. Además, el valor de la variable de interacción con contenidos era también menor que en el cluster 1. Por otro lado la variable de cantidad de intentos para resolver un examen es comparable al del cluster 1 o incluso más alta. Por lo tanto, los usuarios de las cuentas de este clúster lograron altas calificaciones, y pudieron lograrlo viendo muy pocos vídeos, estando activos menos días y con menos envíos que los usuarios de las cuentas del clúster 1. Por lo tanto, nuestra hipótesis es que, o bien los estudiantes que manejan estas cuentas ya tenían conocimientos previos sobre el tema del curso y se limitaron a resolver las actividades requeridas para obtener los certificados, o bien pudieron haber estado realizando algunas acciones ilícitas como parte de la colaboración que les facilitó el camino para obtener un certificado sin mucho esfuerzo.

- **Clúster 3:** El último grupo de cuentas de usuario se distingue claramente de los otros dos grupos por su nota final es mucho más baja que en los otros dos. Esto significa que la mayoría de las cuentas de este clúster no consiguieron aprobar el curso. Los valores de las otras variables también es el más bajo de todos los grupos, indicando que no tuvieron interacción con contenidos, con vídeos o incluso con los exámenes. Estas cuentas no demuestran interés en el curso y pensamos que aquellos estudiantes que hacen parte de este clúster y lograron aprobar el curso, utilizaron algún método de deshonestidad académica que les llevara a conseguir la aprobación.

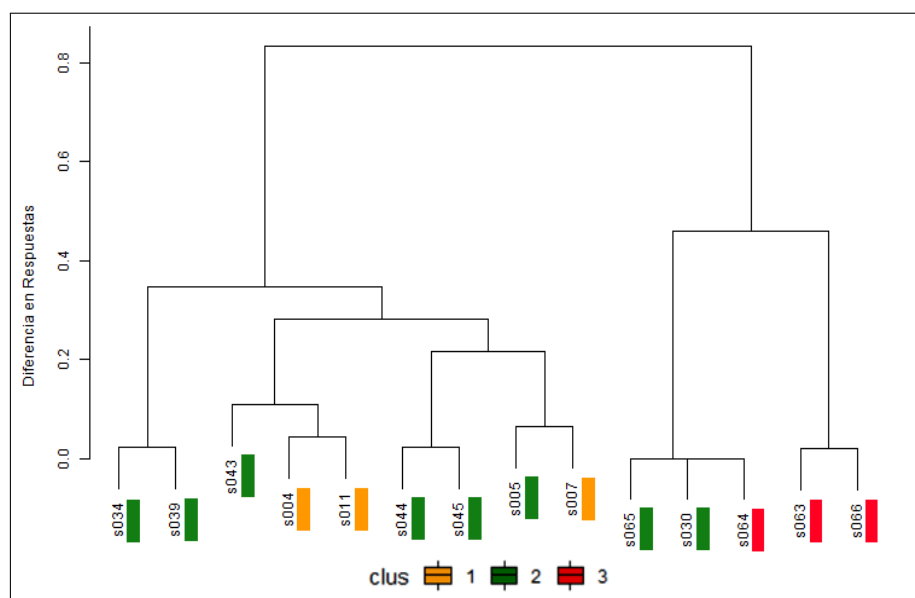
La Tabla 5.2 contiene la cantidad de estudiantes divididos en los tres clústers por curso y cohorte. También muestra el porcentaje de estudiantes en cada clúster por curso.

5.3. Detección de colaboraciones

Para la detección de las colaboraciones en los cursos se utilizó el método descrito en la Sección 3.3.3 para detectar a los estudiantes que tienen las mismas respuestas en los exámenes presentados y la sincronicidad en el envío. Luego a partir de la matriz de distancias obtenida se gráfcaron dendrogramas que muestran la relación entre los estudiantes. Las Figuras 5.5, muestran las relaciones entre los estudiantes en parejas y algunos grupos de más de dos estudiantes y la clasificación según el clúster al cual pertenece el estudiante.

Tabla 5.2: Clusterización de los estudiantes de los cursos

Curso	Cluster 1		Cluster 2		Cluster 3	
<i>Lean Startup -2020</i>	13	17 %	54	71 %	9	12 %
<i>Lean Startup -2021</i>	20	22 %	34	38 %	36	40 %
<i>Textos Argumentativos -2020</i>	8	17 %	29	62 %	10	21 %
<i>Textos Argumentativos -2021</i>	7	13 %	41	75 %	7	13 %
<i>Danza Folclórica - 2020</i>	26	40 %	28	43 %	11	17 %
<i>Danza Folclórica - 2021</i>	21	30 %	33	46 %	17	24 %
<i>LaTeX - 2020</i>	21	36 %	35	59 %	3	5 %
<i>LaTeX - 2021</i>	22	38 %	30	52 %	6	10 %

Figura 5.5: Dendrograma para el curso de *Lean Startup - 2020*

Para el curso de *Lean Startup - 2020* se identificaron cuatro parejas y dos grupos de tres personas. En total 14 estudiantes identificados como sospechosos de deshonestidad académica. De los cuales 3 son clasificados como parte del clúster 1, 8 del clúster 2 y 3 del clúster 3. Figura 5.5.

Para el curso de *Lean Startup - 2021* se identificaron cuatro parejas. En total 8 estudiantes identificados como sospechosos de deshonestidad académica. De los cuales 3 son clasificados como parte del clúster 1, 1 del clúster 2 y 4 del clúster 3. Figura 5.6

Para el curso de *Textos Argumentativos - 2020* se identificaron tres parejas. En total 6 estudiantes identificados como sospechosos de deshonestidad académica, todos parte del clúster 2. Figura 5.7.

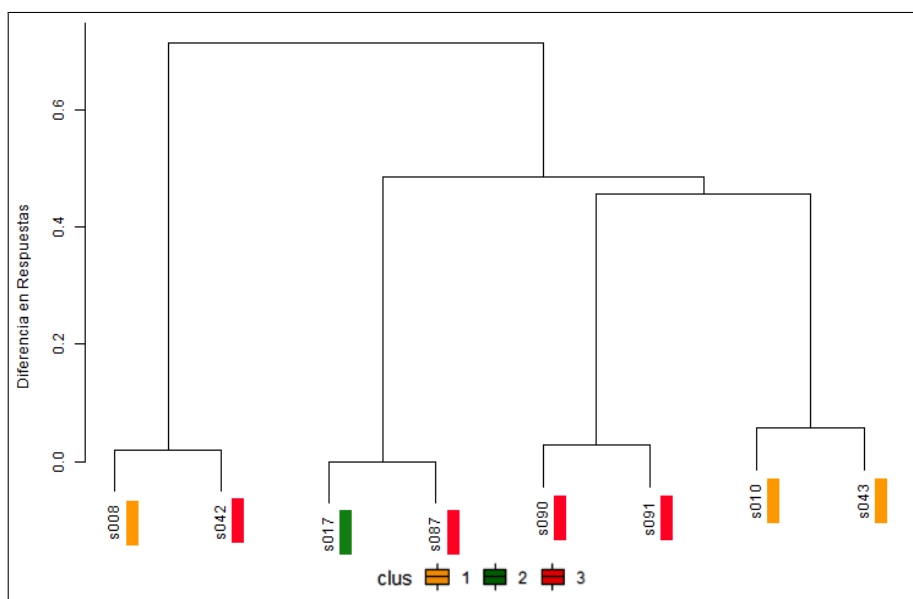


Figura 5.6: Dendrograma para el curso de *Lean Startup* - 2021

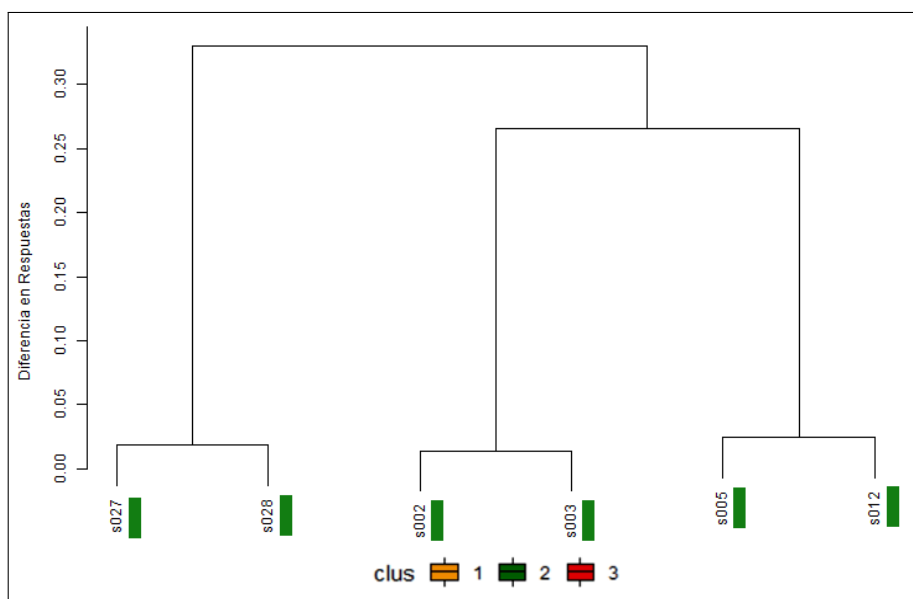


Figura 5.7: Dendrograma para el curso de *Textos Argumentativos* - 2020

Para el curso de *Textos Argumentativos* - 2021 se identificó una pareja, un grupo de tres y un grupo de cuatro estudiantes. En total 9 estudiantes identificados como sospechosos de deshonestidad académica. 8 estudiantes parte del clúster 2 y sólo un estudiante parte del clúster 1. Figura 5.8.

Para el curso de *Danza Folclórica* - 2020 se identificaron 7 parejas. En total 14 estudiantes identificados como sospechosos de deshonestidad académica. 9 estudiantes parte

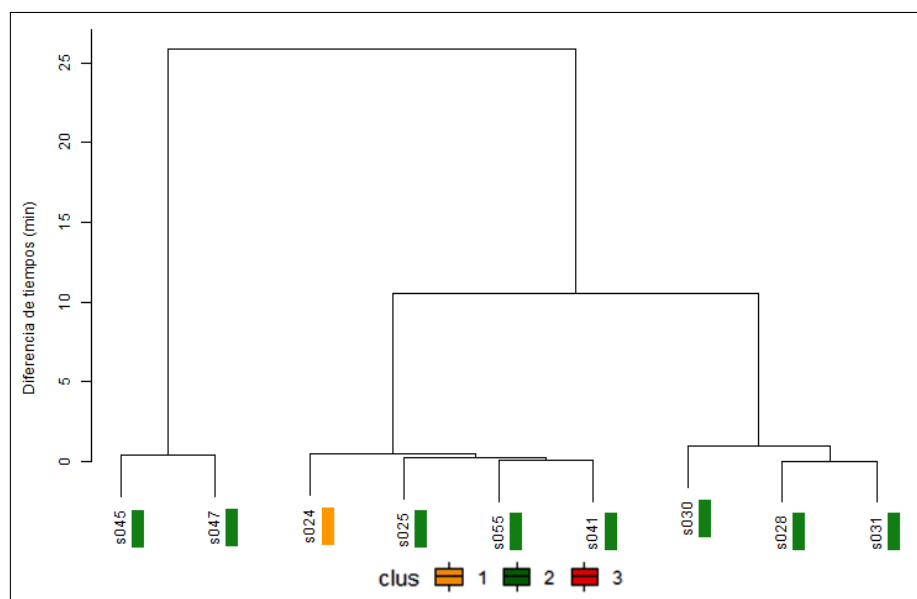


Figura 5.8: Dendrograma para el curso de Textos Argumentativos - 2021

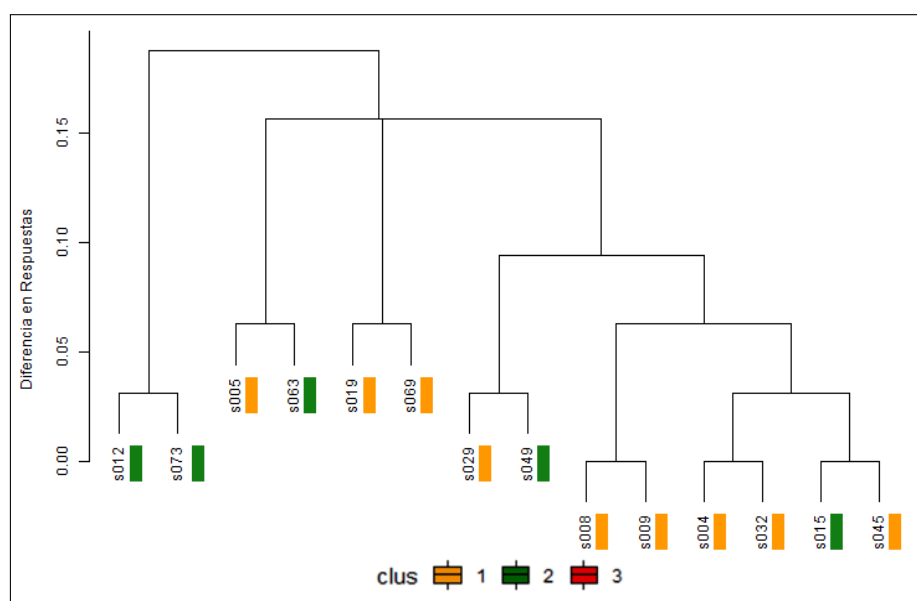


Figura 5.9: Dendrograma para el curso de *Danza Folclórica* - 2020

del clúster 1 y 5 estudiantes parte del clúster 2. Figura 5.9.

Para el curso de *Danza Folclórica* - 2021 se identificó un grupo de 8 estudiantes y una pareja. En total 10 estudiantes identificados como sospechosos de deshonestidad académica. 6 estudiantes parte del clúster 1 y 4 estudiantes parte del clúster 2. Figura 5.10.

Para el curso de *LaTeX* - 2020 se identificaron tres parejas. En total 6 estudiantes iden-

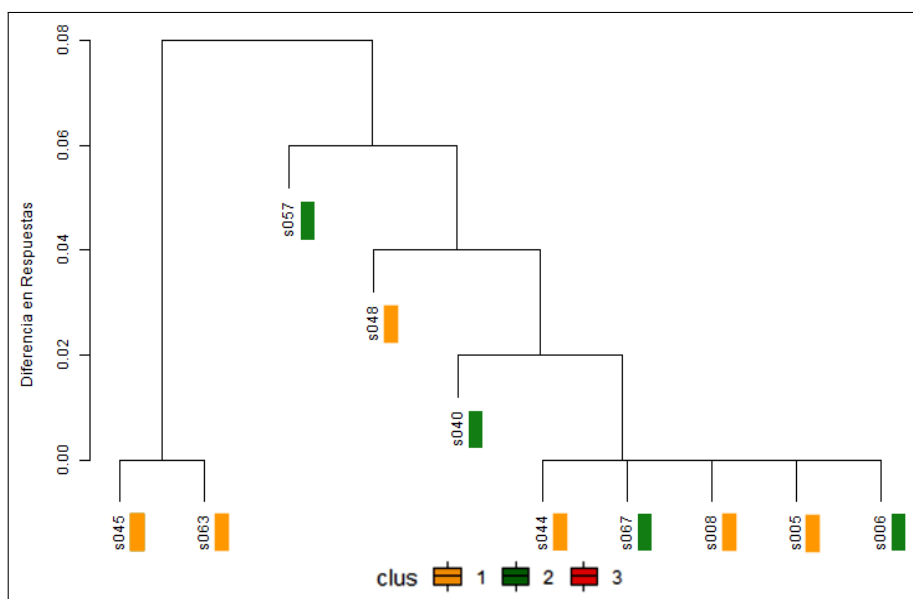


Figura 5.10: Dendrograma para el curso de *Danza Folclórica - 2021*

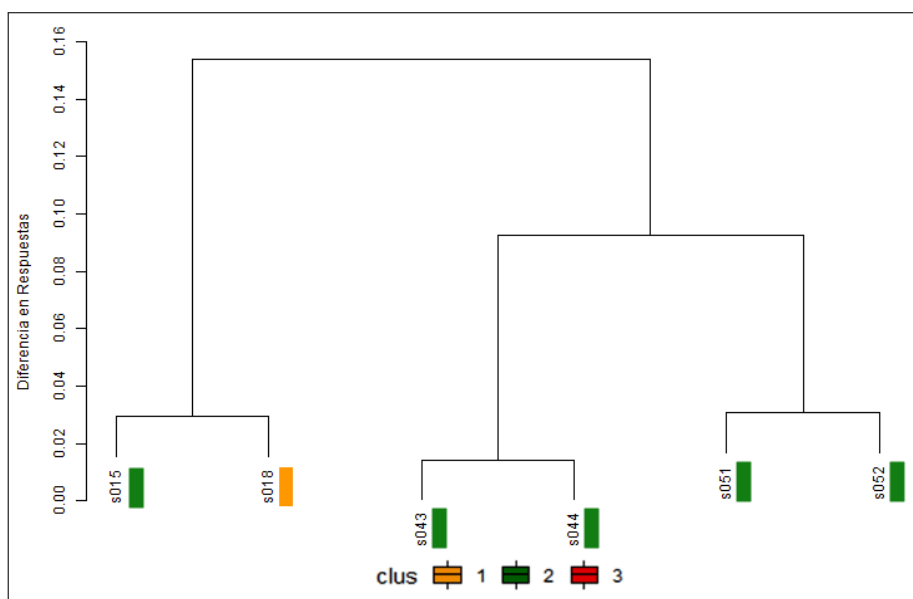


Figura 5.11: Dendrograma para el curso de *LaTeX - 2020*

tificados como sospechosos de deshonestidad académica. 5 estudiantes parte del clúster 1 y un estudiante parte del clúster 2. Figura 5.11.

Para el curso de *LaTeX - 2021* se identificó un grupo de 4 estudiantes. Un estudiante del clúster 1 y 3 estudiantes parte del clúster 2. Figura 5.12.

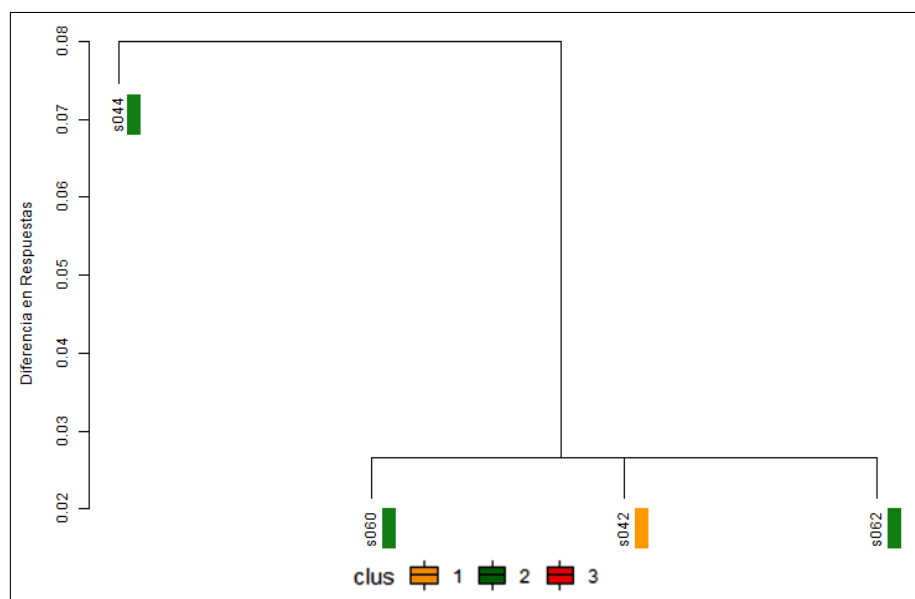


Figura 5.12: Dendrograma para el curso de *LaTeX* - 2021

5.4. Caracterización de las Colaboraciones

En esta subsección se describen las asociaciones de las colaboraciones de estudiantes y su asignación de clúster. Se tuvieron en cuenta asociaciones de parejas y comunidades de más de dos estudiantes.

5.4.1. Asociaciones de parejas

En los cursos nos encontramos con diferentes tipos de asociaciones de parejas. Algunas que trabajan en equipo y otras en las que trabaja un estudiante y comparte su trabajo con el otro. A continuación se describen las asociaciones encontradas:

- Asociación 1 “Colaboración fructífera” (cluster 1 y cluster 1): Esta asociación representa a dos estudiantes del cluster 1 trabajando juntos. Como hemos informado en la subsección anterior, los los estudiantes del clúster 1 se esforzaron mucho en la plataforma para conseguir la aprobación, con valores altos en todas las variables. Por lo tanto, esta asociación podría representar a dos estudiantes que se tomaban el curso en serio, y que colaboraban recíprocamente entre sí para conseguir mejores calificaciones. Sin embargo, dentro de los lineamientos del curso el compartir las respuestas a sus exámenes es considerado un comportamiento de deshonestidad académica.

- Asociación 2 “Free-riding” (clúster 1 y clúster 2) ó (clúster 1 y clúster 3): Esta asociación representa a un estudiante del clúster 1 y con uno del clúster 2 ó un estudiante del clúster 1 y con uno del clúster 3, lo que podría representar una colaboración entre dos estudiantes; sin embargo, esta relación no es equitativa. Según las variables de agrupación elegidas, el clúster 1 tiene una interacción de plataforma más alta que el clúster 2, y aún más que el clúster 3. Sin embargo, ambos estudiante en esta asociación obtienen notas altas. En esta asociación, el estudiante del clúster 1 se esforzaría en su trabajo en la plataforma, mientras que el estudiante del clúster 2 o 3 no se esforzaría mucho pero aún así obtendría un aprobado con la ayuda del estudiante del clúster 1. Esto es lo que en la literatura se lo conoce como “Free-riding”.
- Asociación 3 “Colaboración ilícita” (clúster 2 y clúster 2): En esta asociación ambas cuentas pertenecen al clúster 2, por lo tanto este caso representa dos cuentas que no demostraron mucho esfuerzo en el curso en términos de vídeos o interacción con contenidos, pero aun así pudieron obtener una buena calificación.
- Asociación 4 (clúster 3 y clúster 3): Encontramos dos asociaciones de parejas del clúster 3. Estudiantes que no interactuaron con los contenidos del curso, se copiaron las respuestas y no lograron obtener la calificación para aprobar el curso. Asociaciones que a mitad del curso lo abandonaron.

5.4.2. Asociaciones de comunidades

Se le llamó comunidades a los grupos que se forman por más de dos estudiantes. Pudimos encontrar desde grupos con tres estudiantes hasta grupos con 8 estudiantes. Entre estas comunidades tenemos tres tipos:

- Comunidad 1: Comunidad en dónde sólo hay uno de los estudiantes del clúster 1 y el resto pertenecientes al clúster 2. Las consideramos como comunidades ilícitas y no fructíferas, pues el estudiante perteneciente al clúster 1 es quien realiza todo el trabajo y comparte sus resultados con el resto.
- Comunidad 2: Encontramos comunidades más heterogéneas, en donde hay tanto estudiantes del clúster 1 como del clúster 2. También son comunidades consideradas ilícitas, pues los estudiantes con compromiso en el curso comparten sus resultados con el resto del grupo.

- Comunidad 3: Todos los estudiantes de la comunidad hacen parte del clúster 2. Es decir ninguno de los estudiantes realiza un esfuerzo en el curso, pero obtienen buenas calificaciones. Comunidades que se reúnen sólo con el objetivo de conseguir aprobar el curso.

La Tabla 5.3 contiene la cantidad de asociaciones y comunidades encontradas en los 4 cursos con sus dos cohortes.

Tabla 5.3: Asociaciones y comunidades

Curso	Asociación 1 (Clúster 1 y 1)	Asociación 2 (Clúster 1 y 2) (Clúster 1 y 3)	Asociación 3 (Clúster 2 y 2)	Asociación 4 (Clúster 3 y 3)	Comunidad 1	Comunidad 2	Comunidad 3
Lean Startup -2020		1	2	1		1	1
Lean Startup - 2021	2	1		1			
Textos Argumentativos - 2020			3				
Textos Argumentativos - 2021			1		1		1
Danza Folclórica - 2020	3	3	1				
Danza Folclórica - 2021	1					1	
LaTeX - 2020		1	2				
LaTeX - 2021					1		

Parte III

Propuestas desarrolladas

Capítulo 6

Herramienta de seguimiento a las actividades de aprendizaje

Índice

7.1. Analítica de aprendizaje para detectar comportamientos deshonestos	102
7.1.1. Arquitectura general	102
7.1.2. Características del Dashboard	104
7.2. Estudio de caso de evaluación	106
7.2.1. Cursos	106
7.2.2. Instructores	107
7.2.3. Diseño de la evaluación	108
7.3. Análisis cualitativo	109
7.4. Resultados de las entrevistas	110
7.4.1. RQ1. ¿Qué percepción tienen los instructores sobre las métricas utilizadas para detectar comportamientos deshonestos son adecuadas?	112
7.4.2. RQ2. ¿Cuál es el grado de fiabilidad que los instructores suponen del algoritmo utilizado?	113
7.4.3. RQ3. ¿Qué estrategias utilizaron o propusieron los instructores para disuadir la deshonestidad académica?	114
7.4.4. RQ4. ¿Creen los instructores que es posible reducir la deshonestidad académica mediante el uso del tablero presentado?	116

Dentro del contexto de la educación superior los MOOC permiten a las universidades la oportunidad de ampliar su cobertura y llegar a más alumnos, se presentan como el nuevo camino para cubrir la alta demanda de educación superior que existe actualmente. Además, permiten la expansión del conocimiento, la innovación universitaria, la empleabilidad y el desarrollo sostenible de escenarios de aprendizaje, por lo que muchas universidades apuestan por incluirlos en los programas de formación profesional [46, 83, 138].

Sin embargo, las tecnologías masivas presentan algunos inconvenientes para el apoyo de las relaciones personales entre el tutor y los estudiantes que se necesitan en un escenario más tradicional. Una de las principales dificultades de usar SPOC y MPOC es la falta de herramientas que permiten mantener una relación adecuada entre el tutor y el estudiante, pues es complicado que los tutores tengan un trato personalizado con cada uno de los muchos participantes activos en un curso, dificultando las tareas de seguimiento [26, 86], siendo éstas necesarias en cualquier proceso de aprendizaje y aún más cuando se trata de cursos reconocidos académicamente por una institución de educación superior y/o homologables a créditos académicos.

La Universidad del Cauca en el primer periodo de 2016 comenzó la iniciativa por incorporar los MOOC dentro de los programas de pregrado y mediante la implementación de una instancia de la plataforma Open edX, lanzó dos cursos. El curso “Comprensión de textos argumentativos” en donde se contó con 105 estudiantes inscritos y el curso de “Astronomía cotidiana” que tuvo un alcance de participación de 400 estudiantes por lo cual dentro del contexto de la universidad fue considerado como un MPOC.

La incorporación de estos cursos tuvo ciertos desafíos, entre ellos, la necesidad de tener un mayor control sobre los estudiantes y brindarles un acompañamiento adecuado, pues los cursos son válidos como créditos académicos dentro de los programas de formación que ofrece la universidad. Por eso se optó por la construcción de una herramienta que facilitara al docente el seguimiento a las actividades de aprendizaje de los estudiantes en los cursos SPOC y MPOC ofrecidos sobre la instancia de Open edX. Además, se construyó la herramienta como el elemento inicial para el desarrollo del sistema para la captura y recolección de datos que más adelante fueron utilizados para la identificación de conductas con sospecha de deshonestidad académica.

Este capítulo presenta el diseño y la ejecución de la herramienta que se propone como un apoyo para los docentes que utilicen cursos desplegados en plataformas MOOC para ser impartidos por instituciones de educación superior donde se requiera un seguimiento más personalizado de las actividades que realizan los estudiantes. La herramienta ha sido diseñada para realizar tareas de recolección de datos, interpretación y brindar una

realimentación al instructor a partir de una serie de indicadores significativos para apoyar su enseñanza con este tipo de cursos. Concretamente, se presenta el caso donde se ha utilizado la herramienta para realizar el seguimiento al curso de Astronomía cotidiana de la Universidad del Cauca.

6.1. Descripción de la Herramienta de seguimiento

En esta sección se describen los indicadores seleccionados para realizar el seguimiento de los estudiantes, los cuales surgieron de la revisión bibliográfica sobre el tema, y el diseño e implementación de la herramienta web sobre equipos físicos de la Universidad del Cauca, que luego fueron remplazados por equipos que se obtuvieron gracias al financiamiento del proyecto MOOC-Maker.

6.1.1. Indicadores de seguimiento

Aunque el término de seguimiento es utilizado en muchos ámbitos como la industria, medicina, proyectos, procesos de control, etc. No existe una definición generalizada de este concepto, se aplica de acuerdo al área donde es implementado y con frecuencia no se cuenta con un soporte teórico suficiente que lo defina y lo describa, como es el caso del ámbito educativo, mucho menos en la educación en línea (e-learning), MOOC y SPOC. Sin embargo, se entiende que su tarea es supervisar u observar el proceso por medio del cual un estudiante va adquiriendo conocimientos y habilidades, garantizando que se cumplan los objetivos del curso.

En este trabajo se define como seguimiento el proceso destinado a obtener de manera permanente y sistemática, información que permita saber cómo marchan las actividades de aprendizaje de un estudiante teniendo como base los objetivos e indicadores propuestos para un curso [1, 34, 107]. Este seguimiento, no solo debe dar a conocer el progreso del estudiante, sino además, debe permitir brindar una realimentación que dirija el aprendizaje hacia la situación deseada [34, 107].

La Tabla 6.1 presenta los indicadores para el seguimiento de actividades en cursos SPOC que se han identificado en la bibliografía. Algunos de estos indicadores inspiran los que se han tenido en cuenta en el desarrollo de la herramienta. Dichos indicadores se encuentran clasificados según las actividades de aprendizaje presentes en MOOC presentadas en la Subsección 3.3; interacción estudiante – plataforma de aprendizaje, interacción estudiante - evaluaciones, interacción Estudiante – Contenidos e interacción Estudiante - Herramientas de Comunicación. Además, se presentan como indicadores

Tabla 6.1: Indicadores de seguimiento propuestos

Indicador	Trabajo relacionado	Evaluación			Interacción Estudiante - plataforma			Interacción Estudiante - Contenidos			Interacción Estudiante - Herramientas de Comunicación
		Resultados	Intentos	Tiempo tomado	Ingresos al Curso	Ingresos en Secciones y Unidades	Tiempo tomado	Videos vistos	Videos terminados	Tiempo tomado	Foros creados
Compromiso	[54], [52]	X	X	X	X	X	X	X	X	X	X
Progreso	[47], [91], [54], [27], [69]	X			X	X		X	X		
Rendimiento	[27]-[69]	X									
Interés	[54]	X		X				X	X	X	X
Persistencia	[54]	X	X								
Constancia	[54], [28], [19]	X			X	X		X			X
Abandono	[54]	X		X			X			X	
Eficacia	[54]	X	X					X	X		
Capacidad	[54]	X						X	X		X
Participación	[52], [113], [28]							X	X	X	

compuestos, es decir, indicadores que se relacionan con más de una actividad de aprendizaje. Por ejemplo, la persistencia se relaciona con los resultados obtenidos en evaluaciones y la cantidad de intentos para llegar a ese resultado.

Aunque en esta sección no se presenta una definición de los indicadores, se plantea como trabajo futuro el formalizar su definición y validar el impacto que estos puedan generar en cursos tipo SPOC, para ello se hace necesario profundizar en el ámbito pedagógico que permita definirlos de una manera correcta.

6.1.2. Diseño de la herramienta

El diseño de la herramienta para el seguimiento de las actividades de aprendizaje de estudiantes en SPOCs se realizó teniendo en cuenta la estructura y funcionamiento de la plataforma Open edX y el conjunto de indicadores presentados en la Tabla 6.1. En la Figura 6.1 se presenta un diagrama de despliegue del mecanismo construido. En esta vista se muestra todos los componentes físicos del sistema, así como las conexiones que conforman la solución (incluyendo los servicios).

6.1.2.1. App Recolector

Este es uno de los componentes principales del mecanismo, que se encarga de leer constantemente (cada 5 minutos) los registros de la actividad en la plataforma de aprendizaje. Luego guarda los registro en una base de datos para ser accedidos por el componente App Web.

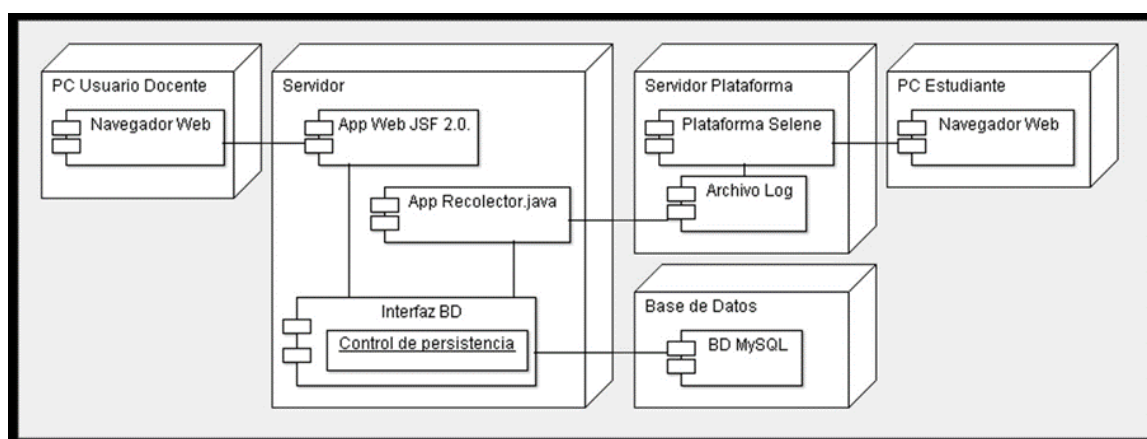


Figura 6.1: Diagrama de despliegue - Herramienta de seguimiento

6.1.2.2. App Web

La App Web es un servicio que permite la consulta y despliegue de los datos a través de un navegador. Esta aplicación se encarga de acceder a la base de datos, sacar estadísticas para organizar información sobre los indicadores de seguimiento y desplegar un servicio web para que los instructores puedan realizar el seguimiento a sus estudiantes.

Este diseño se basó en el mecanismo para la recolección de datos descrito en la Subsección 3.3.2 de extracción de datos para medición de indicadores de interacción con la plataforma de aprendizaje. El diseño completo del mecanismo fue publicado en [63] y la descripción del diseño se encuentra en el Anexo B. Dónde se presenta mediante el modelo “4+1” de Kruchten. Ahí se describen la Vista de escenarios, la vista lógica, la vista de procesos, vista física y vista de despliegue.

6.1.3. Implementación de la herramienta

En esta sección se presenta la implementación de un prototipo basado en la arquitectura anteriormente definida, la cual pretende mostrar las siguientes características:

- Permitir ver los ingresos de los estudiantes al curso, a sus secciones, subsecciones y unidades.
- Permitir ver las interacciones de los estudiantes con los contenidos digitales (vídeos).
- Permitir ver las interacciones de los estudiantes con los foros del curso.

- Permitir ver las interacciones de los estudiantes con las evaluaciones.
- Y por ultimo, permitir ver los indicadores compuestos que sean propuesto en las anteriores secciones.

La instalación de la aplicación se hace mediante la utilización de Jboss, un servidor de aplicaciones gratuito basado en estándares de J2EE. Es una aplicación muy potente, la cual se puede descargar de manera gratuita y es multiplataforma, implementa todo el paquete de servicios J2EE. La aplicación estuvo en funcionamiento hasta el año 2019 sobre el puerto 8005 en el dominio selene.unicauca.edu.co y luego fue reemplazada por la aplicación para la detección y seguimiento de comportamientos de fraude que se presenta en el capítulo 7.

En la Figura 6.2 se muestra una consulta de ingresos en un curso de Astronomía cotidiana desarrollado en el 2016. Se tiene 1523 ingresos en el curso para el mes de marzo, provenientes de un total de 344 estudiantes que estuvieron activos en dicho periodo.

HISTORIAL DE INGRESOS EN EL CURSO: Unicauca + CS001 + 2016-I		
Estudiante	Fecha:	Hora:
Yonay_Alexandra_Agredo	2016-03-31	00:12:38
Julian_Andreas_Garrojal_Sanchez	2016-03-31	00:28:30
Juliana_Andreas_Carrojal_Sanchez	2016-03-31	00:33:24
Enedy_Nicolis_Dobolazaar_Lopez	2016-03-31	00:35:00
Anny_Marcela_Cojiao_Santacruz	2016-03-31	00:37:11
Jana_Fernando_Pena_Delgado	2016-03-31	00:40:19
Loisy_Humberto_Guarana_Serrano	2016-03-31	00:51:46
Julian_Andreas_Hurtado_Gomez	2016-03-31	00:53:56
Nandy_Andrea_Acosta_Dustamante	2016-03-31	01:09:07
Fongier_Alexander_Micropotes_Pizo	2016-03-31	01:19:28

Estudiantes	Ingresos
344	1523

Figura 6.2: Captura de pantalla de la herramienta de seguimiento. Historial de ingresos en el curso de Astronomía cotidiana 2016.

En la figura 6.3 se muestra la captura de pantalla de la herramienta en donde se puede identificar a los estudiantes según su constancia en el curso. Es posible ordenar los resultados por el número de ingresos en el curso, número de ingresos en los contenidos del curso, número de vídeos reproducidos, participación en los foros y número de evaluaciones realizadas.

CURSO: Unicauca+CS001+2016-I

Estudiante	Ingresos al Curso ^	Ingresos a los contenidos ⇅	Reproducciones de Videos ⇅	Participación en Foros ⇅	Intentos en evaluación ⇅
Andrés_Gatton_Medina_Morales	1	3	0	2	0
Diego_Elipo_Villalba	1	4	10	0	0
Gabriela_Marcela_Salazar	1	5	21	0	0
Jhony_Guilia_Beltrán_Gloria	1	6	12	0	0
Juan_Carlos_Díaz_González	1	3	3	0	0
Maria_Silvana_Garcenas	1	9	8	1	0
Sofía_Dé	1	0	0	1	0
Victor_Ceron_Narvaez	1	2	1	0	0
Carlos_Fernán_Ortíz_Alegria	2	6	8	0	0
Geotisa_Fernando_Lopez_Ceron	2	8	5	0	0

Figura 6.3: Consulta de la constancia en el curso de Astronomía cotidiana 2016

6.2. Desarrollo de caso de estudio

Para realizar las pruebas de uso del mecanismo se utilizó el curso virtual de Astronomía Cotidiana ofrecido en el 2016 a estudiantes de la Universidad del Cauca. Curso descrito en la sección 3.2 y mediante el cual obtienen créditos dentro de los programas de formación de la universidad.

6.2.1. Resultados de seguimiento

A continuación se describen mediante ejemplos las diferentes consultas que permitieron al docente obtener información del seguimiento a las actividades de los estudiantes.

6.2.1.1. Ingresos en el curso y sus contenidos

El menú de navegación se encuentra ubicado en la parte izquierda de la página; dicho menú permite el acceso a los diferentes indicadores de seguimiento de la herramienta. Dentro de la pestaña ingreso (Figura 6.4) se despliegan dos ítems que permiten acceder a los ingresos al curso y a los ingresos en los contenidos del curso.

Una de las inquietudes de los instructores fue conocer los ingresos de los estudiantes los primeros días del curso, para saber si el procedimiento definido para inscripción de estudiantes a la plataforma era el correcto y si los estudiantes habían recibido las comunicaciones orientadoras al comienzo de la asignatura. En la Figura 6.5, se muestran los ingresos al curso para el mes de febrero, mes en que dio inicio el curso de astronomía.

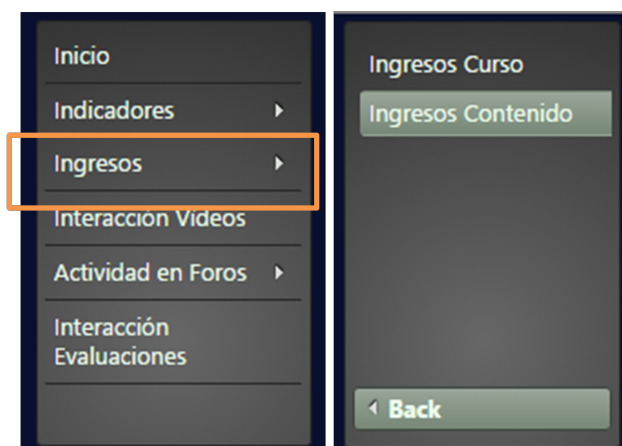


Figura 6.4: Menú de navegación herramienta de seguimiento

Cómo se observa, se registraron 1508 ingresos en el curso y el número de estudiantes fue 401.

HISTORIAL DE INGRESOS EN EL CURSO: Unicauca+CS001+2016-I		
Estudiante	Fecha:	Hora:
Diana_Isabel_IbanezPino	2016-02-29	00:03:54
Diana_Isabel_IbanezPino	2016-02-29	00:12:14
Quiana_Isabel_IbanezPino	2016-02-29	00:20:05
Martha_Vanessa_Giluentes	2016-02-29	00:26:06
Eduardo_Alejandro_Velez_Mosquera	2016-02-29	00:28:19
Isabel_David_Maldina_Sandoval	2016-02-29	00:53:28
Alexis_Javier_Manrique_Zuniga	2016-02-29	01:04:39
Wesnei_L_Alegría_Riquitos	2016-02-29	01:32:05
Alvaro_Javier_Manrique_Zuniga	2016-02-29	01:35:08
Dejaravilla	2016-02-29	01:45:00
Estudiantes		Ingresos
401		1508

Figura 6.5: Ingresos al curso piloto mes de febrero

El panel de la derecha permite la consulta seleccionando el curso y el mes, además de poder escoger un estudiante específico. La tabla del centro, contiene opciones que permiten filtrar los resultados por estudiante, fecha y hora.

La Figura 6.6, muestra el historial de ingresos obtenidos para el mes de febrero. En la semana 1 (día 15 al 19) es donde se presentan mayores ingresos, debido a la expectativa del curso. Se ve que para fechas posteriores los ingresos disminuyen y se mantienen. De

la misma forma la herramienta permite ver los contenidos que se han consultado para fechas específicas.



Figura 6.6: Historial de ingresos al curso en el mes de febrero

6.2.1.2. Actividad en vídeos

En cuanto a las interacciones de los estudiantes con los vídeos la plataforma permite ver la información de cada vídeo y también de cada estudiante. La pestaña “Interacción en Vídeos” en el menú de navegación es el acceso a las dichas consultas. En la Figura 6.7 se presenta la consulta hecha al finalizar el curso para el video “GQQg83MSuvY” (identificador del video en Youtube), pero además se puede ver la sección y unidad de ubicación del vídeo dentro del curso.

Se puede ver que el vídeo obtuvo 168 interacciones, lo que corresponde a que fue reproducido desde cero 113 veces (“Vistos”) y en 55 ocasiones lo miraron completamente (“Stops”). Las interacciones fueron realizadas por 109 estudiantes de los cuales 51 reprodujeron el vídeo hasta el final.

Además, es posible la consulta de las interacciones de los vídeos por cada estudiante. En la Figura 6.8 se muestra un ejemplo donde un estudiante, registró 226 interacciones que corresponden a 127 reproducciones de las cuales 99 veces llegó hasta el final del vídeo. Miró 106 vídeos diferentes, de los cuales 90 fueron observados en su totalidad. También la herramienta entrega la lista de los diferentes vídeos vistos por el estudiante. La consulta fue realizada al finalizar el curso.

CURSO: Unicauca - CS001 - 2016-I					
Estudiante	Video:	Acción:	Fecha:	Hora:	
Yerson_Mendoza_Hernandez_Lasso	GQqg83MSuVY	stop_video	2016-04-07	14:12:21	Unidad
Hermano_Castillo_José	GQqg83MSuVY	stop_video	2016-04-02	00:21:01	Unidad
Angellin_Martin_Nieto_Ley	GQqg83MSuVY	play_video	2016-04-02	04:49:30	Unidad
Angellin_Martin_Nieto_Ley	GQqg83MSuVY	play_video	2016-04-02	05:53:25	Unidad
Ludwing_Vivale_Balazano	GQqg83MSuVY	play_video	2016-04-01	00:57:09	Unidad
Jhony_Josue_Colada_Ayala	GQqg83MSuVY	play_video	2016-04-01	01:08:41	Unidad
Argente_Ivana_Paulinas	GQqg83MSuVY	play_video	2016-04-01	01:54:46	Unidad
Cristóbal_González_Bolanos	GQqg83MSuVY	play_video	2016-04-01	02:49:25	Unidad
Dany_Alejandro_Martínez	GQqg83MSuVY	play_video	2016-04-01	02:51:33	Unidad
Diego_Alejandro_Martínez	GQqg83MSuVY	stop_video	2016-04-01	02:59:45	Unidad

Estudiante	Interacciones	Vistos	# Plays	Estudiantes	# Stops	Terminados
109	168	1	113	109	55	51

Figura 6.7: Consulta interacciones de los estudiantes con los vídeos del curso

6.2.1.3. Actividad en Foros

En cuanto a los foros se puede obtener el total de las participaciones de los estudiantes, dichas participaciones corresponden a creaciones de temas nuevos (“hilos”), respuestas en dichos temas y comentarios de las mismas del curso en general, además, de la misma forma que para los otros indicadores, es posible hacer consultas por cada estudiante. En la Figura 6.9 se muestra los datos de las participaciones del curso piloto en donde se contó con la participación de 330 estudiantes. Consulta realizada al final del curso.

6.2.1.4. Interacciones con Evaluaciones

Conocer el número de estudiantes que realizaron un determinado examen es de vital importancia en cualquier curso, más aún si el curso es en modalidad no presencial. El prototipo obtiene la información de todas las interacciones de los estudiantes con las evaluaciones programadas en la plataforma de aprendizaje, de esta forma es posible conocer además del número de estudiantes que participaron, conocer la hora y la fecha de la interacción, y el número de repeticiones realizadas.

La característica de la cantidad de repeticiones se agregó a la herramienta debido a que Open edX permite configurar en los exámenes el número de intentos, es decir, el número de veces que un estudiante puede repetir un examen o ejercicio.

En la Figura 6.10 se muestran los resultados obtenidos para el examen final del curso piloto. Se puede apreciar como de 403 estudiantes inscritos, 358 presentaron el examen final del curso, también se muestran el número de participaciones y cuantas veces fue

CURSO: Unicauca+ CS001+ 2016-I					
<div style="text-align: center;"> << < 1 2 3 > >> </div>					
Estudiante	Video:	Acción:	Fecha:	Hora:	
Adrian_Felipe_Vargas_Arias	9vk6Kn2VXhQ	play_video	2016-03-13	01:23:30	Unidad
Adrian_Felipe_Vargas_Arias	9vk6Kn2VXhQ	play_video	2016-03-13	01:23:38	Unidad
Adrian_Felipe_Vargas_Arias	9vk6Kn2VXhQ	stop_video	2016-03-13	01:26:06	Unidad
Adrian_Felipe_Vargas_Arias	iFeBJULpkA	play_video	2016-03-13	01:26:07	Unidad
Adrian_Felipe_Vargas_Arias	LoSOZ9r_UvU	play_video	2016-03-13	01:33:58	Unidad
Adrian_Felipe_Vargas_Arias	xluMl0YWm4E	play_video	2016-03-13	01:34:03	Unidad
Adrian_Felipe_Vargas_Arias	LoSOZ9r_UvU	play_video	2016-03-13	01:53:34	Unidad
Adrian_Felipe_Vargas_Arias	LoSOZ9r_UvU	stop_video	2016-03-13	02:04:34	Unidad
Adrian_Felipe_Vargas_Arias	Z2yCHMBFNYM	play_video	2016-03-13	02:35:51	Unidad
Adrian_Felipe_Vargas_Arias	xluMl0YWm4E	stop_video	2016-03-13	02:37:06	Unidad

Estudiante	Interacciones	Vistos	# Plays	Estudiantes	# Stops	Terminados
Adrian_Felipe_Vargas_Arias	226	106	127	1	99	90

Figura 6.8: Consulta de las interacciones de un estudiante específico con vídeos

repetido el examen.

CURSO: Unicauca+ CS001+ 2016-I				
<div style="text-align: center;"> << < 1 2 3 4 5 6 7 8 9 10 > >> </div>				
Estudiante	Foro:	Fecha:	Hora:	
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	00:22:59	Aplicaciones prácticas
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	00:28:44	Recurso de astronomía
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	01:21:57	Relación de programación
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	03:06:39	CALENTAMIENTO
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	05:00:03	, \u00bfqu\u00e9 es la astronomía?
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	05:11:15	Calentamiento general
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	05:17:43	ARTE EN LA ASTRONOMÍA
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	05:51:48	astronomía en imágenes
Deniel_Andrea_Yañez_Auriz	edx.forum.thread.created	2016-06-16	15:34:31	peque\u00f1a serie de imágenes
Lisbeth_Vandera_Vergara_Zumaya	edx.forum.thread.created	2016-06-16	15:57:02	TURISMO ASTRONÓMICO

Estudiantes	Hilos	Respuestas	Comentarios	Participaciones
330	1340	1005	65	2410

Figura 6.9: Consulta de las interacciones en el foro del curso

CURSO: Unicauca+CS001+2016-I				
Estudiante		Fecha:	Hora:	
<input type="text"/>		<input type="text"/>	<input type="text"/>	
Jorge_Eduardo_Hernandez		2016-06-16	18:05:06	
Jose_Ignacio_Romero_Urbano		2016-06-16	18:14:29	
Rafael_Alberto_Guerra_Patillas		2016-06-16	18:16:13	
Eduardo_Jose_Cadizuez_Hurtado		2016-06-16	18:16:40	
Javier_Cabrera_Zambrano_Mendez		2016-06-16	18:16:47	
Sandra_Luzmarido_Esquivel_Noguera		2016-06-16	18:18:18	
David_Eduardo_Urrea_Galvis		2016-06-16	18:19:03	
Werner_Alejo_Ricoes		2016-06-16	18:19:17	
Mario_Alejandra_Garcia		2016-06-16	18:20:28	
Janry_Ezquerro_Gil_Cifuentes		2016-06-16	18:21:11	
Estudiante	Intentos	Repeticiones	Unidades realizadas	Estudiantes
	400	42	1	358

Figura 6.10: Consulta de las interacciones en el examen final del curso

6.2.1.5. Indicadores compuestos

Los indicadores compuestos son la información más relevante para el docente. En la pestaña indicadores se encuentran tres ítems, Figura 6.11.

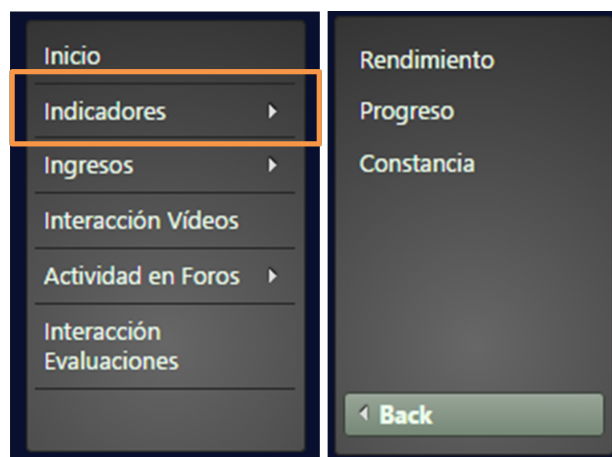


Figura 6.11: Menú navegación Indicadores compuestos

- **Rendimiento:** Muestra las evaluaciones obtenidas por los estudiantes.
- **Progreso:** Muestra la cantidad de actividades realizadas y el porcentaje avanzado.
- **Constancia:** Muestra la actividad del estudiante en la plataforma de aprendizaje.

En la Figura 6.12 se muestra la consulta de rendimiento de los estudiantes de curso piloto. La información es presentada en forma de tabla en donde se tiene la lista de los estudiantes y las calificaciones obtenidas en los exámenes. Los botones en los identificadores de los exámenes permiten ordenar los resultados y de esta forma conocer quiénes son los estudiantes que tienen alto y bajo rendimiento. Esta es una de las características que diferencia la herramienta de seguimiento de Insights de Edx y Analyse, con dichas herramientas es más difícil identificar los grupos de bajo y alto rendimiento.

CURSO: Unicauca+CS001+2016-I					
◀ ◀◀ 1 2 3 4 5 6 7 8 9 10 ▶▶ ▶					
Estudiante	Nota Final ▲	Nota 1 ◆	Nota 2 ◆	Nota 3 ◆	Nota 4 ◆
Peyer_Ignacio_Risarcos_Bogota	0	0	0	0	0
Sofo_De	0	0	0	0	0
Susy_Juliana_Rosero_Arias	0	0	0	0	0
Yuley_Catherine_Bezana	0	0	0	0	0
Brenda_Gaardo_Elmer_Mangas	8	21	0	0	0
David_Eduardo_Umea_Galvis	12	0	0	38	0
Daniel_Felipe_Rendon_Sorrito	15	0	0	50	0
Johan_Fabian_Ulve_Beltrano	20	0	57	0	0
Maria_Alejandra_Clarke_Gortes	20	57	0	0	0
Gesar_Otiven_Ancane_Pino	20	0	57	0	0

Figura 6.12: Consulta del rendimiento de los estudiantes del curso

Un ejemplo para los resultados del indicador de progreso se muestra en la Figura 6.13. Se presentan tres indicadores antes mencionados, la cantidad de exámenes realizados por los estudiantes, la cantidad de vídeos vistos y las unidades del contenido visitadas. Además por cada uno se presenta el porcentaje de avance, calculado con base en el total de exámenes, vídeos y contenidos.

También es posible ordenar los resultados y así poder identificar a los estudiantes con mayores y menores avances, además de dar una idea de cuál es el avance del curso en general. En la Figura 6.14 se muestra una consulta para la constancia de los estudiantes del curso piloto. En este caso la información presentada hace referencia a la actividad del estudiante en la plataforma y que permitirá saber que tan comprometido se encuentra el estudiante con el curso.

La información presentada es un consolidado de las actividades del estudiante, total de ingresos al curso y contenidos, total de interacciones con vídeos, participación en los foros y el número de intentos realizado en las evaluaciones.

La forma en cómo se pueden ordenar los resultados de las consultas, hace que la he-

herramienta sea muy útil para SPOC y MPOC donde la cantidad de estudiantes es limitada pero mayor que en cursos tradicionales y el docente no puede realizar seguimiento a las actividades por sí mismo.

Que las otras herramientas sean directamente destinadas a cursos MOOC hace que no incorporen este tipo de presentación de datos, pues al ser cursos demasiados grandes el docente ya no puede tener una relación con los estudiantes. Pero en cursos tipo SPOC y MPOC además de ser posible dicha relación es necesaria y el uso de la herramienta facilita al docente las tareas de acompañamiento al estudiante.

6.3. Evaluación de usabilidad y utilidad

En esta sección se presenta una evaluación de la herramienta que tiene como objetivo entender su utilidad y usabilidad. Para esto se desarrolló una prueba siguiendo una metodología heurística [50], un método de evaluación por inspección que es llevado a cabo por un grupo de evaluadores a partir de unos principios (denominados “heurísticos”) previamente establecidos, que buscan medir la calidad de cualquier sistema en su relación para ser aprendido y utilizado por un determinado grupo de usuarios en un contexto dado [50].

La evaluación se realizó con 13 profesores de educación superior: 8 miembros de la Pontificia Universidad Católica de Chile (PUC) y 5 docentes de la Universidad del Cauca (Unicauca). La prueba contó con tres secciones: usabilidad, utilidad y opinión. Se agregó la tercera sección (opinión), para poder obtener las opiniones del grupo, como una

Estudiante	Exámenes Realizados	Porcentaje	Videos Reproducidos	Porcentaje	Unidades Vistas	Porcentaje
Adrián_Itálopez_Virgas_Artes	4	100 %	106	77 %	23	100 %
Adriana Acebo_Virgos_Bautista	4	100 %	12	8 %	13	56 %
Adriana_Paula_Lectamo_Campo	4	100 %	16	11 %	14	60 %
Adriana_Paula	1	25 %	0	0 %	12	52 %
Aléxan_Steven_Hoyos_Chilangana	3	75 %	41	29 %	23	100 %
Alejandra_Díaz	2	50 %	30	21 %	8	34 %
Alisón_Guerrero_Manoz	4	100 %	6	4 %	14	60 %
Alizara_Carolina_Botinos_Abril	3	75 %	25	18 %	14	60 %
Alvarez_Javier_Mc.Nique_Zuriga	3	75 %	7	5 %	7	30 %
Amanda_Beatriz_Naguera_Burbano	0	0 %	9	6 %	10	43 %

Figura 6.13: Consulta del progreso de los estudiantes del curso

CURSO: Unicauca+CS001+2016-I					
1 2 3 4 5 6 7 8 9 10					
Estudiante	Ingresos al Curso	Ingresos a los contenidos	Reproducciones de Videos	Participación en Foros	Intentos en evaluación
Majalante	232	468	23	67	2
Daguanmálica	127	144	66	14	2
Laverde, Valeria, Hernández, Lasso	117	256	196	20	9
Lucero, Cruz, Lopez	96	200	180	7	4
Lucía, Elizabeth	75	124	160	13	5
Jokyr, Marcela, Sambony, Ortega	68	115	196	12	5
Lucy, María, Araujo, Goydelo	63	103	94	3	5
Natália, Brígida, Orihentes	62	119	162	9	11
Sonia, Murgoth, Tondroy, Mabeoy	62	129	259	18	3
Lery, Humberto, Guzman, Serrano	61	113	67	15	5

Figura 6.14: Consulta de la constancia de los estudiantes del curso

realimentación que sirva para realizar cambios y mejoras a la herramienta.

Para la evaluación se desarrolló un cuestionario basado en [50] compuesto de 22 preguntas organizadas en tres secciones: una para evaluar usabilidad de la herramienta, otra para analizar su utilidad y otra parte abierta para que los profesores ofrecieran su opinión acerca de la herramienta y sus posibles mejoras. Dicho cuestionario fue validado antes de su ejecución por el grupo de profesores de la PUC.

El cuestionario estaba acompañado de una serie de instrucciones que guiaban a los evaluadores a través de las distintas funcionalidades de la herramienta. Las instrucciones se agregaron con el fin de permitirles a los evaluadores, familiarizarse con la herramienta.

El cuestionario se trabajó mediante afirmaciones, el rango de cada afirmación se situaba de 1 a 5, siendo 1 el valor para “Muy poco de acuerdo” y 5 para los participantes que estuvieran “Muy de acuerdo” con la afirmación presentada. Se incluyeron en total 20 preguntas cerradas y 2 preguntas abiertas (Anexo A).

6.3.1. Resultados en cuanto a Usabilidad

En la Tabla 6.2, se presentan los resultados obtenidos de aplicar el cuestionario en el grupo con los 13 evaluadores participantes. En la parte izquierda se encuentran las afirmaciones utilizadas para evaluar la usabilidad de la herramienta, en la parte central se encuentran el promedio de los resultados que se obtuvieron por institución y en la parte derecha se encuentra el promedio total obtenido de cada afirmación.

Como se puede observar, la mayoría de las afirmaciones obtuvieron una buena res-

Tabla 6.2: Resultados sobre la usabilidad de la herramienta

Afirmaciones de Usabilidad	Evaluadores		Promedio
	Unicauca	PUC	
1 - La herramienta es fácil de utilizar	4,3	4,0	4,2
2 - Me gusta la apariencia de la herramienta	3,5	2,9	3,2
3 - Considero que gráficamente el sitio está equilibrado	3,7	3,2	3,4
4 - El lenguaje utilizado es comprensible	4,5	4,3	4,4
5 - El objetivo de cada característica de la interfaz es claro	4,0	4,0	4,0
6 - Los resultados mostrados en pantalla son fáciles de interpretar	4,0	3,9	4,0
7 - El contenido del menú es suficientemente descriptivo de la información que se ofrece	4,0	4,0	4,0
8 - Es fácil realizar una consulta	4,3	4,2	4,3
9 - Los elementos dentro de las páginas, permiten saber dónde me encuentro dentro del sitio y cómo navegar en él	4,3	4,3	4,3
10 - Los resultados de las consultas aparecen enseguida	4,8	4,7	4,8

puesta por parte de los evaluadores. Destacándose entre ellas la velocidad con que aparecen los resultados. En general la evaluación en cuanto a usabilidad del sitio fue buena. Los usuarios destacan la facilidad de uso, que la herramienta permite que se ingrese y se navegue en el sitio sin dificultades, que la velocidad de entrega de resultados es muy buena y que el lenguaje utilizado para describir los contenidos es adecuado. Características que ayudan a que la herramienta tenga una mayor aceptación.

Sin embargo, los resultados del cuestionario (afirmaciones 2 y 3), indican que la interfaz gráfica de la herramienta debe mejorar, pues a los evaluadores no les gusta la apariencia y consideran que gráficamente el sitio no está equilibrado. Hecho que perjudica de cierta manera que los resultados mostrados sean fácilmente interpretados. Esto sugiere que en próximas versiones se tenga más en cuenta la apariencia visual.

6.3.2. Resultados en cuanto a Utilidad

Para tener una idea acerca de la utilidad de la herramienta, en el cuestionario se establecieron 5 afirmaciones, que iban encaminadas a saber si el objetivo de la herramienta es claro y saber cuál es el nivel de facilidad de la herramienta para permitir reconocer el progreso, constancia, persistencia y rendimiento de los estudiantes. En la Tabla 6.3, se observan los resultados obtenidos.

Con estas afirmaciones, se quería evaluar la utilidad de la herramienta para dar a conocer indicadores como el rendimiento, la persistencia, la constancia y el progreso de los estudiantes. El resultado que más se destaca es la posibilidad para dar a conocer el progreso de un estudiante específico, pues la herramienta permite conocer el progreso

Tabla 6.3: Resultados sobre la utilidad de la herramienta

Afirmaciones de Utilidad	Evaluadores		Promedio
	Unicauca	PUC	
1 - Queda claro cuál es el propósito de la herramienta	4,3	4,5	4,4
2 - Es fácil identificar los estudiantes con mejor rendimiento	4,2	4,0	4,1
3 - Es posible reconocer los estudiantes más persistentes en un curso	4,2	4,2	4,2
4 - Es posible conocer el progreso de un estudiante específico	4,5	4,5	4,5
5 - Es posible conocer a los estudiantes con menos constancia en un curso	4,3	4,2	4,3

Tabla 6.4: Opinión sobre la herramienta

Afirmaciones de Opinión	Evaluadores		Promedio
	Unicauca	PUC	
1 - Es útil para realizar seguimiento a las actividades de aprendizaje de los estudiantes en cursos SPOC y MPOC	3,7	4,2	3,9
2 - Puede aportar información útil en cursos reconocidos académicamente por una institución de educación y/o homologables a créditos académicos	3,7	4,2	3,9
3 - Brinda la información necesaria para el seguimiento de los estudiantes de SPOC y MPOC	4,0	4,2	4,1
4 - Utilizaría la herramienta	4,2	4,2	4,2
5 - Recomendaría el uso la herramienta	4,2	4,2	4,2

individual de cada estudiante, el progreso en general del curso y conocer cuáles de los estudiantes tienen un mayor o menor progreso. Los evaluadores están de acuerdo en que el objetivo de la herramienta es claro y que los indicadores en ella mostrados son de gran utilidad para apoyar a los docentes en cuanto al seguimiento en SPOCs y MPOCs.

6.3.3. Resultados de Opinión

Para esta sección en el cuestionario se incluyeron 5 afirmaciones entre las cuales se tuvieron en cuenta dos afirmaciones para poder conocer si se utilizaría y recomendaría la herramienta. Además, se propusieron dos preguntas con las cuales se quería captar lo destacado de la herramienta, como también, las sugerencias en cuanto a mejoras futuras.

En la Tabla 6.4, se muestran los resultados obtenidos. Según estos resultados, los evaluadores utilizarían y recomendarían la herramienta. Además, opinan que la herramienta brinda la información necesaria para el seguimiento de los estudiantes en SPOC y MPOC. Sin embargo, no están completamente de acuerdo con que pueda aportar información útil en cursos reconocidos académicamente por una institución de educación superior.

Las dos preguntas abiertas incorporadas al cuestionario fueron: ¿Qué destaca de la herramienta? Y ¿Qué se puede mejorar de la herramienta?

Entre las respuestas, los evaluadores destacan que la herramienta logra presentar de

forma sencilla, eficaz y clara la información de las actividades realizadas por los estudiantes durante el desarrollo de un curso, con comentarios como: “La herramienta logra presentar de forma sencilla la información de las actividades realizadas por los estudiantes durante el desarrollo de un curso” y “La herramienta otorga de manera eficaz información sobre el progreso de los estudiantes”. También, manifiestan que es de gran utilidad en contextos en los que se usa un MOOC como complemento a una clase presencial.

Por otra parte, se destaca que la herramienta ofrece indicadores predefinidos que son de gran apoyo para MOOC, en especial para el caso de los SPOC y MPOC, en donde se requiere que haya un Seguimiento de las actividades de los estudiantes y que tiene mucho potencial.

Sin embargo, se propone mejorar en diseño, en la visualización de la información y en la cantidad de información proporcionada al docente, incorporando gráficas que faciliten la interpretación de los resultados. También, expresan que: “quizás sería bueno ofrecer indicadores libres para que el usuario genere sus propias consultas al sistema”. Se plantea en futuros desarrollos tener en cuenta todas las recomendaciones resultado de la evaluación preliminar, además, se busca que la herramienta pueda hacer el seguimiento automático de las actividades y proporcionar información relevante al instructor, sin depender de las consultas del docente en la herramienta.

Capítulo 7

Dashboard: Herramienta para la identificación y seguimiento

Índice

8.1. Algoritmo para la detección de comportamientos con sospecha de deshonestidad académica	119
8.2. Caracterización de las colaboraciones encontradas	122
8.3. Herramienta de seguimiento	124
8.4. Evaluación de la herramienta de detección de conductas con sospecha de deshonestidad académica	125

La preocupación constante para permitir que los MOOCs, SPOCs y otros cursos en línea sean reconocidos académicamente debido a la deshonestidad académica presente [2,64], han motivado el desarrollo de mecanismos para la identificación de estos comportamientos. Además, los actuales entornos virtuales de aprendizaje no cuentan con las herramientas tecnológicas necesarias para apoyar a los instructores en la identificación de la deshonestidad de los estudiantes [103,105].

Por ello, en este estudio, se trabajó en la mejora de este tema, mediante el desarrollo de un tablero de control que contiene las métricas para la identificación y seguimiento de que fueron presentadas en en la Sección 3.3 para identificar a los alumnos sospechosos de deshonestidad académica. El tablero de mandos está desarrollado para mostrar las métricas de forma comprensible para los instructores y que éstos puedan también intervenir durante el desarrollo del curso. Más concretamente, nuestro cuadro de mando incluye la métrica de similitud en las respuestas de los exámenes y la métrica de la cercanía temporal a la hora de entregar las respuestas de los exámenes. Con esto, presentamos

visualizaciones sobre estas métricas y otros indicadores de comportamiento.

Lo que se busca resolver con este capítulo, es que si ponemos estas métricas a disposición de los instructores a través de un panel de control, ¿pueden los docentes o instructores a cargo de los cursos, entender las métricas y utilizarlas para intervenir y disminuir la deshonestidad académica? La mayor parte de la literatura anterior que aborda la evaluación de los cuadros de mando de analítica del aprendizaje se ha centrado en otros aspectos, como el seguimiento de los estudiantes [62], las métricas de autorregulación [94, 139] o las métricas de prevención del abandono escolar [114]. Sin embargo, en el contexto de la deshonestidad académica, creemos que esta es la primera evaluación de este tipo.

Por lo tanto, este capítulo presenta una evaluación de un cuadro de mando que incorpora métricas para detectar la deshonestidad académica con cuatro instructores. El contexto de la investigación es dentro de los SPOCs con reconocimiento académico desarrollados en la Universidad del Cauca. La evaluación valorará el potencial de la herramienta para disminuir la deshonestidad académica a partir del estudio de caso con los instructores y de una entrevista semiestructurada.

7.1. Analítica de aprendizaje para detectar comportamientos deshonestos

Esta sección describe nuestro Dashboard propuesto, teniendo en cuenta cómo se implementó, qué métricas se utilizan para la detección de la deshonestidad académica, y las características incluidas.

7.1.1. Arquitectura general

El propósito general del dashboard era permitir a los profesores saber qué estudiantes estaban compartiendo potencialmente las respuestas de los exámenes y apoyar la monitorización de dichas actividades, facilitando la detección de comportamientos deshonestos. La arquitectura general del dashboard se muestra en la figura 7.1.

Nuestro dashboard tiene dos componentes principales, un módulo de procesamiento de datos y un módulo de visualizaciones. Dentro del módulo de procesamiento de datos, hay un submódulo desarrollado en Python que se encarga de recuperar los registros de las interacciones de los alumnos con la plataforma casi en tiempo real, con apenas un retraso de unos cinco minutos, para luego almacenar esos registros en una base de datos

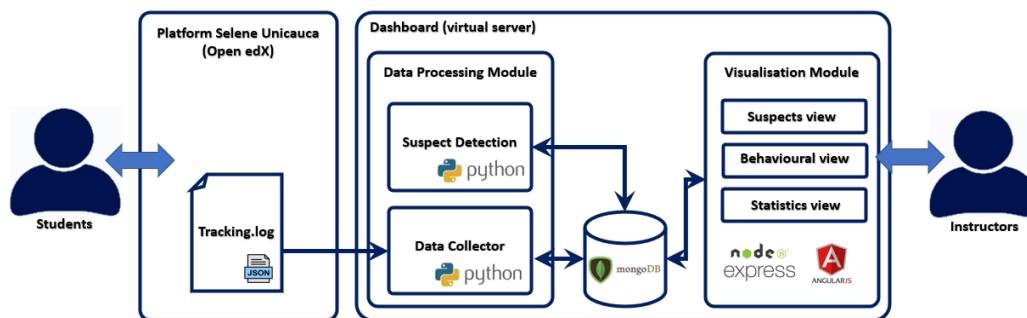


Figura 7.1: Dashboard architecture for the detection of academically dishonesty behaviours.

MongoDB. Estos registros se están recuperando continuamente del archivo en formato JSON llamado `tracking.log` alojado en la plataforma Selene, que es un archivo de registro estándar que todas las instancias de Open edX generan para registrar todas las acciones de los alumnos a través de la plataforma de aprendizaje. Por otro lado, todas las métricas que se computan, incluyendo las de detección de fraude, son procesadas por un submódulo Python diferente. Las métricas se calculan también cada cinco minutos, justo después de que se hayan recuperado los nuevos registros. El objetivo de este enfoque es facilitar al profesor la posibilidad de controlar el progreso de los estudiantes casi en tiempo real con sólo cinco minutos de retraso, de modo que el tablero de instrumentos también se puede utilizar durante un curso real. Esta es una de las principales mejoras con respecto a estudios anteriores retrospectivos que utilizaban algoritmos para detectar la deshonestidad académica, en lugar de apoyar la detección en vivo para los instructores.

El módulo de visualizaciones está desarrollado con NodeJS, Express y Angular, que muestra el dashboard como una aplicación web. Este módulo se encarga de acceder a la base de datos y generar la interfaz para que los instructores interactúen con las métricas mediante el uso de visualizaciones. Tiene tres componentes, un componente que gestiona los datos para las visualizaciones, un componente que se encarga de mostrar las diferentes vistas, y un componente que implementa la lógica de negocio de la aplicación.

De este modo, nuestro cuadro de mandos hace un seguimiento de los estudiantes, analiza sus registros para detectar a los estudiantes sospechosos de fraude en tiempo real y muestra las visualizaciones a través de una aplicación web que se entrega a los instructores del curso Selene.

7.1.2. Características del Dashboard

El cuadro de mandos se desplegó en el mismo servidor y dominio donde está instalada la plataforma Selene. El acceso a la aplicación está controlado por el administrador de la plataforma, que es el responsable de crear las cuentas de los instructores.

El dashboard tiene tres vistas:

- Vista de sospechosos
- Vista de estadísticas
- Vista de comportamiento

7.1.2.1. Vista de sospechosos

La figura 7.2 muestra una captura de pantalla de la vista de sospechosos, que permite al instructor consultar los alumnos sospechosos de deshonestidad académica. Muestra un dendrograma que agrupa a los alumnos según la similitud de sus respuestas y un dendrograma que los agrupa según sus marcas de tiempo de envío. El instructor también tiene la opción de comparar dos estudiantes seleccionados, mostrando las preguntas de la prueba, la interferencia en esas respuestas entre los dos estudiantes y la distancia en las marcas de tiempo de envío. Los instructores pueden utilizar estas métricas previamente validadas en entornos de investigación para detectar a aquellos estudiantes que podrían haber cometido comportamientos deshonestos con la ayuda de otros compañeros.

Esta vista muestra todos los alumnos que son detectados como sospechosos por el algoritmo. Además, es posible consultar los alumnos sospechosos por cada curso o por examen por separado. En esta vista, sólo se muestran los alumnos detectados con comportamientos fraudulentos sospechosos. En concreto, aquellos que tienen respuestas muy similares (más del 95 %) o que presentaron los exámenes muy cerrados en el tiempo (menos de 20 segundos)

7.1.2.2. Vista de seguimiento

La segunda vista es la de seguimiento, la figura 7.3 muestra las estadísticas generales donde el instructor puede visualizar el número de interacciones con los contenidos del curso y con los vídeos, el número de comentarios publicados en los foros, y el número de exámenes presentados. También muestra un gráfico del historial de interacciones, donde es posible ver cómo se han comportado los alumnos a lo largo del curso. Además, es posible consultar las estadísticas de cada alumno por separado.

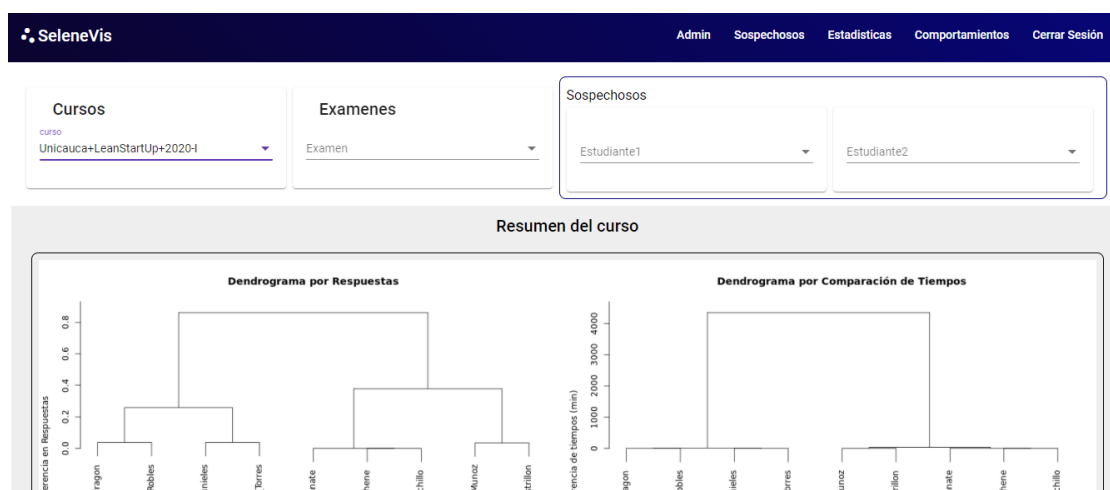


Figura 7.2: Screenshot of the view of students suspected of academic dishonesty in the tool

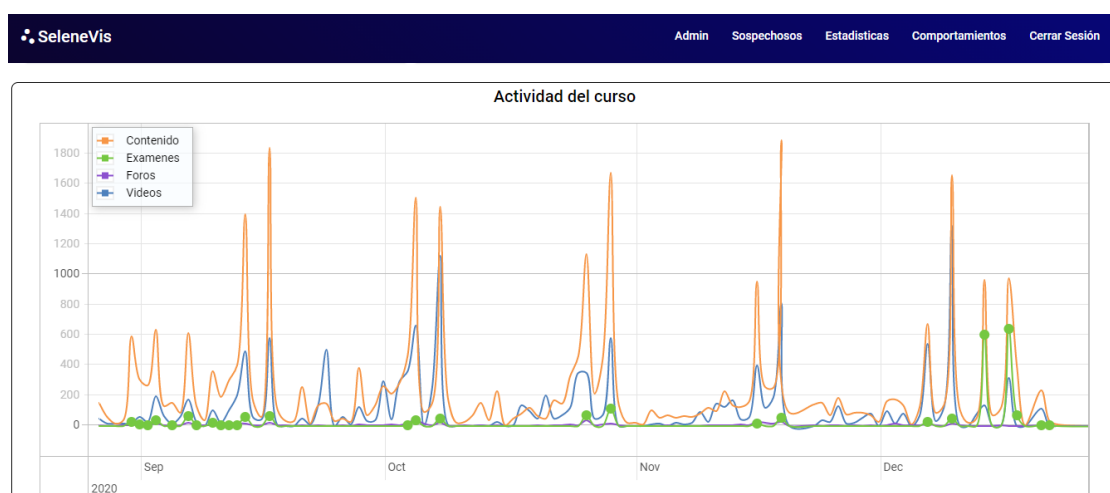


Figura 7.3: Screenshot of the tool's statistics view.

7.1.2.3. Vista de comportamientos

Por último, la Figura 7.4 muestra la captura de pantalla de la vista de comportamientos. Esta vista muestra cómo se ha comportado un alumno en un día concreto o en una sesión determinada. El comportamiento se representa como un gráfico que muestra la secuencia de interacciones que un alumno ha realizado en esa sesión concreta. Estas dos últimas vistas pueden ser utilizadas por los instructores para revisar las interacciones con el curso de aquellos alumnos que están etiquetados en la primera vista como sospechosos de comportamientos deshonestos.

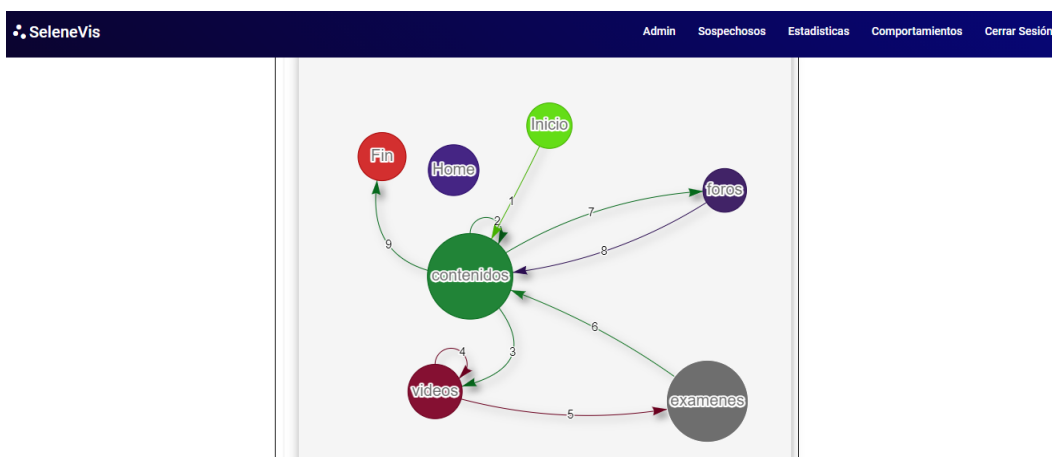


Figura 7.4: Screenshot of the tool's behaviours view.

7.2. Estudio de caso de evaluación

Esta sección está dividida en dos apartados. Por un lado muestra los cursos y los instructores que están a cargo de los cursos con los que forman los casos de estudio. Después presenta el método utilizado para la evaluación del cuadro de mando.

7.2.1. Cursos

Los cursos utilizados en este estudio se denominan: “Introducción al emprendimiento Lean Startup” (Curso Lean Startup), “Comprensión de textos argumentativos” (Curso de textos), “La danza folclórica como patrimonio cultural” (Curso de danza), e “Introducción a la edición de textos científicos y literarios con LaTeX” de la Universidad del Cauca ofrecidos a través de la plataforma Selene para obtener créditos. Además, utilizan contenidos, courseware y horarios similares a los que se encuentran frecuentemente en los MOOCs. Para este estudio, hemos seleccionado las últimas iteraciones de cursos que se ofrecieron durante el primer semestre de 2020.

Estos cursos se han organizado a través de unidades temáticas divididas en módulos. La evaluación se ha llevado a cabo según las directrices del reglamento del estudiante, con un examen por módulo del curso. Las evaluaciones se han realizado a través de exámenes en línea con preguntas tipo test con una sola respuesta correcta. Cada evaluación constaba de 12 a 40 preguntas con la posibilidad de realizar el examen en dos fechas diferentes. Cada examen era abierto y estaba disponible para los estudiantes sólo durante una ventana de tiempo de 60 minutos para minimizar las posibilidades de comportamientos deshonestos o de compartir las respuestas. Además, había pruebas de ensayo, que no

contaban para la calificación del curso. Los instructores organizaron un examen de práctica inicial para que los estudiantes se familiarizaran con la mecánica de las evaluaciones del curso.

7.2.2. Instructores

Los instructores que colaboraron en la evaluación del tablero fueron cuatro en total. Fueron elegidos porque cada uno de ellos estaba a cargo de uno de los cuatro SPOC ofrecidos por la Universidad del Cauca en el primer semestre de 2021. Nos pusimos en contacto con cada uno de ellos, les explicamos el objetivo de la evaluación y aceptaron participar en este estudio. A continuación describimos el perfil de cada uno de ellos. Para anonimizar los datos de los instructores, hemos decidido darles un seudónimo.

- Carlos es el profesor con más experiencia en educación. Lleva 22 años trabajando como profesor de la Universidad de Cauca y desde 2005 trabaja en la línea de investigación de cursos online, donde ha participado en el desarrollo del primer curso online de la universidad. También ha estado a cargo de uno de los cursos de SPOC desde que se implementó Selene en 2016. Es informático y tiene una amplia experiencia en el análisis de datos.
- John no tiene experiencia en cursos presenciales, pero ha estado a cargo de cursos en línea desde 2013 y se unió al equipo de Selene en 2019. Su principal área de trabajo es la telemática y su principal experiencia es la administración de servidores web y entornos virtuales de aprendizaje.
- Lucía es profesora universitaria desde hace 11 años y trabaja como instructora de cursos en línea desde que se implementó la plataforma Selene en 2016. Es instructora de humanidades y tiene muy poca experiencia trabajando con tecnologías educativas y no está relacionada con el análisis de datos.
- Sofía ha trabajado como instructora de escuela secundaria desde 2009 y en 2015 comenzó a trabajar en la educación superior y se unió al grupo Selene en 2018, donde tuvo su primer acercamiento a la educación en línea. Es instructora de humanidades y tiene poca experiencia en tecnologías educativas y análisis de datos.

Hay dos perfiles arquetípicos distintos de instructores. Dos instructores que dominan las tecnologías informáticas y el análisis de datos, y dos instructores que no están relacionados con la materia. Además, también tenemos instructores con amplia experiencia

en cursos en línea e instructores que se están iniciando en esta área. Esto podría ayudar a generalizar los resultados obtenidos, ya que es muy común encontrar estos dos perfiles entre los instructores.

7.2.3. Diseño de la evaluación

La revisión de la literatura tuvo como objetivo revisar el estado de los diferentes enfoques para detectar y monitorear la deshonestidad académica dentro de los MOOCs, SPOCs y otros cursos de aprendizaje en línea. Concluimos que hay dos características importantes: las métricas utilizadas para la detección de los comportamientos deshonestos y la fiabilidad de la herramienta. Además, queremos abordar qué estrategias pueden aplicar los instructores con la herramienta y si es posible reducir la deshonestidad académica. Contribuimos a la literatura sobre el análisis del aprendizaje y la deshonestidad académica abordando las siguientes Preguntas de Investigación (RQs):

- RQ1. ¿Cuál es la percepción de los instructores de las métricas implementadas en el tablero?
- RQ2. ¿Qué piensan los instructores sobre la fiabilidad de las métricas del cuadro de mandos?
- RQ3. ¿Qué estrategias utilizaron o propusieron los instructores para disuadir el fraude?
- RQ4. ¿Creen los instructores que es posible reducir la deshonestidad académica con la ayuda de las métricas presentadas en el cuadro de mando?

Como parte de la evaluación, organizamos tres sesiones con cada uno de los instructores por separado. Todas las sesiones y entrevistas con los instructores se realizaron a través de videollamadas, lo que nos permitió grabar las sesiones para su posterior análisis. A continuación describimos el enfoque de cada una de estas sesiones.

7.2.3.1. fase de introducción

Para esta primera fase nos pusimos en contacto con los instructores encargados de los cuatro cursos ofrecidos en Selene en el primer semestre de 2021 antes de que cada uno de los cursos comenzara. Se programó una sesión privada con cada uno de ellos, en la que se les dio acceso al cuadro de mando y se les presentó un tutorial de las funcionalidades del mismo. La idea de esta fase era que el instructor se familiarizara con el cuadro de mando y pudiera utilizarlo en el curso que tenía a su cargo.

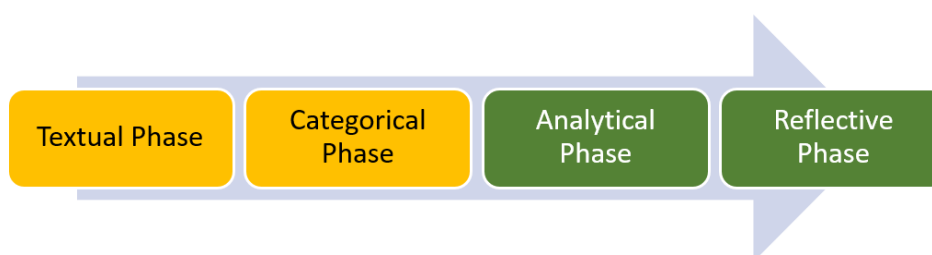


Figura 7.5: Fase de análisis cualitativo

7.2.3.2. Fase de mejora del cuadro de mando

En esta fase se realizó una breve entrevista a cada instructor con la intención de obtener datos para mejorar la herramienta y resolver dudas. Esta fase tuvo lugar aproximadamente a mitad del curso. Se preguntó a los instructores sobre la usabilidad del cuadro de mando y sobre su opinión respecto a las mejoras que se podrían realizar en el mismo. También se respondió a las dudas que pudieran haber surgido después de utilizar el cuadro de mando durante la mitad del curso.

7.2.3.3. Fase de evaluación

Esta última fase fue la principal, ya que nos propusimos responder a nuestras RQs mediante la realización de una entrevista semiestructurada a cada uno de los instructores. Diseñamos un protocolo con preguntas orientadoras para responder a las RQs, pero también permitiendo que las preguntas y respuestas surgieran de forma natural, de manera que la sesión fuera una conversación informal que nos llevara a obtener información más profunda sobre el punto de vista de los instructores respecto a las cuatro RQs.

7.3. Análisis cualitativo

Para el análisis de los datos recogidos en las entrevistas semiestructuradas, hemos seguido un proceso de análisis de contenido cualitativo [81], que es un enfoque de análisis textual cualitativo sistemático, basado en reglas, en el que se establecen categorías y codificación. Es un proceso iterativo, ya que los códigos pueden modificarse o cambiarse a medida que se revisan los datos. Este proceso consta de cuatro etapas (Figura 7.5). Las llevaron a cabo dos personas, un ingeniero electrónico con experiencia en el análisis de datos y un profesor universitario con experiencia de investigación en humanidades.

La fase textual consistió en el preprocesamiento de los datos. En esta fase se realizó una transcripción de cada una de las grabaciones de las sesiones que tuvimos con los

instructores y se seleccionó la información más relevante para nuestro estudio, dejando de lado las conversaciones que no eran relevantes o que no contribuían a responder las siguientes preguntas guía:

- ¿Cuál es su opinión sobre las métricas utilizadas para detectar el comportamiento de los delincuentes?
- ¿Qué grado de fiabilidad considera que tiene el panel de detección?
- ¿Cuáles son las estrategias que ha utilizado o propone para disuadir el fraude?
- ¿Es posible reducir la deshonestidad académica con la ayuda del cuadro de mandos que ha utilizado?

En la fase categórica, las transcripciones se organizaron de acuerdo con las preguntas guía. Las preguntas directamente relacionadas con las RQs y para cada una de ellas se propuso una variable antes de las entrevistas. Luego, para cada variable, se obtuvieron categorías de forma inductiva, donde estas categorías debían explicar la variable y dar respuesta a la pregunta guía (Tabla 7.1). Las categorías se obtuvieron etiquetando palabras que representan temas importantes y recurrentes asociados a cada variable. Por ejemplo, la pregunta RQ1 dice “¿Consideran los instructores que los parámetros utilizados para detectar comportamientos deshonestos son adecuados? En este caso, la variable es “métrica” y se comprobó que, en general, la forma en que se visualizan las métricas es adecuada, que las métricas son precisas, que las métricas se complementan entre sí y que las métricas revelan asociaciones entre sospechosos.

A continuación se realizó una fase de análisis en la que se analizaron las relaciones entre las unidades de texto seleccionadas y las variables y categorías identificadas. Los resultados obtenidos en esta fase se describen en los resultados (Sección 7.4). Por último, la fase reflexiva da lugar a interpretaciones de los resultados que también permiten contrastar los resultados con otros estudios. Los resultados obtenidos de esta fase se describen en la discusión (Sección 8.4).

7.4. Resultados de las entrevistas

Los resultados se organizan con base en las cuatro RQs presentadas en la sección 7.3, con una subsección centrada en cada una de ellas.

Tabla 7.1: Variables, categories, and examples

Variable	Category	Example
Metrics	Visualisations	“The visualisations are very accurate...”
	Appropriate metrics	“The metrics seem to me very adequate...”
	Complementary metrics	“The two metrics work well together, they complement each other...”
	Suspect associations	“It is easy to interpret which students are suspected of fraud generate associations...”
Reliability	Accurate platform	“I think it is very reliable because all the data are recorded on the learning platform and the platform is very accurate...”
	Fraud detection	“There are very low chances that students will submit exactly the same mistakes in their responses, that to me that is academic dishonesty.”
	Partially reliable	“I am not sure about the reliability of the dishonesty detection, but I am confident about the suspicious behaviours...”
Strategies	Different exams	“I always have several different exams, so it is not complicated for me to do that...”
	Question bank	“I think that having a large question bank is a successful strategy...”
	Communication with students	“I would choose to send dissuasive messages...”
	Warnings	“In the first instance, I would send a warning.”
Fraud mitigation	Student ethics	“The dashboard can detect dishonest actions, but it cannot diminish them because these behaviours are part of the students’ ethics...”
	Fraud difficult	“...it is possible to create strategies that make it more difficult for the students...”
	University policies	“I think that there should be an institutional policy for those cases...”

7.4.1. RQ1. ¿Qué percepción tienen los instructores sobre las métricas utilizadas para detectar comportamientos deshonestos son adecuadas?

De acuerdo con la información obtenida a través de las entrevistas, encontramos que todos los instructores están de acuerdo en que las métricas son adecuadas y que pueden ser utilizadas para obtener la información requerida para evaluar la deshonestidad académica de los estudiantes. John mencionó lo siguiente “Las métricas me parecen muy adecuadas y las visualizaciones también, porque la comparación de las respuestas de los alumnos se entiende fácilmente y el uso del dendrograma para visualizar las métricas es muy bueno”.

Además, los instructores que tienen más experiencia con las plataformas educativas y el análisis de datos dijeron que la información proporcionada se procesa fácilmente y que los gráficos se entienden bien. Por lo tanto, los instructores están de acuerdo en que las métricas presentadas por la herramienta son comprensibles y que los gráficos son autoexplicativos.

Por otro lado, según la experiencia del instructor Carlos, considera que la métrica de similitud de respuestas es adecuada porque en varios cursos ofrecidos por él se han dado casos de estudiantes que comparten sus respuestas en el momento del examen. El instructor afirmó lo siguiente: “Compartir las respuestas de los exámenes en los cursos online no es un secreto, lo que ocurre cuando hay personas que pueden comunicarse entre sí e intercambiar las respuestas rápidamente...” Otro instructor, John, reforzó esta idea de la utilidad de esta métrica diciendo: “Desde el punto de vista del instructor, encuentro más útil la similitud de las respuestas que la cercanía de las marcas de tiempo de envío. Por ejemplo, cuando un estudiante hace un examen y pasa la solución a otros, los exámenes no se entregan necesariamente al mismo tiempo”.

En cuanto a las métricas y la importancia que tienen, los instructores destacaron la cuestión de las asociaciones que hacen los estudiantes. Sofía mencionó que la forma en que se visualizan las métricas le permite saber cómo se asocian los alumnos y cómo trabajan en pareja o en grupo para hacer trampa. “Las visualizaciones de las métricas muestran la información de forma sencilla, de manera que las métricas y las visualizaciones son fácilmente comprensibles sin ser un experto en análisis de datos”.

A partir de los resultados obtenidos de la pregunta RQ1, identificamos cuatro aspectos generales. Por un lado, que la presentación de las métricas de forma gráfica ha sido muy acertada, ya que facilita su comprensión. También, que compartir las respuestas entre los estudiantes es un comportamiento que se ha visto reflejado en los SPOCs propuestos por

la Universidad de X, ya que los profesores que los imparten han identificado este comportamiento en los cursos que han tenido a su cargo. Por otro lado, se ha comprobado que las métricas permiten identificar las asociaciones de los alumnos, ya que la forma de graficar las métricas (dendrogramas), permite visualizar si los alumnos trabajan en parejas o en grupos más grandes. Entendemos por colaboraciones ilícitas el hecho de compartir las respuestas de los exámenes o hacerlos en pareja o en grupo, que son acciones no permitidas en los SPOC de la Universidad del Cauca. Creemos que este es un beneficio útil de la herramienta, ya que no sólo permite identificar a los estudiantes sospechosos de deshonestidad académica, sino también saber con quiénes trabajan. Por último, se ha destacado que las métricas se complementan entre sí, ya que se ha comprobado que una de las características de los SPOCs que ofrece la Universidad de X es que los alumnos comparten las respuestas de los exámenes, pero como no hay una retroalimentación inmediata en cuanto a la calificación, los alumnos envían sus exámenes muy cerca en el tiempo, por lo que las asociaciones que se identifican con cada una de las métricas son las mismas, haciendo que las métricas se complementen. Esto es lo contrario de lo que ocurre en los cursos con retroalimentación, en los que los alumnos esperan a ver el resultado de uno de ellos para corregir los errores y enviar el siguiente examen.

7.4.2. RQ2. ¿Cuál es el grado de fiabilidad que los instructores suponen del algoritmo utilizado?

Hemos comprobado que los cuatro instructores declaran que los mecanismos utilizados para detectar a los alumnos potencialmente deshonestos funcionan. Afirmaron que han podido verificar manualmente los resultados mostrados por el cuadro de mando en cuanto a la coincidencia de las respuestas de los exámenes con la hora y la fecha del envío. La instructora Lucía mencionó que “el punto de vista del algoritmo, la herramienta es fiable, porque efectivamente, la herramienta detecta a los alumnos que tienen exámenes similares y que los han presentado muy cerca en el tiempo.”

Además, comprobamos que el cuadro de mando es fiable porque los instructores confían en las herramientas informáticas y en los resultados obtenidos. La instructora Sofía mencionó que “es muy fiable porque todos los datos se registran en la plataforma de aprendizaje y la herramienta es muy precisa, todo está bien organizado, lo que permite que la herramienta sea bastante fiable”.

Por otra parte, hay pruebas de que no sólo sobre el algoritmo se considera fiable, sino también sobre la detección de los comportamientos deshonestos. Es decir, aunque no hay una prueba sólida de que hayan sido deshonestos, hay una gran posibilidad de

que lo hayan sido. Carlos afirmó que según su experiencia en cursos online, este tipo de acciones pueden ser denotadas como deshonestidad académica, el dijo: “ Hay muy pocas posibilidades de que los estudiantes presenten exactamente los mismos errores en sus respuestas, eso, para mí, es deshonestidad académica”

Contrariamente a las declaraciones del instructor Carlos, el instructor John menciona que la herramienta sólo identifica a los estudiantes sospechosos y por lo tanto no es posible estar seguro de que los estudiantes han cometido deshonestidad académica. El instructor mencionó que sería necesario confirmar, de alguna manera, si hubo o no deshonestidad académica, lo cual es difícil de hacer: “No estoy seguro de la fiabilidad de la detección de la deshonestidad, pero estoy seguro de los comportamientos sospechosos, porque la herramienta presenta dos métricas que me llevaron a confirmar que había comportamientos sospechosos”.

En cuanto a la pregunta RQ2, hemos comprobado que la herramienta y la forma en que identifica a los estudiantes sospechosos de deshonestidad académica son muy fiables. En primer lugar, porque hemos comprobado manualmente que la información presentada por la herramienta es correcta, es decir, que el algoritmo de cálculo funciona correctamente. Después, porque los datos y los cálculos realizados son ejecutados por una máquina informática, lo que la hace muy fiable. También que la noción de confiabilidad para detectar a los estudiantes que cometen deshonestidad académica depende mucho de la experiencia del instructor. En otras palabras, el instructor Carlos afirmó sin dudar que los comportamientos identificados por la herramienta eran deshonestidad académica. El instructor lleva mucho tiempo impartiendo cursos online y ya tiene la experiencia necesaria para saber cómo se comportan sus alumnos. Por otro lado, los instructores con menos experiencia en cursos en línea no se atrevieron a afirmar que los comportamientos identificados por la herramienta eran deshonestos. Incluso dudaron de la fiabilidad de la herramienta al concluir que ésta no detecta la deshonestidad académica, sino los comportamientos sospechosos.

7.4.3. RQ3. ¿Qué estrategias utilizaron o propusieron los instructores para disuadir la deshonestidad académica?

Dos de los instructores comenzaron a aplicar estrategias para disuadir la deshonestidad académica después de usar la herramienta, pues gracias a ella, se dieron cuenta de los comportamientos de deshonestidad de sus estudiantes. Una de las estrategias fue dar diferentes versiones de los exámenes a los estudiantes que fueron identificados por la herramienta, y una segunda estrategia fue utilizar un banco de preguntas para aleatorizar

las preguntas de modo que los exámenes fueran diferentes para cada estudiante. Estas estrategias se emplearon en el último examen después de utilizar la herramienta durante los exámenes iniciales. La instructora Sofía del curso de danza aplicó exámenes diferentes a los alumnos que formaban parejas y grupos de trabajo, mencionó: “Siempre tengo varios exámenes diferentes, así que no me resulta complicado hacerlo...”

Por otro lado, Carlos del curso de emprendimiento generó un gran banco de preguntas para el último examen del curso con el doble de preguntas de lo habitual. Él encontró que esta estrategia fue exitosa y algunos estudiantes se abstuvieron de tomar el último examen, él dijo: “Creo que tener un gran banco de preguntas es una estrategia exitosa, pero también aumenta la carga de trabajo de manera significativa”.

Además, todos los instructores coincidieron en afirmar que la principal estrategia de intervención sería intentar acercarse a los alumnos identificados y advertirles sobre su comportamiento. John dijo: “En primer lugar, les enviaría una advertencia. En un segundo momento, pasaría a las acciones de hecho. Prestaría más atención a los exámenes, corroboraría los datos y pensaría en medidas disciplinarias.” Por último, el instructor Carlos dijo que optaría por enviar mensajes disuasorios, diciéndoles a los estudiantes que fueron detectados y recomendándoles que no realicen esas conductas.

También coincidieron en que se podría implementar una estrategia de avisos automáticos antes y durante los exámenes. Es decir, informar a los alumnos de que existe un sistema para detectar comportamientos deshonestos y también enviar una notificación a los alumnos detectados. Sofía dijo: “si me dijeran que hay un sistema que puede detectar actividades sospechosas, no me arriesgaría a cometer deshonestidad académica...”

Encontramos que las estrategias aplicadas no produjeron resultados significativos porque no hubo una planificación previa, y los instructores comenzaron a aplicar correctivos o estrategias en el curso después de ver en la herramienta el comportamiento sospechoso de algunos de los estudiantes. Los instructores no estaban preparados para hacer frente a las sospechas de comportamiento deshonesto de algunos alumnos y creemos que, en las nuevas versiones del curso, los instructores planificarán con antelación y aplicarán así nuevas estrategias. Esperamos que esto pueda ayudar a reducir la prevalencia del mal comportamiento de los estudiantes en los SPOC y que también pueda aplicarse en los cursos en línea tradicionales. Además, pudimos identificar que los instructores necesitan políticas claras que les permitan actuar en este tema. Los cuatro instructores afirmaron que la universidad necesita apoyarlos y definir cómo proceder cuando se identifica a los estudiantes que han cometido deshonestidad académica. Todos están de acuerdo en que estos comportamientos deben ser castigados, pero como no existe una política universitaria sobre la deshonestidad académica no saben cómo proceder.

7.4.4. RQ4. ¿Creen los instructores que es posible reducir la deshonestidad académica mediante el uso del tablero presentado?

Encontramos dos posturas sobre si la herramienta podría ayudar a disminuir la deshonestidad académica. Dos de los instructores estuvieron de acuerdo en que la deshonestidad académica no se puede disuadir con una herramienta porque los comportamientos deshonestos son inherentes a los estudiantes y es difícil disuadirlos. Juan afirmó que "la herramienta puede detectar acciones deshonestas, pero no puede disminuirlas porque estos comportamientos forman parte de la ética del estudiante". Además, Carlos mencionó que "no es posible reducir la deshonestidad académica, pero sí crear estrategias que la dificulten..."

También informaron que la herramienta puede reducir la deshonestidad académica en los cursos en línea, siempre y cuando exista un mecanismo de acción hacia los estudiantes identificados o al menos una notificación o comunicación con el estudiante informando sobre que existe un sistema de detección de deshonestidad académica. Esto se debe a que la sensación de ser vigilado hace que la persona limite sus comportamientos deshonestos. Sofía mencionó que: "informar desde el principio del curso que existe un sistema de detección de deshonestidad académica muy probablemente inhibirá a los estudiantes de compartir sus respuestas".

Además, los dos instructores que informaron que el tablero de control puede ayudar a disminuir la deshonestidad académica, también mencionaron que debería haber políticas dentro de la institución para castigar estos comportamientos. Lucía mencionó que: "actualmente la universidad está otorgando créditos académicos a los estudiantes que fueron identificados por la herramienta, ¿qué piensa hacer la universidad con ellos? Creo que debería haber una política institucional para esos casos..."

Con respecto a la RQ4, se obtuvo que la herramienta no puede reducir por sí misma los comportamientos deshonestos con sus funcionalidades actuales, ya que hasta ahora identifica estos comportamientos, y muestra esta información, pero no hay ningún mecanismo para reducirlos. Estamos de acuerdo con los instructores en que la herramienta carece de un sistema de alerta que pueda enviar automáticamente mensajes informando a los estudiantes de que han sido detectados como sospechosos. Este tipo de estrategia podría mejorar la herramienta y ayudar a disuadir estos comportamientos, sin embargo, también sería un reto gestionar las falsas alarmas, o los estudiantes que juegan con el sistema una vez que saben cómo funciona.

Parte IV

Conclusiones, limitaciones y trabajos futuros

Capítulo 8

Discusión de resultados

Índice

9.1. Conclusiones	129
9.1.1. Observaciones finales	129
9.1.2. Limitaciones	131
9.2. Estancias de investigación	134
9.3. Trabajos futuros	135

En este capítulo se discuten algunos de los resultados de los capítulos anteriores 4, 5, 6 y 7. Más concretamente, en la sección 8.1 se habla de los hallazgos encontrados a partir de la implementación del algoritmo propuesto para la identificación de comportamientos deshonestos, en la sección 8.2 debatimos sobre los perfiles de los alumnos que encontramos en el capítulo 5 y las posibles aplicaciones y resultados para el proceso de aprendizaje. En la sección 8.3 se discute sobre los primeros resultados obtenidos con la herramienta de seguimiento y finalmente, la sección 8.4 presenta y discute nuestros resultados de la evaluación de la eficacia, la utilidad y usabilidad de la herramienta para la identificación y seguimiento de comportamientos deshonestos.

8.1. Algoritmo para la detección de comportamientos con sospecha de deshonestidad académica

Varios estudios anteriores han utilizado enfoques algorítmicos para detectar trampas en entornos en línea [16,32,88,100]. Sin embargo, todos estos estudios se centraron en los MOOC, que tienen importantes diferencias clave con nuestro contexto. En estos MOOC,

los estudiantes pueden optar por recibir un certificado del curso si el curso se completó con éxito, pagando la tarifa de inscripción para acceder al modo verificada; sin embargo, estos cursos no otorgaban crédito para obtener un título. Otra diferencia clave debido al contexto es que los estudiantes de MOOC pueden crear varias cuentas, por lo que una sola persona física puede estar manejando más de una cuenta que podría usarse para explotar las propiedades de retroalimentación del sistema. Sin embargo, en este caso, la creación de cuentas está centralizada como parte del sistema de la Universidad del Cauca, por lo que los estudiantes solo pueden tener una cuenta y luego se ven obligados a depender de trabajar con sus amigos. Es más, el anonimato de los ambientes MOOC es útil para emprender actividades tan poco éticas, sin embargo, los estudiantes de la plataforma Selene Unicauca saben que sus cuentas están vinculadas a su expediente académico. Si bien la metodología de detección algorítmica es similar, hasta donde sabemos, esta es la primera vez que este tipo de metodologías basadas en datos que se basan en el seguimiento de registros de datos de cursos con reconocimiento académico.

Nuestro enfoque algorítmico basado en la similitud entre las soluciones y las marcas de tiempo de envío ha detectado un total de 15 estudiantes en el estudio presentado en el capítulo 4 para el Curso Lean Startup, seis de ellos organizados en tres parejas, y nueve de ellos organizados en tres grupos de 3, y 10 estudiantes para el Curso de Textos, organizado en 5 parejas. La dinámica de estas diferentes comunidades de estudiantes puede representar diferentes comportamientos, en algunos casos pueden tener un trabajo compartido similar y realmente realizar un esfuerzo conjunto, en otros un estudiante puede liderar la carga de trabajo y pasar las respuestas a sus amigos, estudio que se realizó en el capítulo 5. Trabajo anterior [10] indicaron diferentes perfiles de alumnos de CAMEO, en algunos casos realizando trampas más deliberadas mientras que en otros usándolas como plan de respaldo o ayuda. Si bien no hemos profundizado en el análisis de las diferentes dinámicas de comportamiento, consideramos que es una línea prometedora que debe explorarse en profundidad. Los 25 estudiantes detectados, representan un 17 % de los estudiantes que han sido incluidos en el estudio, por lo que este es un alto porcentaje de estudiantes que cometen deshonestidad académica para aprobar un curso de crédito, más aún teniendo en cuenta que nuestro diseño algorítmico se encuentra en el lado conservador. y esperaríamos que más estudiantes realicen deshonestidad académica. Estos porcentajes globales están por encima del orden de magnitud cuando se comparan con otros estudios, como el 13 % de estudiantes de Alexandron et al. (2019) [10] o 1.3 % de estudiantes en Northcutt et al. (2015) [88], a pesar de que estos estudios también indicaron que sus estimaciones eran conservadoras. Creemos que en nuestro caso de estudio estamos detectando un mayor porcentaje de estudiantes debido a que estos cursos son

reconocidos por créditos como parte de un título. Cabe señalar que esta alta prevalencia de deshonestidad académica en estos cursos también está influenciada por los modelos de valoración y evaluación basados en cuestionarios. Estos exámenes son fáciles de copiar tanto en persona como en línea, pero es especialmente sencillo en los exámenes en línea si no se tienen en cuenta las buenas prácticas de diseño (como la aleatorización, grandes grupos de preguntas, etc.). Hay otras formas de implementar evaluaciones de cursos que son menos propensas a presentar problemas de deshonestidad.

En cuanto a los niveles de interacción y las características conductuales, todos los estudios previos [10, 87, 100], coinciden en que los estudiantes que realizaron estos comportamientos académicos deshonestos lograron aprobar el curso con significativamente menos esfuerzo que el resto de los estudiantes. Sin embargo, en este estudio no encontramos los mismos resultados con el Curso Lean Startup, ya que los estudiantes que son detectados por el algoritmo interactuaron cantidades similares o incluso más con el curso (pero la diferencia no fue estadísticamente significativa). Una posible interpretación es que los estudiantes se reúnan para resolver los trabajos de forma conjunta porque quieren sacar buenas notas o porque suelen trabajar con estos equipos, por lo que discuten ejercicios, acuerdan una solución y la entregan juntos, e interactúan mucho con los contenidos al buscando las respuestas a los exámenes. Este comportamiento podría tener raíces motivacionales diferentes a las más deliberadas que se encuentran en los artículos de CAMEO. Los patrones temporales muestran claramente picos de actividad más abruptos cerca del examen en el caso de los estudiantes detectados realizando deshonestidad académica, lo que podría explicarse por reuniones de grupos de trabajo los mismos días para estudiar y resolver el examen juntos.

Por otro lado, el comportamiento de los estudiantes en el Curso de Textos muestra patrones ligeramente diferentes. En este caso los alumnos detectados por el algoritmo han interactuado un poco menos con los contenidos en contraste con los alumnos detectados en el Curso Lean Startup. Esto significa que, aunque los cursos están diseñados siguiendo una estructura y metodología de evaluación similar, observamos diferencias en el comportamiento de los alumnos que detectamos. En general, la cantidad de interacciones en el Curso de Textos es menor que en el Curso Lean Startup. Puede haber múltiples variables que pueden afectar estos hallazgos; como los diferentes temas, y por ende el tipo de contenidos y su dificultad, definitivamente pueden influir en el comportamiento de los estudiantes. Por ejemplo, en nuestro caso de estudio el Curso de Textos tiene menos contenido de video y la mayor parte de su contenido son lecturas. Esto puede explicar por qué los estudiantes detectados como sospechosos de fraude en el Curso de Textos revisan los contenidos antes de una prueba mientras que los estudiantes sospechosos en el Curso

Lean Startuptener muchas interacciones durante la prueba, ya que en este caso los estudiantes pueden usar más fácilmente la funcionalidad de búsqueda en los contenidos para obtener respuestas. Creemos que estos hallazgos resaltan la importancia del equilibrio existente entre el diseño pedagógico, el comportamiento del alumno y la deshonestidad académica.

8.2. Caracterización de las colaboraciones encontradas

En este estudio, detectamos diferentes comportamientos de colaboración entre estudiantes, y planteamos la hipótesis de que algunos de ellos podrían estar fuertemente relacionados con la deshonestidad académica en los MOOC, mientras que otros podrían ser beneficiosos para los estudiantes (Colaboraciones fructíferas). En primer lugar, implementamos el enfoque de agrupación en la subsección 3.3.4 para caracterizar a los estudiantes y luego aplicamos el algoritmo descrito en la subsección 3.3.3 para detectar a los colaboradores. Observamos que la idea principal subyacente para esta caracterización era que las parejas o comunidades de estudiantes detectadas por nuestro método siempre habían presentado sus tareas muy cerca en el tiempo entre sí y que compartían sus respuestas; por lo tanto, este era un comportamiento sospechoso y posiblemente ilícito.

Un hallazgo importante fue que, a pesar de que aplicamos el análisis de conglomerados a todos los cursos por separado, obtuvimos resultados similares aunque los cursos tuvieran diferentes temas, lo que sugiere que este hallazgo podría generalizarse más allá de los cursos presentados en este estudio. Sin embargo, la transferibilidad de los resultados de la agrupación a otros cursos debería analizarse con mayor profundidad, ya que en este estudio sólo se utilizaron con una misma metodología propuesta por la Universidad del Cauca. Además, podemos observar algunas diferencias en los valores de las calificaciones o notas finales para algunos clusters. Esto podría deberse a que la dificultad de los cursos es diferente. Por lo tanto, las características del curso, como la dificultad del tema, deben tenerse en cuenta al cambiar el contexto.

El método de agrupación detectó tres clústers diferentes con características distintas. Las estudiantes del clúster 1 recibieron un buenas calificaciones invirtiendo un gran esfuerzo, interactuaron con los vídeos y contenidos del curso, realizaron aportes en los foros de discusión y presentaron todos sus exámenes. Los estudiantes del clúster 2 se esforzaron poco, casi no interactuaron con vídeos y contenidos, y tampoco participaron en los foros. Aun así, consiguieron buenas calificaciones. Por último, los estudiantes del clúster 3 tuvieron participación nula en el curso y malas calificaciones, a excepción de los estudiantes que se asociaron con estudiantes de los otros clústers.

Las asociaciones 1 “Colaboración fructífera” están compuestas por cuentas que potencialmente trabajaron juntas y tuvieron altos grados de compromiso. Estas asociaciones pueden representar potencialmente a dos estudiantes que se esforzaron en el curso viendo vídeos, trataron de aprender y comprender los contenidos, y se reunieron para presentar sus exámenes juntos, resolviendo potencialmente juntos de forma colaborativa los cuestionarios en una especie de relación equitativa. La motivación aquí puede ser la ambición de mejorar las notas, y podríamos argumentar que esta relación no representa un problema grave para el proceso de aprendizaje de estos estudiantes. Pues Los dos estudiantes trabajarían juntos para conseguir altas calificaciones y de la misma manera, mostrarían un esfuerzo en la plataforma. Este es el único tipo de asociación que demostró un comportamiento que tiene las características más positivas del aprendizaje colaborativo [72]. Sin embargo, es una conducta no permitida en este tipo de cursos.

Las asociaciones 2 “Free-riding” podrían representar una colaboración no equitativa. En este caso, hubo una interacción menos equilibrada en la que los estudiantes del clúster 1 tienen potencialmente una actitud pasiva y pasan las respuestas a los estudiantes del clúster 2 o clúster 3 (potencialmente un amigo o conocido), de modo que este último cuenta con la posibilidad de aprobar el curso sin invertir mucho esfuerzo, practicando el comportamiento conocido como free-riding [119, 128]. De hecho, la literatura ha reportado que un comportamiento típico hacia la deshonestidad académica, es que uno copia del otro (‘activo’), y el otro permite que los otros copien (‘pasivo’) [38]. Además, dejar que otros copien de ti se considera menos grave que copiar realmente de otros [135]. En el caso de esta asociación específica, el impacto en el proceso de aprendizaje de los alumnos del clúster 2 y 3 es obviamente más grave.

De manera similar, en las Asociaciones 3 “Colaboración ilícita”, los estudiantes tuvieron una interacción muy limitada con la plataforma, pero ambos aprobaron el curso. Así, podemos suponer que estas asociaciones realizaron algún tipo de estrategia que les permitió obtener respuestas a las preguntas del examen sin estudiar los contenidos del curso. En este caso, los estudiantes podrían haber estado aplicando estrategias de “gaming the system”, en las que un alumno intenta tener éxito en un entorno educativo explotando las propiedades de la ayuda y la retroalimentación del sistema en lugar de intentar aprender el material para conseguir un aprobado sin invertir el esfuerzo necesario [14]. Este tipo de comportamiento puede ser grave para el proceso de aprendizaje, ya que en varios estudios los autores encontraron que los comportamientos de jugar con el sistema están asociados con malos resultados de aprendizaje [15, 30]. Esto también puede afectar a las creencias y actitudes futuras de estos estudiantes, ya que pueden llegar a pensar que son capaces de lograr los objetivos sin poner mucho esfuerzo.

También detectamos un número de comunidades con más de dos cuentas que colaboran juntas. Sin embargo, es difícil caracterizar sistemáticamente las asociaciones en cada comunidad porque había numerosos estudiantes.

Se necesitará más trabajo para evaluar si los estudiantes se conocían antes de comenzar el curso o si se conocieron en línea en grupos de estudio [73], y luego decidieron participar en una ‘colaboración no ética’. Debido a que son estudiantes de la misma universidad y los cursos son electivos, es muy probable que los estudiantes se conocieran antes de registrarse en el curso.

8.3. Herramienta de seguimiento

En la mayoría de los trabajos sobre MOOC se presentan soluciones a los problemas relacionados con la evaluación. Incluso ya se cuentan con algunas técnicas que permiten evaluar a los estudiantes de manera automática [7, 56, 70, 125]. Sin embargo, en cuanto al asesoramiento y al seguimiento, los trabajos de investigación son muy escasos, y las tecnologías MOOC existentes aún no cuentan con las herramientas adecuadas para realizar este tipo de tareas, a pesar de ser importantes en cualquier proceso de aprendizaje [21, 22], en especial el seguimiento, pues es parte fundamental del proceso educativo, ya que contribuye a alcanzar el cumplimiento de los objetivos establecidos y de esta forma aporta a la mejora de la calidad de educación que se ofrece con a través de los MOOC [21, 29, 129].

Además, la falta de seguimiento es la causa de algunos de los problemas y desafíos que se presentan en los MOOC. Al no existir seguimiento, se obliga a que el estudiante realice todas sus actividades de aprendizaje con poco acompañamiento o supervisión por parte de un tutor, provocando en los estudiantes bajo rendimiento y falta de interés [22, 85], lo que perjudica la calidad de educación que se pretende impartir, además, de ser una de las razones por las cuales los MOOC tienen un alto porcentaje de abandono [22, 85, 137].

Se encontró que en la mayoría de los trabajos se realiza la monitorización de algunas de las interacciones con la plataforma con el objetivo de mejorar el diseño y los contenidos de los cursos, pero no con el fin de brindar un acompañamiento al estudiante. Además, el proceso de recolección de los datos se realiza cuando el curso ya ha terminado y no muestran una adecuada monitorización de las actividades durante el proceso de aprendizaje del estudiante.

En trabajos que cuentan con realimentación se facilitan enlaces a contenidos complementarios e información sobre el uso de la plataforma. Sin embargo, la realimentación no

está relacionada al seguimiento del proceso de aprendizaje del estudiante, es decir su rendimiento, estilo de aprendizaje, interés en los cursos, dificultades, etc. Son utilizados para proporcionar contenidos y no recomendaciones en cuanto al rendimiento del estudiante.

Que otras herramientas se encuentren directamente destinadas al alojamiento de MOOC hace que no se centren en el seguimiento a las actividades de los estudiantes, pues al ser cursos demasiados grandes el docente ya no puede tener una relación personalizada con los estudiantes. Pero en SPOC y MPOC además de ser posible dicha relación es necesaria y el uso de la herramienta facilita al docente las tareas de acompañamiento al estudiante.

Para solucionar el inconveniente, en este trabajo propusimos un prototipo de una herramienta web que permite la monitorización de las actividades de aprendizaje de los estudiantes en cursos MOOC ofrecidos en la Universidad del Cauca a través de una instancia de Open edX, validada por 13 docentes que están relacionados directamente con cursos virtuales.

Se destaca que la herramienta ofrece indicadores predefinidos que son de gran apoyo para MOOC, en especial para el caso de los SPOC y MPOC, en donde se requiere que haya un Seguimiento de las actividades de los estudiantes y que tiene mucho potencial. Además, la forma en cómo se pueden ordenar los resultados de las consultas, hace que la herramienta sea muy útil para cursos en donde la cantidad de estudiantes es limitada pero mayor que en cursos tradicionales y el docente no puede realizar seguimiento a las actividades por sí mismo.

Se propone como trabajo futuro, el estudio de herramientas que brinden retroalimentación automática al estudiante, basadas en un adecuado seguimiento de las actividades de aprendizaje de estudiantes en cursos MOOC y que permitan proveer al estudiante cierto grado de acompañamiento.

Por otra parte se proponer utilizar el seguimiento a las actividades de los estudiantes para identificar sus preferencias de aprendizaje, estilos de aprendizaje o patrones de comportamiento y con base en ellos diseñar mecanismos para la adaptabilidad y personalización de los curso enfocados en el estudiantes.

8.4. Evaluación de la herramienta de detección de conductas con sospecha de deshonestidad académica

Para los resultados obtenidos de la pregunta RQ1 identificamos cuatro aspectos generales. Por un lado, que la presentación de las métricas de forma gráfica ha sido muy

acertada, ya que facilita su comprensión. Por otro lado, que compartir las respuestas entre los estudiantes es un comportamiento que se ha visto reflejado en los SPOCs propuestos por la Universidad de X, ya que los profesores que los imparten los han identificado a lo largo de todas las versiones de los cursos que han tenido a su cargo. Esto hace que la métrica sea muy adecuada. Por otro lado, se encontró que las métricas permiten identificar asociaciones de estudiantes, ya que la forma en que se grafican las métricas (Dendrogramas), permite visualizar si los estudiantes trabajan en parejas o en grupos más grandes y esto le agrega utilidad a la herramienta, ya que no sólo permite identificar a los estudiantes sospechosos de deshonestidad académica, sino también saber con quiénes trabajan. Por último, se ha destacado que las métricas se complementan entre sí, ya que se ha comprobado que una de las características de los SPOCs que ofrece la Universidad de X es que los alumnos comparten las respuestas de los exámenes, pero como no hay una retroalimentación inmediata en cuanto a la calificación, los alumnos envían sus exámenes muy cerca en el tiempo, por lo que las asociaciones que se identifican con cada una de las métricas son las mismas, haciendo que las métricas se complementen. Esto es lo contrario de lo que ocurre en los cursos con retroalimentación, donde los alumnos esperan a ver el resultado de uno de ellos para corregir los errores y enviar el siguiente examen.

En cuanto a la pregunta RQ2, hemos comprobado que la herramienta y la forma en que identifica a los estudiantes sospechosos de deshonestidad académica son muy fiables. En primer lugar, porque hemos comprobado manualmente que la información presentada por la herramienta es correcta, es decir, que el algoritmo de cálculo funciona correctamente. Después, porque los datos y los cálculos realizados son ejecutados por una máquina informática, lo que la hace muy fiable. También que la noción de confiabilidad para detectar a los estudiantes que cometen deshonestidad académica depende mucho de la experiencia del instructor. En otras palabras, el instructor Carlos afirmó sin dudar que los comportamientos identificados por la herramienta eran deshonestidad académica. El instructor lleva mucho tiempo impartiendo cursos online y ya tiene la experiencia necesaria para saber cómo se comportan sus alumnos. Por otro lado, los instructores con menos experiencia en cursos en línea no se atrevieron a afirmar que los comportamientos identificados por la herramienta eran deshonestos. Incluso dudaron de la fiabilidad de la herramienta al concluir que ésta no detecta la deshonestidad académica, sino los comportamientos sospechosos.

Por otro lado, encontramos que las estrategias aplicadas no dieron resultados porque no hubo una planificación previa y los instructores simplemente las aplicaron después de que los cursos estuvieran en marcha y pudieran ver a algunos de los estudiantes de-

tectados. Los instructores no estaban preparados para hacer frente a las sospechas de comportamiento deshonesto de algunos estudiantes y creemos que, en las nuevas versiones del curso, los instructores podrán planificar con antelación y así poder aplicar sus estrategias. Esperamos que esto reduzca los beneficios del mal comportamiento de los estudiantes en los cursos SPOC y que también pueda aplicarse en los cursos tradicionales en línea. Además, pudimos identificar que los instructores necesitan políticas claras que les permitan actuar en este tema. Los cuatro instructores afirmaron que la universidad necesita apoyarlos y definir cómo proceder cuando se identifica a los estudiantes que han cometido deshonestidad académica. Todos están de acuerdo en que estos comportamientos deben ser castigados, pero como no existe una política universitaria sobre la deshonestidad académica no saben cómo proceder.

Con respecto a la pregunta Q4, encontramos que la herramienta no puede reducir las conductas deshonestas, ya que hasta ahora las identifica y no tiene ningún mecanismo para evitarlas o tratar de reducirlas. Coincidimos con los instructores en que la herramienta carece de un sistema de alerta que envíe automáticamente mensajes informando a los estudiantes de que han sido detectados como sospechosos. Este tipo de estrategia mejoraría la herramienta y le permitiría ser un mejor soporte para disuadir estos comportamientos.

Además, hemos conseguido que se hable de la posibilidad de implantar políticas de deshonestidad académica en los cursos online. Creemos que esto se debe a que, a diferencia de otros trabajos, hemos pasado de un enfoque de análisis de datos a un tablero de control que funciona en tiempo real y facilita a los instructores el seguimiento del comportamiento deshonesto de los estudiantes. Se ha mencionado que este tablero de control proporciona un apoyo para que los cursos en línea sean reconocidos académicamente con mayor seguridad y facilitaría la aplicación de políticas dentro de la universidad.

Capítulo 9

Conclusiones, Limitaciones y Trabajo futuro

Este último capítulo de la tesis comprende las conclusiones finales (Sección 9.1), resultados de las estancias de investigación en la Sección 9.2 y algunas ideas sobre el trabajo futuro (Sección 9.3).

9.1. Conclusiones

Esta sección está dividida primero en la Subsección 9.1.1 que presenta algunas observaciones finales y luego la Subsección 9.1.2 describe las limitaciones de nuestro trabajo.

9.1.1. Observaciones finales

En principio, logramos desarrollar un método basado en datos para la detección de comportamientos deshonestos en el aprendizaje en línea que se basó en trabajos anteriores pero que también introdujo nuevas características para una detección más confiable. Este método fue probado en un curso de crédito impartido en la plataforma Selene Unicauca y encontramos que el 17 % de los estudiantes han realizado acciones académicas deshonestas y el 100 % por ciento de los estudiantes que fueron detectados realizando deshonestidad académica aprobaron el curso, mientras que en el caso de los estudiantes que no fueron detectados, solo el 62 % aprobaron el curso. También reportamos características de comportamiento significativamente diferentes entre estas dos cohortes. Estos datos nos llevaron a profundizar más en el comportamiento de los estudiantes y sus razones para la deshonestidad académica.

Es por eso, que nos centramos en caracterizar los comportamientos de los estudiantes con sospecha de deshonestidad académica y encontramos que los estudiantes detectados tienen colaboraciones en parejas o grupos más grandes. Y aunque las colaboraciones pueden surgir de forma espontánea porque las personas pueden encontrarse en los foros o en los grupos de trabajo virtuales, o porque los amigos deciden hacer un curso juntos, especialmente en los SPOCs que son ofrecidos de manera privada, donde es más probable que los estudiantes puedan conocerse, estas colaboraciones pueden no ser éticas.

Además, aunque generalmente las colaboraciones en los diferentes entornos de aprendizaje son positivas para el proceso de aprendizaje, en esta disertación, se ha revelado que no todas las colaboraciones de los estudiantes pueden considerarse buenas o beneficiosas. Este fenómeno no es nuevo, y en los cursos presenciales tradicionales, los investigadores y los profesionales han informado con frecuencia de colaboraciones no equitativas o deshonestas [57, 74, 131]. Y en este estudio ha ampliado el estado del arte mediante la implementación de una caracterización basada en datos de los diferentes tipos de colaboración presente en SPOCs con reconocimiento académico.

La colaboración puede ser un factor importante en los resultados de los estudiantes en cualquier tipo de curso, ya que el aprendizaje puede surgir de las conexiones espontáneas entre los estudiantes y en muchos de los trabajos que hemos encontrado se destacan las ventajas de la colaboración. Sin embargo, la mayoría de las asociaciones que hemos detectado han mostrado un bajo interés en el aprendizaje del curso y conductas deshonestas explícitas. Por lo tanto, sostenemos que todavía es necesario estudiar más profundamente los tipos de colaboración que pueden surgir en los MOOCs y SPOCs, y otros tipos de cursos en línea, para entender realmente cuáles de ellos pueden ser positivos para los resultados de aprendizaje de los estudiantes.

Por otra parte, en este estudio también presentamos propuestas para el seguimiento de las actividades de aprendizaje de los estudiantes y para la detección de conductas de deshonestidad académica. Creemos que nuestra herramienta de detección es la primera de su clase, ya que no se han encontrado otras similares en la literatura, pues la mayoría de los estudios sobre algoritmos de deshonestidad académica han presentado resultados de análisis retrospectivos. En cambio, nuestra propuesta ofrece las métricas en tiempo real y pueden ser consultadas por los instructores durante el desarrollo del curso para analizar el comportamiento de los alumnos.

En la evaluación de la herramienta para la detección de deshonestidad académica se constató que los instructores consideraban la herramienta muy fiable. Además, el algoritmo utilizado fue probado en varias cohortes posteriores en las que se han encontrado los mismos comportamientos de los estudiantes, lo que indica que los comportamien-

tos detectados se generalizan en todos los cursos. Sin embargo, es necesario mejorar la herramienta para hacerla más explicativa, interpretable y añadir métricas adicionales para verificar que aquellos casos sospechosos de deshonestidad académica no son falsas alarmas.

Por último, nos gustaría señalar que las titulaciones y programas basados en MOOCs se están convirtiendo en una tendencia importante, y muchas universidades están experimentando con metodologías de créditos tanto semipresenciales como totalmente online. Nuestros hallazgos y los presentados en otros estudios indican claramente la gravedad de la deshonestidad académica en estos entornos, y que requieren más investigación, experimentos de intervención y la alineación de las prácticas docentes y las funcionalidades de las plataformas con los hallazgos de la investigación, para que toda la comunidad pueda proponer mejoras sobre el tema. De lo contrario, esta situación podría poner en peligro el futuro del reconocimiento académico en el aprendizaje en línea.

9.1.2. Limitaciones

En cuanto a las limitaciones del estudio, la primera y más obvia es que no tenemos pruebas sólidas (como la transmisión de vídeo) de que los estudiantes estén realizando tal deshonestidad académica juntos, sin embargo, las evidencias que se mostraron en la sección 4 indican comportamientos con sospecha de deshonestidad académica. Solo hemos probado el algoritmo en SPOCs, y por lo tanto, no podemos argumentar que esto podría generalizarse a MOOCs con reconocimiento académico. Además, con base en nuestros hallazgos, creemos que las conductas de deshonestidad académica pueden estar influenciadas de manera importante por múltiples variables y no pueden generalizarse de un curso a otro, especialmente cuando hay cambios en el diseño pedagógico, los contenidos y la temática. Por ejemplo, en cursos prácticos con evaluaciones continuas bien diseñadas y sin exámenes autocalificados, la prevalencia de la deshonestidad académica se reduciría significativamente. Una clara debilidad del método de detección sería realizar ataques de confrontación, lo que significa que si los estudiantes conocen cómo funciona el algoritmo de detección, podrían enviarlos en diferentes momentos y seleccionar algunas respuestas diferentes para agregar ruido que dificultaría su detección.

Nuestro trabajo se ha centrado en caracterizar una serie de colaboraciones siguiendo un enfoque basado en datos. Sin embargo, creemos que hay algunas limitaciones en los hallazgos reportados en este trabajo. Primero, tuvimos un valor umbral claro en nuestra metodología de detección, y podría haber otros tipos de colaboraciones que no han sido capturados por el enfoque algorítmico utilizado en este estudio; por lo tanto, tener un

umbral diferente de detección de los estudiantes que categorizamos como colaboradores tendría un impacto en la precisión del algoritmo. Además, aunque los colaboradores descubiertos aportan pruebas sólidas, ya que las diferencias son estadísticamente significativas, no disponemos de una verdad de base que nos ayude a refinar el algoritmo y a evaluar su calidad real. De hecho, podría haber otras explicaciones potenciales a los resultados que reportamos de colaboraciones. Además, el contexto puede ser un fuerte determinante para la existencia de diferentes colaboraciones y comportamientos [67]. La temática y el diseño del curso [48], la plataforma en la que se realizó [122], y la audiencia a la que se dirige el curso [66], podrían tener una influencia importante que podría llevar a comportamientos deshonestos y colaboraciones diferentes a los aquí presentados. Por lo tanto, sería necesario un estudio más amplio con diferentes contextos para poder generalizar los hallazgos que reportamos.

Una parte del objetivo de este trabajo era arrojar nueva luz sobre la comprensión de las colaboraciones en línea de los estudiantes, sus comportamientos, motivaciones y necesidades. Esto también puede ayudar a comprender mejor el papel que la colaboración en línea puede tener en los resultados del aprendizaje y proporcionar al instructor herramientas que le permitan mejorar de alguna manera el diseño y el desarrollo de los MOOCs para promover más colaboración. Sin embargo, la alta prevalencia de colaboraciones en las que los alumnos aprobaban claramente un curso gracias a algún tipo de "free-riding" limita las conclusiones positivas que podemos extraer del estudio. Un trabajo anterior [9] descubrió que algunos aspectos del diseño del curso, como la aleatorización, podrían ayudar mucho a disuadir la deshonestidad académica. Existen numerosas opciones de diseño que pueden ayudar a minimizar estos problemas y el trabajo futuro debería invertir tiempo en la creación de directrices útiles para los diseñadores y profesionales de los cursos en línea.

Por otra parte, creemos que la herramienta de detección y seguimiento presentada, es la primera de este tipo, ya que los trabajos anteriores se han centrado en análisis retrospectivos de la deshonestidad académica [9,10,98]. La herramienta presentada permite tener una visualización en tiempo real de los alumnos y grupos detectados como académicamente deshonestos con indicadores de comportamiento que apoyan esta detección. Sin embargo, dado que se trata de la primera herramienta creada con este fin, también debemos evaluar las necesidades de los instructores en términos de alfabetización en datos y analítica que puedan permitir comprender las métricas presentadas. Por ejemplo, nosotros incluimos gráficos de comportamiento o dendrogramas con las comunidades de deshonestidad académica; por lo tanto, los instructores no podrán llegar a las conclusiones adecuadas interpretando parte de esta información. Por ello, creemos que más allá

de la fiabilidad del algoritmo, un aspecto muy importante es la interpretabilidad de la información que mostramos, que permita al instructor tomar una buena decisión. Evaluar esta interpretabilidad es parte de las direcciones futuras.

En cuanto a las limitaciones, la métrica de la similitud en las respuestas de los estudiantes fue diseñada para funcionar sólo con preguntas de opción múltiple hasta ahora. Aunque creemos que puede funcionar con respuestas abiertas cortas y, en general, con respuestas cortas que tienen un número limitado de respuestas potenciales, hay que trabajar más para que funcione así. Por otro lado, la métrica de cercanía en el envío de la marca de tiempo probablemente no funcionará tan bien si se implementan exámenes cronometrados, porque la mayoría de los estudiantes enviarán muy cerca del tiempo en cualquier caso. Además, el rendimiento podría variar también si se implementa la aleatoriedad de preguntas en los exámenes. En nuestro caso, al haber exámenes en dos fechas diferentes, es posible que los alumnos que se examinen en la primera fecha compartan las preguntas y respuestas con los alumnos que se examinen en la segunda fecha. En esos casos, esto podría ser un potencial comportamiento deshonesto que no sería detectado por nuestras métricas actuales. Por lo tanto, todavía hay margen de mejora y nuevos casos de estudio como parte del trabajo futuro.

Además, también existe la posibilidad de que algunos estudiantes envíen todas las respuestas correctas en un corto período de tiempo y sean identificados erróneamente como sospechosos de deshonestidad académica en lo que se conoce como una falsa alarma. Los profesores no identificaron ninguno de estos casos durante el estudio de casos, pero es totalmente viable. Por lo tanto, implementar sistemas completamente automáticos podría ser un reto, ya que los algoritmos no son perfectos. De ahí que abogemos por que los profesores realicen indagaciones adicionales, consulten el histórico de calificaciones y analicen los factores contextuales del alumno para validar los resultados de los algoritmos.

Además, la evaluación se llevó a cabo con cuatro instructores que estaban a cargo de los cuatro SPOC puestos en marcha en la Universidad del Cauca, lo que podría ser una muestra limitada para generalizar los resultados del estudio. Sin embargo, creemos que este estudio piloto inicial muestra la aplicabilidad de la herramienta en un entorno real. También, los trabajos futuros deberán evaluar cuáles son las posibilidades de intervención, que actualmente se han omitido. Pues, creemos que nuestro estudio ha dado un paso más en comparación con los trabajos anteriores, y esperamos que esto conduzca a nuevos estudios que evalúen los algoritmos de deshonestidad académica desplegados en cursos en vivo.

9.2. Estancias de investigación

En el marco de la investigación realizada durante este trabajo, el autor realizó cuatro estancias de investigación en las siguientes universidades de alta calidad:

- La primera estancia se realizó del 13 de octubre al 12 de diciembre del 2017 en la Universidad Galileo de Guatemala, bajo la asesoría del Doctor Héctor Roderico Amado Salvatierra coordinador del área de investigación del GES (Galileo Educational System). Las actividades desarrolladas fueron:
 - Revisión bibliográfica sistémica para la identificación del estado del arte para la tesis doctoral.
 - Formación en la plataforma Open edX, plataforma que actualmente funciona y ofrece SPOCs en la Universidad del Cauca.
 - Presentación del trabajo titulado: “Características del comportamiento de los estudiantes en un MPOC con reconocimiento académico sobre una instancia Open edX” en la conferencia internacional MOOC-Maker celebrada los días 16 y 17 de noviembre del 2017.

- La segunda estancia se realizó entre el 7 de octubre y 10 de noviembre del 2018 en la Universidad de Chile, bajo la asesoría del Doctor Oscar Jeréz director del Centro de Enseñanza y Aprendizaje. En dicha estancia se desarrollaron las siguientes actividades:
 - Finalización de la escritura y sometimiento del artículo “Extraction of Student Interaction Data from an Open edX Platform” publicado en la Revista Ingenierías Universidad de Medellín en categoría B por Publindex para el año 2021.

- La tercera estancia se realizó de manera virtual entre el 9 de diciembre y el 12 de marzo del 2021 en el Departamento de ciencias de la enseñanza del Instituto de Ciencias Weizmann, bajo la asesoría del Doctor Giora Alexandrón. Durante este tiempo se realizaron las siguientes actividades:
 - Finalización de escritura del artículo “Data-driven detection and characterization of communities of accounts collaborating in MOOCs” publicado en la revista Future Generation Computer Systems categorizada como una JCR Q1.

- Diseño y desarrollo del caso de estudio del Capítulo 7: Evaluación de la herramienta de detección y seguimiento de comportamientos de deshonestidad académica.
- La cuarta estancia se realizó entre el 24 de abril y 23 de julio del 2021 en el Departamento de Ingeniería de la Información y Comunicaciones de la Universidad de Murcia, bajo la tutoría del PhD. José Antonio Ruipérez Valiente. Se desarrolló la siguiente actividad:
 - Escritura del artículo “Evaluating a Learning Analytics Dashboard to Detect Dishonest Behaviours: A Case Study in SPOCs with Academic Recognition” aceptado para publicación en la revista Journal of Computer Assisted Learning clasificada como JCR Q1.

9.3. Trabajos futuros

Como parte del trabajo futuro, planeamos utilizar las lecciones aprendidas de nuestro trabajo y de otros sobre la influencia del diseño del curso para proponer directrices a los instructores que puedan reducir las conductas de deshonestidad académica realizadas por los estudiantes. Y además, promover dentro de nuestra universidad la creación de lineamientos a tener en cuenta para los casos de deshonestidad académica en cursos en línea. Debido a que al ser una universidad presencial hacen falta políticas para proceder ante conductas deshonestas en SPOCs con reconocimiento académico.

También, tenemos previsto añadir nuevas fuentes de datos a nuestro análisis para mejorar los conocimientos y la caracterización. Por ejemplo, las interacciones más detalladas del foro podrían aportar información sobre cómo interactúan los estudiantes y podríamos contrastar esta información con estos resultados. Además, podríamos detectar interacciones saludables a partir de un análisis de minería de textos entre un grupo de estudiantes que pertenecen al clúster 1, lo que daría más información sobre una colaboración saludable y fructífera. Por otra parte, los estudios de métodos mixtos que implican la recogida de datos cualitativos, como entrevistas y grupos de discusión, podrían ayudar a validar algunas de las inferencias realizadas en este trabajo. Aunque los estudiantes que participaron en comportamientos deshonestos podrían ser reacios a revelar los detalles de su comportamiento, los estudios cualitativos, al menos, podrían ser beneficiosos para corroborar las conclusiones sobre los vínculos de colaboración saludables identificados en este estudio.

Por ultimo, planeamos utilizar las lecciones aprendidas de este estudio de caso y el trabajo previo sobre la influencia del diseño del curso en la deshonestidad académica, para proponer directrices de diseño a los instructores que puedan reducir la cantidad de deshonestidad académica. También planeamos seguir trabajando en la robustez de la herramienta implementando más métricas para detectar la deshonestidad académica para aumentar su fiabilidad, como el análisis de la dirección IP de los envíos de los estudiantes, el análisis del rendimiento de los estudiantes, o incluso añadir la posibilidad de ajustar el umbral de detección de ambas métricas. Esto sería un gran apoyo para el instructor en aquellos casos límite.

Apéndice A

Prueba usabilidad, utilidad y opinión

Los siguientes cuestionarios tienen como propósito realizar una evaluación preliminar de la herramienta. Previo a contestar esta encuesta se solicita haber probado la herramienta, por lo que se proponen algunas tareas a realizar. Nota: Todos los datos recopilados en la presente encuesta únicamente serán utilizados con fines de investigación y mejora de la herramienta. Las respuestas de cada persona serán de carácter confidencial.

Cuestionario de Usabilidad:

Conteste de 1 a 5, cuán de acuerdo está con las siguientes afirmaciones

1. La herramienta es fácil de utilizar.
2. Me gusta la apariencia de la herramienta.
3. El lenguaje utilizado es comprensible.
4. El objetivo de cada característica de la interfaz es claro.
5. Los resultados mostrados en pantalla son fáciles de interpretar.
6. El contenido del menú es suficientemente descriptivo de la información que se ofrece.
7. Es fácil realizar una consulta.
8. Los elementos dentro de las páginas, permiten saber dónde me encuentro dentro del sitio y cómo navegar en él.
9. Los resultados de las consultas aparecen enseguida.
10. Considero que gráficamente el sitio está equilibrado.

Cuestionario de utilidad:

Después de interactuar con la herramienta, conteste de 1 a 5, cuán de acuerdo está con las siguientes afirmaciones.

1. Queda claro cuál es el propósito de la herramienta.
2. Es fácil identificar los estudiantes con mejor rendimiento.
3. Es posible reconocer los estudiantes más persistentes en un curso.
4. Es posible conocer el progreso de un estudiante en particular.
5. Es posible conocer a los estudiantes con menos constancia en un curso.

Cuestionario de opinión:

Conteste de 1 a 5, cuán de acuerdo está con las siguientes afirmaciones y responda a las preguntas abiertas.

1. Es útil para realizar seguimiento a las actividades de aprendizaje de los estudiantes en cursos SPOC.
2. Puede aportar información útil en cursos reconocidos académicamente por una institución de educación y/o homologables a créditos académicos.
3. Brinda la información necesaria para el seguimiento de los estudiantes de un curso SPOC.
4. Utilizaría la herramienta.
5. Recomendaría el uso de la herramienta.
6. ¿Qué destaca de la herramienta?
7. ¿Qué se puede mejorar?

Apéndice B

Diseño del mecanismo para el seguimiento de las actividades de aprendizaje de los estudiantes en MOOC

El diseño del mecanismo para el seguimiento de las actividades de aprendizaje de estudiantes en cursos en línea, abiertos y masivos se realizó teniendo en cuenta la estructura y funcionamiento de la plataforma Open edX. En la Figura B.1 se presenta la arquitectura general del mecanismo construido.

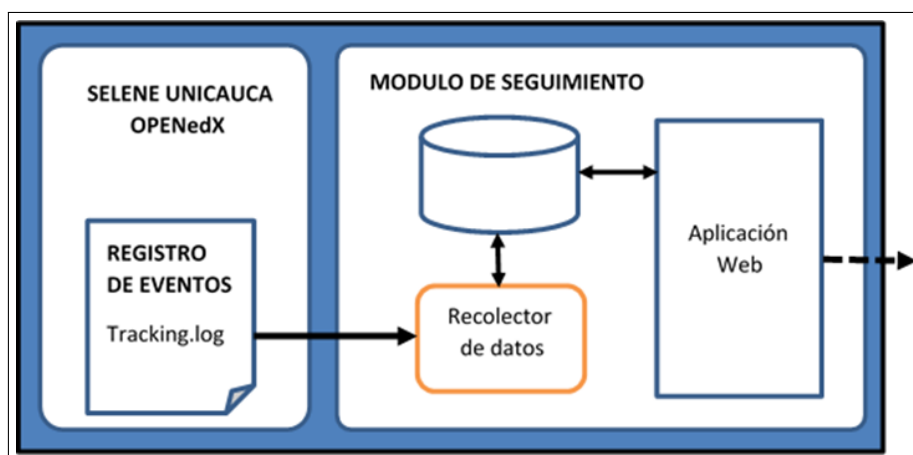


Figura B.1: Arquitectura general de la herramienta de seguimiento

El archivo tracking.log es en donde se registran las interacciones de los estudiantes con la plataforma de aprendizaje Selene, una instancia de Open edX. Su ubicación dentro

del servidor es: /edx/var/log/tracking/. El Recolector de datos, es el encargado de obtener constantemente una copia del archivo tracking.log, leer su contenido, extraer la información relevante para el seguimiento de las actividades de aprendizaje y guardarla en la base de datos. Por su parte, la aplicación web, permite mostrar a los usuarios en este caso a los docentes del curso, los resultados del seguimiento registrado en la base de datos del mecanismo.

En las siguientes secciones se explica con mayor detalle el funcionamiento de la herramienta para realizar seguimiento de las actividades de aprendizaje de los estudiantes en cursos MOOC.

B.1. Vista de escenarios

La descripción de la arquitectura en esta vista se hace mediante Diagramas de Casos de uso, a partir de los cuales se relacionan las cuatro vistas restantes. La vista de Escenarios es obligatoria cuando se utiliza el modelo 4+1 vistas, ya que todos los elementos de la arquitectura se derivan de los requerimientos que ahí se presentan. A continuación, se listan los casos de uso del sistema:

- Iniciar Sesión
- Ingresar usuario y contraseña
- Ver interacciones
- Ver ingresos al curso
- Ver ingresos a Contenidos
- Ver interacciones Vídeos
- Ver interacciones Foros
- Ver interacciones Exámenes
- Seleccionar Curso
- Seleccionar Sección
- Seleccionar Sección
- Seleccionar Unidad

- Seleccionar Unidad
- Seleccionar Estudiante

En la Figura B.2, se presenta el Diagrama de Casos de Uso que orienta el diseño arquitectónico de este trabajo.

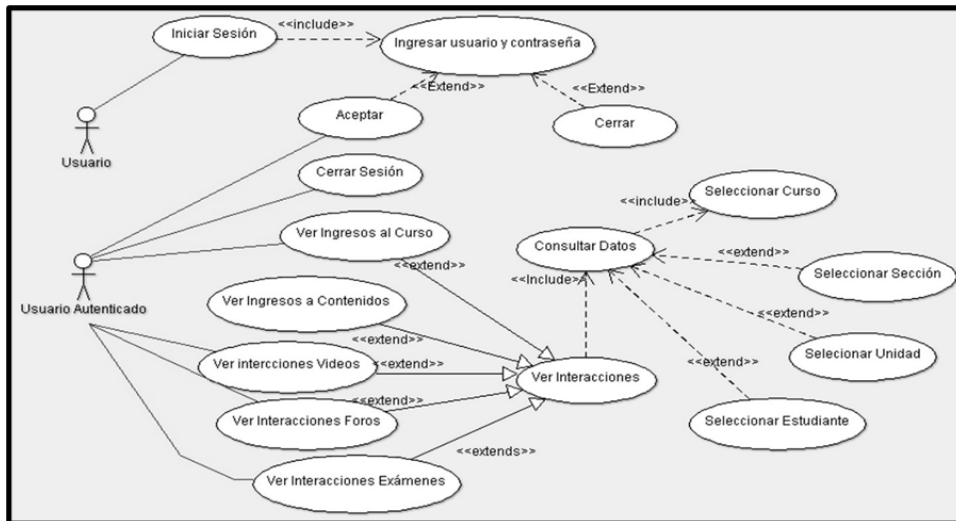


Figura B.2: Diagrama de casos de uso

La herramienta debe permitir a los docentes realizar consultas sobre los datos de las interacciones de los estudiantes con la plataforma de aprendizaje Selene. Los principales datos presentados se relacionan con los ingresos de los estudiantes al curso, navegación por las secciones y unidades del curso, publicaciones y votos realizadas en los foros, interacciones con los videos e interacciones con las evaluaciones. Para las consultas los usuarios registrados deben poder escoger el curso, la sesión, la unidad y el estudiante, y de esta forma realizar un seguimiento a las actividades de los estudiantes, es decir, ver cuáles son las interacciones que los estudiantes tienen con la plataforma de aprendizaje.

B.2. Vista lógica

En esta vista se representa la funcionalidad que el sistema proporcionará a los usuarios finales. Es decir, lo que el sistema debe hacer, y las funciones y servicios que ofrece. La representación se realiza mediante Diagramas de Clases. En la Figura B.3 se presenta a continuación un diagrama general para el mecanismo de seguimiento, de manera que permita entender cómo funciona el mecanismo de una manera más simple.

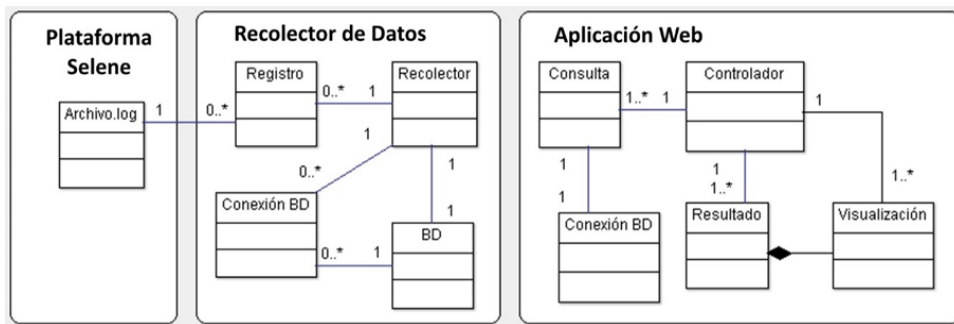


Figura B.3: Diagrama de clases

B.3. Vista de procesos

En esta vista se muestran los procesos que hay en el sistema y la forma en la que se comunican estos procesos. En esta parte se describen los flujos de trabajo del negocio y del sistema por medio de Diagramas de Actividades directamente relacionados con los Casos de Uso definidos anteriormente. Para tener una adecuada descripción del sistema se tienen en cuenta dos procesos: Monitorización de las actividades de aprendizaje (obtención de registros) y Visualización.

Monitorización de las actividades de aprendizaje:

En la Figura B.4 se presenta el diagrama de actividades asociado al proceso de obtención de los registros que se generan en la plataforma de aprendizaje Selene (Open edX).

La plataforma Selene actualiza constantemente un archivo con extensión Log en el cual se registran las interacciones del estudiante con la plataforma de aprendizaje. De aquí, mediante el manejo del archivo es posible realizar una monitorización constante de las actividades de aprendizaje. Como se muestra, el archivo es leído por el mecanismo, se filtra la información necesaria, es decir, los registros de las actividades a las cuales se interesa realizar seguimiento, se verifica que la información aún no se encuentre en la base de datos y finalmente se guarda dichos registros.

La herramienta captura desde el archivo tracking.log todos los registros de los estudiantes. Dicho archivo se encuentra en un formato JSON y en donde cada una de las actividades se registra como un elemento.

Visualización

Una vez los registros de las actividades de los estudiantes se encuentren en la base de datos, es necesario que se puedan visualizar. En la Figura B.5 se presenta el proceso asociado a la visualización de los datos obtenidos.

Para esto, se propuso la construcción de una aplicación web que permite al docente

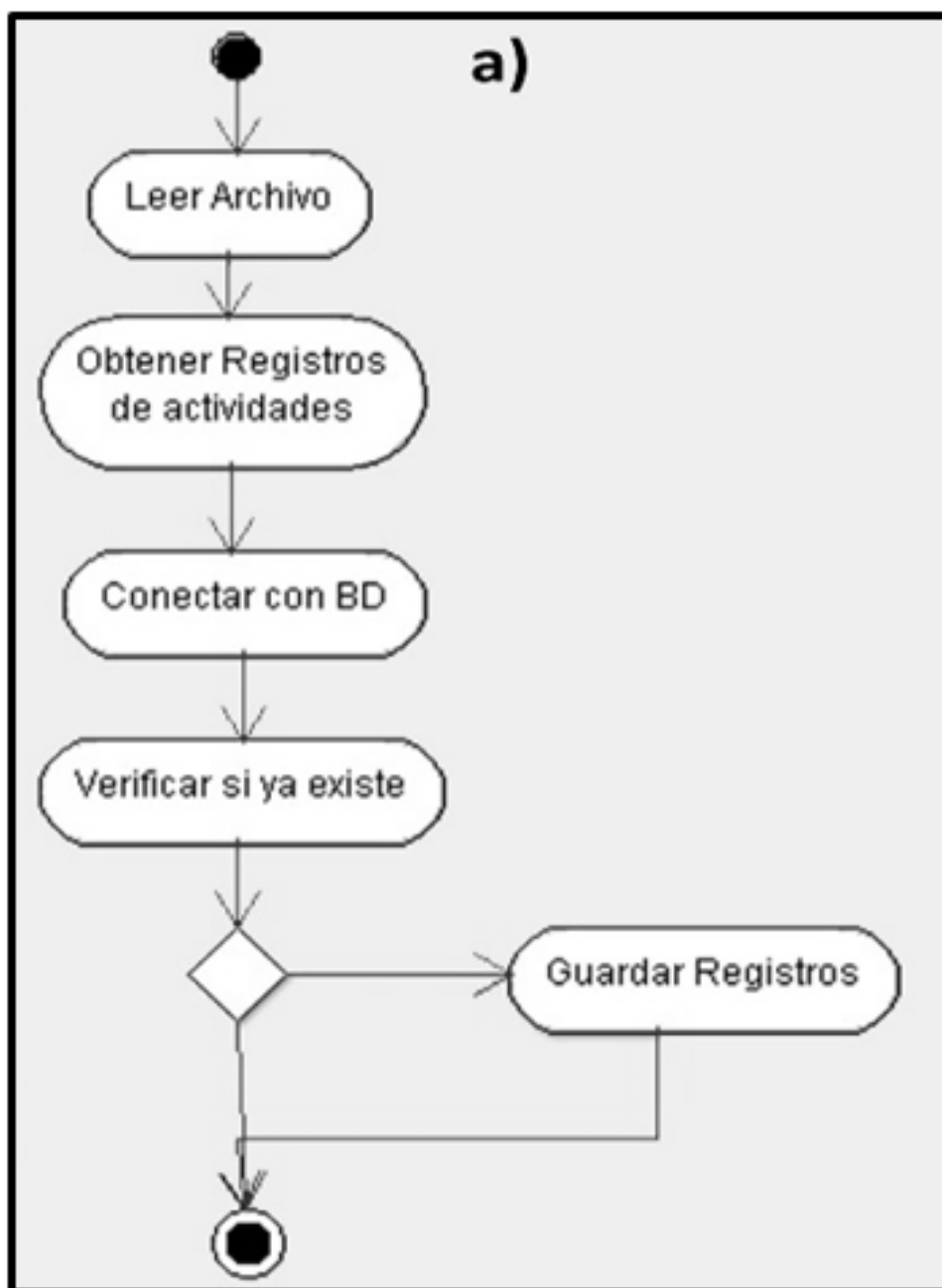


Figura B.4: Diagrama de actividades (Monitorización)

ver los resultados del seguimiento en cualquier momento y lugar.

B.4. Vista de implementación

Esta vista se ocupa de la gestión del software; en otras palabras, se muestra el sistema software mediante sus componentes y sus relaciones. Como se describe en la arquitectura

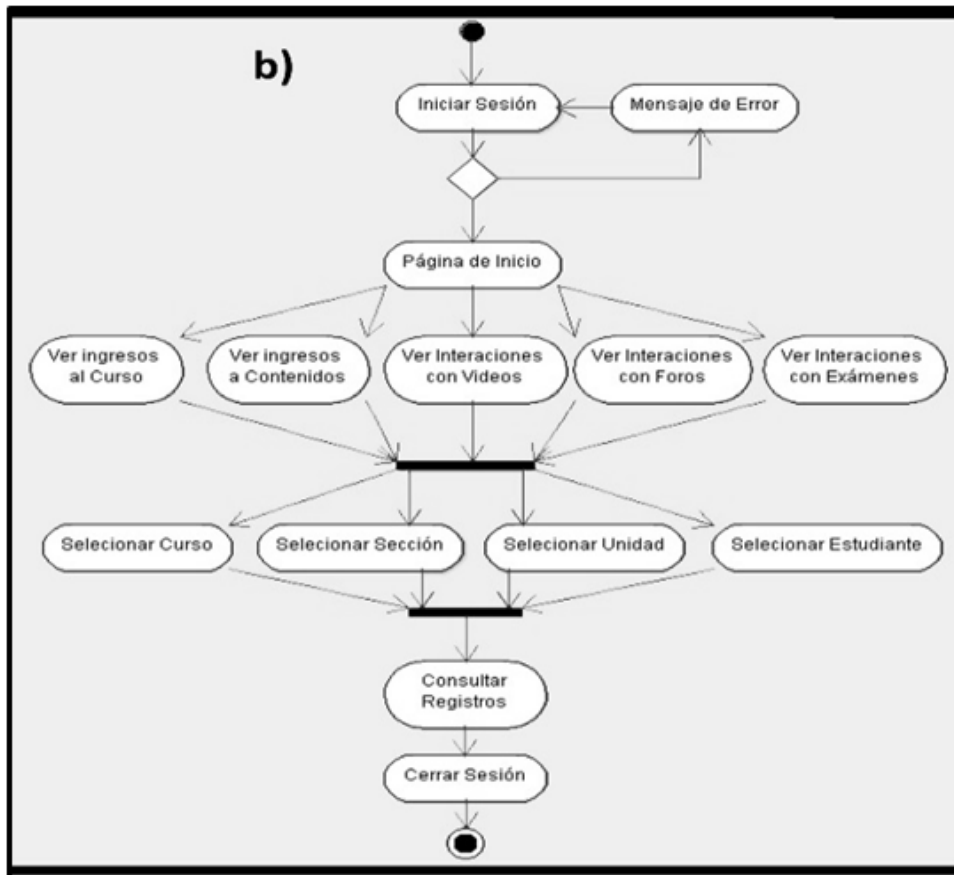


Figura B.5: Diagrama de actividades (Visualización)

general, el mecanismo se encuentra dividido en dos módulos importantes, el módulo de recolección de datos y el módulo de Visualización (Aplicación web). En la Figura B.6 y B.7 se muestran los Diagramas de Componentes para estos módulos.

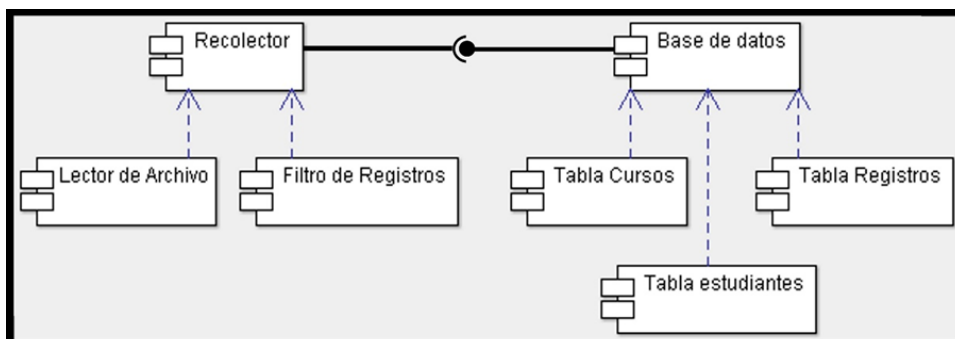


Figura B.6: Diagrama de componentes Módulo de monitorización

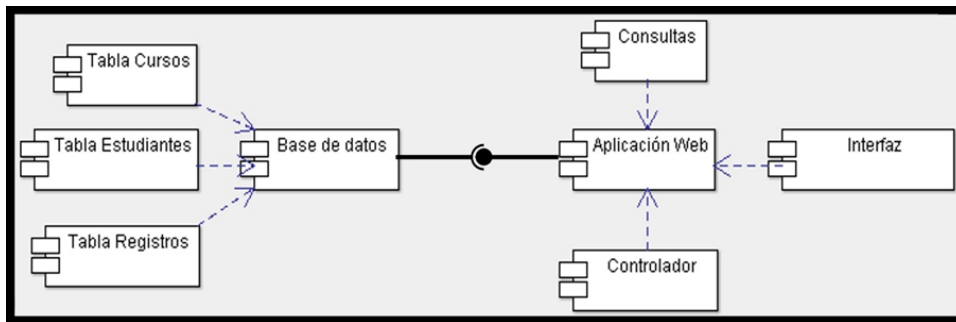


Figura B.7: Diagrama de componentes Módulo de visualización

B.5. Vista de despliegue

En esta vista se muestra todos los componentes físicos del sistema, así como las conexiones que conforman la solución (incluyendo los servicios). En la Figura B.8, se muestra el Diagrama de Despliegue del mecanismo para el seguimiento.

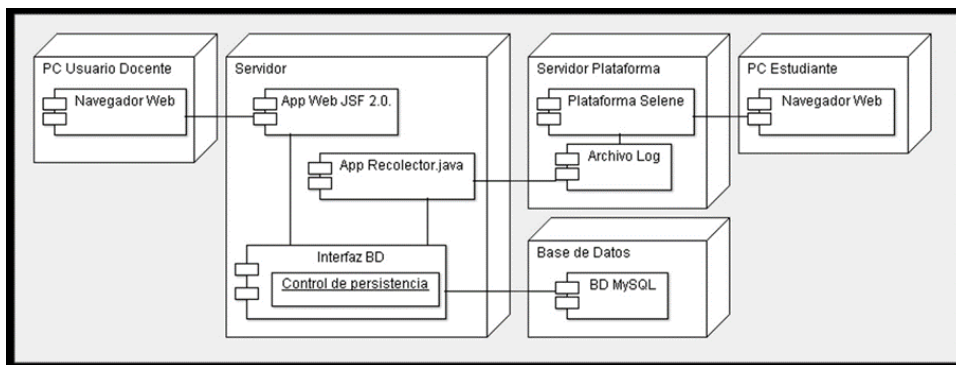


Figura B.8: Diagrama de despliegue

Bibliografía

- [1] Miríada X, Oct. 2016.
- [2] S. Abramovich, C. Schunn, and R. M. Higashi. Are badges useful in education?: it depends upon the type of badge and expertise of learner. *Educational Technology Research and Development*, 61(2):217–232, Apr. 2013.
- [3] I. Aguaded and R. Medina-Salguero. Certificación de los MOOC y su reconocimiento en créditos universitarios. *International Studies on Law and Education*, 23:39–50, 2016.
- [4] R. Al-Shabandar, A. J. Hussain, P. Liatsis, and R. Keight. Analyzing Learners Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach. *IEEE Access*, 6:73669–73685, 2018. Conference Name: IEEE Access.
- [5] C. Alario-Hoyos, I. Estévez-Ayres, M. Pérez-Sanagustín, C. D. Kloos, and C. Fernández-Panadero. Understanding Learners’ Motivation and Learning Strategies in MOOCs. *The International Review of Research in Open and Distributed Learning*, 18(3), May 2017.
- [6] C. Alario-Hoyos, M. Pérez-Sanagustín, C. Delgado-Kloos, H. A. P. G, and M. Muñoz-Organero. Delving into Participants; Profiles and Use of Social Tools in MOOCs. *IEEE Transactions on Learning Technologies*, 7(3):260–266, Sept. 2014.
- [7] S. Alber and L. Debiasi. *Automated Assessment in Massive Open Online Courses*. 2013.
- [8] G. Alexandron, S. Lee, Z. Chen, and D. E. Pritchard. Detecting cheaters in moocs using item response theory and learning analytics. In F. Cena, M. C. Desmarais,

- and D. Dicheva, editors, *Late-breaking Results, Posters, Demos, Doctoral Consortium and Workshops Proceedings of the 24th ACM Conference on User Modeling, Adaptation and Personalisation (UMAP 2016), Halifax, Canada, July 13-16, 2016*, volume 1618 of *CEUR Workshop Proceedings*. CEUR-WS.org, 2016.
- [9] G. Alexandron, J. A. Ruipérez-Valiente, Z. Chen, P. J. Muñoz-Merino, and D. E. Pritchard. Copying@scale: Using harvesting accounts for collecting correct answers in a MOOC. *Computers & Education*, 108:96–114, 2017.
- [10] G. Alexandron, J. A. Ruipérez-Valiente, and D. E. Pritchard. Towards a General Purpose Anomaly Detection Method to Identify Cheaters in Massive Open Online Courses. EdArXiv, 2019.
- [11] A. Arturo Amaya and M. V. Alvarez. Beneficios de los MOOC en la Educación Superior. *Memorias del Encuentro Internacional de Educación a Distancia*, 0(4), Dec. 2015.
- [12] H. S. G. Asep and Y. Bandung. A Design of Continuous User Verification for Online Exam Proctoring on M-Learning. In *2019 International Conference on Electrical Engineering and Informatics (ICEEI), Bandung, Indonesia, July 9-10*, pages 284–289, July 2019.
- [13] Y. Atoum, L. Chen, A. X. Liu, S. D. H. Hsu, and X. Liu. Automated Online Exam Proctoring. *IEEE Transactions on Multimedia*, 19(7):1609–1624, July 2017.
- [14] R. Baker, J. Walonoski, N. Heffernan, I. Roll, A. Corbett, and K. Koedinger. Why students engage in "gaming the system" behavior in interactive learning environments. *Journal of Interactive Learning Research*, 19(2):185–224, 2008.
- [15] R. S. J. d. Baker, A. T. Corbett, I. Roll, and K. R. Koedinger. Developing a generalizable detector of when students game the system. *User Modeling and User-Adapted Interaction*, 18(3):287–314, Aug. 2008.
- [16] Y. Bao. *Detecting Multiple-Accounts Cheating in MOOCs*. PhD thesis, TU Delft, 2017.
- [17] D. Bartalesi-Graf. From a Small Liberal Arts College to the World: Our Blended Courses, SPOC, and MOOCs in Italian Studies. In *Digital Education: Out to the World and Back to the Campus*, Lecture Notes in Computer Science, pages 3–9, Cham, 2017. Springer International Publishing.

- [18] Z. Bocheng, H. Ding, Y. Wang, and R. Huang. MOOC and digital logic course teaching under the complete credit system. In *2015 10th International Conference on Computer Science Education (ICCSE)*, pages 810–813, July 2015.
- [19] C. G. Brinton, M. Chiang, S. Jain, H. Lam, Z. Liu, and F. M. F. Wong. Learning about Social Learning in MOOCs: From Statistical Analysis to Generative Model. *IEEE Transactions on Learning Technologies*, 7(4):346–359, Dec. 2014.
- [20] C. Brooks, C. Stalburg, T. Dillahunt, and L. Robert. Learn with friends: The effects of student face-to-face collaborations on massive open online course activities. In *Proceedings of the Second (2015) ACM Conference on Learning @ Scale - L@S '15*, pages 241–244. ACM Press, 2015.
- [21] M. S. Bulbarela and J. M. M. Fierro. *Tesis Doctoral. Monitorización del progreso en el aprendizaje*. PhD thesis, Universidad Politecnica de Cataluña, Barcelona, Mar. 2008.
- [22] J. Cabero, C. Llorente, and A. Vázquez. MOOC's typologies: Design and educational implications. *Revista de currículum y formación del profesorado*, 18:13–26, 2014.
- [23] E. V. Cano and E. L. Meneses. Los MOOC y la educación superior: la expansión del conocimiento, 2014.
- [24] J. Caudill. The Emerging Formalization of MOOC Coursework: Rise of the MicroMasters. In *EdMedia + Innovate Learning*, pages 1–6, Waynesville, NC, June 2017. Association for the Advancement of Computing in Education (AACE).
- [25] B. Chen, Y. Fan, G. Zhang, and Q. Wang. Examining motivations and self-regulated learning strategies of returning moocs learners. In *Proceedings of the Seventh International Learning Analytics & Knowledge Conference*, pages 542–543, 2017.
- [26] X. Chen, D. Barnett, and C. Stephens. Fad or future: The advantages and challenges of massive open online courses (MOOCs). *Research-to Practice Conference in Adult and Higher Education*, Sept. 2014.
- [27] K. Chorianopoulos and M. Giannakos. Merging learner performance with browsing behavior in video lectures. *WAVE 2013 workshop LAK '13*, 2013.

- [28] I. Claros, R. Cobos, E. Guerra, J. d. Lara, A. Pescador, and J. Sánchez-Cuadrado. Integrating open services for building educational environments. pages 1147–1156. IEEE, Mar. 2013.
- [29] I. Claros, A. Garmendía, L. Echeverría, and R. Cobos. Towards a collaborative pedagogical model in MOOCs. pages 905–911. IEEE, Apr. 2014.
- [30] M. Cocea, A. HersHKovitz, and R. S. J. d. Baker. The Impact of Off-task and Gaming Behaviors on Learning: Immediate or Aggregate? In *Proceedings of the 2009 conference on Artificial Intelligence in Education: Building Learning Systems that Care: From Knowledge Representation to Affective Modelling*, pages 507–514, NLD, July 2009. IOS Press.
- [31] A. Cohen, U. Shimony, R. Nachmias, and T. Soffer. Active learners’ characterization in MOOC forums and their generated knowledge. *British Journal of Educational Technology*, 50(1):177–198, 2019.
- [32] H. Corrigan-Gibbs, N. Gupta, C. Northcutt, E. Cutrell, and W. Thies. Deterring cheating in online environments. 22(6):28:128:23, 2015.
- [33] M. Cote, F. Jean, A. B. Albu, and D. Capson. Video summarization for remote invigilation of online exams. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1–9, Lake Placid, NY, USA, Mar. 2016.
- [34] Coursera. Introducing signature track, 2013.
- [35] M. Dadashzadeh. The Online Examination Dilemma: To Proctor or Not to Proctor? *Journal of Instructional Pedagogies*, 25:1–11, Jan. 2021.
- [36] S. Dendir and R. S. Maxwell. Cheating in online courses: Evidence from online proctoring. *Computers in Human Behavior Reports*, 2:1–10, Aug. 2020.
- [37] E. G. dice. Coursera introduce ‘Signature Track’, una forma de obtener un certificado que verifica la identidad del alumno., Jan. 2013.
- [38] J. Eisenberg. To cheat or not to cheat: effects of moral perspective and situational variables on students’ attitudes. *Journal of Moral Education*, 33(2):163–178, 2004.
- [39] J. P. Espada, C. C. Rodríguez, V. García-Díaz, and R. G. Crespo. Method for analysing the user experience in MOOC platforms. pages 157–162. IEEE, Nov. 2014.

- [40] A. Ezen-Can, K. E. Boyer, S. Kellogg, and S. Booth. Unsupervised modeling for understanding mooc discussion forums: A learning analytics approach. In *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge*, LAK '15, page 146–150, New York, NY, USA, 2015. Association for Computing Machinery.
- [41] G. Fenu, M. Marras, and L. Boratto. A multi-biometric system for continuous student authentication in e-learning platforms. *Pattern Recognition Letters*, 113:83–92, Oct. 2018.
- [42] R. Ferguson and D. Clow. Examining engagement: analysing learner subpopulations in massive open online courses (moocs). In *Proceedings of the fifth international conference on learning analytics and knowledge*, pages 51–58, 2015.
- [43] A. Fox. From MOOCs to SPOCs. *Commun. ACM*, 56(12):38–40, Dec. 2013.
- [44] S. Fu, J. Zhao, H. F. Cheng, H. Zhu, and J. Marlow. T-Cal: Understanding Team Conversational Data with Calendar-based Visualization. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems - CHI '18*, pages 1–13, Montreal QC, Canada, 2018. ACM Press.
- [45] S. Fu, J. Zhao, W. Cui, and H. Qu. Visual Analysis of MOOC Forums with iForum. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):201–210, Jan. 2017.
- [46] C. M. M. García. Diseño e implementación de cursos abiertos masivos en línea (MOOC): expectativas y consideraciones prácticas, 2013.
- [47] J. Gascon, J. Ruipérez, P. Muñoz, H. Pijeira, J. Ruiz, and C. D. Kloos. ANALISE: A Learning Analytics Tool for Open edX, July 2016.
- [48] D. Gašević, S. Dawson, T. Rogers, and D. Gasevic. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. *The Internet and Higher Education*, 28:68–84, 2016.
- [49] M. Gea, R. Montes, B. Rojas, and R. Bergaz. Comunidades Activas de Aprendizaje: hacia la Formación Abierta en las Universidades. *Revista Iberoamericana de Tecnologías del Aprendizaje*, 2:3, 11, Mar. 2014.

- [50] M. P. González, J. Lorés, and A. Pascual. Evaluación Heurística de Sitios Web Académicos Latinoamericanos dentro de la Iniciativa UsabAIPO. *Universitat de Lleida*, pages 143–153, 2010.
- [51] K. K. d. L. Guedes, H. C. Davis, and J. Schulz. Integrating MOOCs into traditional higher education modules: a MOOC-based blend framework. *Research in Learning Technology*, 30, May 2022.
- [52] P. J. Guo, J. Kim, and R. Rubin. How video production affects student engagement: an empirical study of MOOC videos. pages 41–50. ACM, Apr. 2014.
- [53] W. Guo. From SPOC to MPOC – The Effective Practice of Peking University Online Teacher Training. In *2014 International Conference of Educational Innovation through Technology (EITT)*, pages 258–264, Oct. 2014.
- [54] S. Halawa, D. Greene, and J. Mitchell. Dropout prediction in moocs using learner activity features. *Proceedings of the second European MOOC stakeholder summit*, 37(1):58–65, 2014.
- [55] Haus Goffredo, Benvenuto Pasquinelli, Scaccia Daniela, and Scarabottolo Nello. ONLINE WRITTEN EXAMS DURING COVID-19 CRISIS. 2020.
- [56] A. Hegde, N. Ghosh, and V. Kumar. Multiple Choice Questions with Justifications. pages 176–177. IEEE, Dec. 2014.
- [57] A. Hellas, J. Leinonen, and P. Ithantola. Plagiarism in take-home exams: Help-seeking, collaboration, and systematic cheating. In *Proceedings of the 2017 ACM Conference on Innovation and Technology in Computer Science Education - ITiC-SE '17*, pages 238–243. ACM Press, 2017.
- [58] R. Hernández and H. R. Amado-Salvatierra. Towards Full Engagement for Open Online Education. A Practical Experience for a MicroMaster. In *Digital Education: Out to the World and Back to the Campus*, Lecture Notes in Computer Science, pages 68–76, Cham, 2017. Springer International Publishing.
- [59] P. M. Institute, editor. *A guide to the project management body of knowledge / Project Management Institute*. PMBOK guide. Project Management Institute, Newtown Square, PA, sixth edition edition, 2017.
- [60] D. Jaramillo, M. Solarte, and G. Ramirez. Estrategia de seguimiento a las actividades de aprendizaje de los estudiantes en cursos en línea masivos y privados

- (MPOC) con reconocimiento académico en la Universidad del Cauca. In *Septima Conferencia de Directores de Tecnología de Información TICAL 2017*, San Jose, Costa Rica, July 2017.
- [61] D. Jaramillo-Morillo, J. Ruipérez-Valiente, M. F. Sarasty, and G. Ramírez-Gonzalez. Identifying and characterizing students suspected of academic dishonesty in SPOCs for credit through learning analytics. *International Journal of Educational Technology in Higher Education*, 17(1):45, Nov. 2020.
- [62] D. Jaramillo-Morillo, M. S. Sarasty, G. R. Gonzalez, and M. Perez-Sanagustin. Follow-Up of Learning Activities in Open edX: A Case Study at the University of Cauca. In *Digital Education: Out to the World and Back to the Campus*, Lecture Notes in Computer Science, pages 217–222. Springer, Cham, May 2017.
- [63] D. A. Jaramillo-Morillo and Mario Fernando Solarte. Architectural approach for automatic follow up of learning activities in massive open online courses. *Sistemas & Telemática*, 14(37):57–72, Aug. 2016.
- [64] W. Jobe. No university credit, no problem? Exploring recognition of non-formal learning. In *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*, pages 1–7, Oct. 2014. ISSN: 2377-634X.
- [65] W. Jobe and P. Hansson. Mobile Learning for Human Rights in Kenya: The Haki Zangu Case For Non-Formal Learning. In *ICT for Anti-Corruption, Democracy And Education In East Africa*, pages 67–82, Africa, Dec. 2013. Katja Sarajeva.
- [66] S. Joksimović, A. Manataki, D. Gašević, S. Dawson, V. Kovanović, and I. F. De Kereki. Translating network position into performance: importance of centrality in different network configurations. In *Proceedings of the sixth international conference on learning analytics & knowledge*, pages 314–323, 2016.
- [67] S. Joksimović, O. Poquet, V. Kovanović, N. Dowell, C. Mills, D. Gašević, S. Dawson, A. C. Graesser, and C. Brooks. How do we model learning at scale? a systematic review of research on moocs. *Review of Educational Research*, 88(1):43–86, 2018.
- [68] M. Khalil and M. Ebner. Clustering patterns of engagement in massive open online courses (moocs): the use of learning analytics to reveal student categories. *Journal of computing in higher education*, 29(1):114–132, 2017.

- [69] C. D. Kloos, P. Muñoz-Merino, and M. Muñoz-Organero. Extendiendo Google Course Builder mediante Proyectos Realistas en un Curso de Master. *Universidad Carlos III de Madrid*, 2014.
- [70] C. D. Kloos, P. J. Muñoz-Merino, M. Muñoz-Organero, C. Alario-Hoyos, M. Pérez-Sanagustín, H. A. P. G, J. A. Ruipérez, and J. L. Sanz. Experiences of running MOOCs and SPOCs at UC3M. pages 884–891. IEEE, Apr. 2014.
- [71] M. Klusener and A. Fortenbacher. Predicting students’ success based on forum activities in MOOCs. In *2015 IEEE 8th International Conference on Intelligent Data Acquisition and Advanced Computing Systems: Technology and Applications (IDAACS)*, pages 925–928, Warsaw, Poland, Sept. 2015. IEEE.
- [72] M. Laal and M. Laal. Collaborative learning: what is it? *Procedia - Social and Behavioral Sciences*, 31:491–495, 2012.
- [73] C. Lampe, D. Y. Wohn, J. Vitak, N. B. Ellison, and R. Wash. Student use of facebook for organizing collaborative classroom activities. *International Journal of Computer-Supported Collaborative Learning*, 6(3):329–347, 2011.
- [74] A. S. Lan, A. E. Waters, C. Studer, and R. G. Baraniuk. Sparse Factor Analysis for Learning and Content Analytics. *Journal of Machine Learning Research*, pages 1959–2008, July 2013. arXiv: 1303.5685.
- [75] N. Li, H. Verma, A. Skevi, G. Zufferey, J. Blom, and P. Dillenbourg. Watching MOOCs together: investigating co-located MOOC study groups. *Distance Education*, 35(2):217–233, 2014.
- [76] Y. Li and H. Li. Mooc-frs: A new fusion recommender system for moocs. In *2017 IEEE 2nd Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, pages 1481–1488. IEEE, 2017.
- [77] T. Liyanagunawardena, S. Williams, and A. Adams. The impact and reach of MOOCs:a developing countries’ perspective. *eLearning Papers*, (33), May 2013.
- [78] T. Liyanagunawardena, S. Williams, and A. Adams. MOOCs: A Systematic Study of the Published Literature 2008-2012. *International review of research in open and distance learning*, 14:202–227, 2013.

- [79] H. Lynda, B.-D. Farida, B. Tassadit, and L. Samia. Peer assessment in moocs based on learners' profiles clustering. In *2017 8th International Conference on Information Technology (ICIT)*, pages 532–536. IEEE, 2017.
- [80] J. J. Maldonado, R. Palta, J. Vázquez, J. L. Bermeo, M. Pérez-Sanagustín, and J. Muñoz-Gama. Exploring differences in how learners navigate in MOOCs based on self-regulated learning and learning styles: A process mining approach. In *2016 XLII Latin American Computing Conference (CLEI)*, pages 1–12, Oct. 2016.
- [81] P. Mayring. Qualitative Content Analysis: Theoretical Background and Procedures. In *Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods*, Advances in Mathematics Education, pages 365–380. Springer Netherlands, Dordrecht, 2015.
- [82] P. McGee. Supporting Academic Honesty in Online Courses. *Journal of Educators Online*, 10(1), Jan. 2013.
- [83] E. L. Meneses, J. L. S. Sánchez-Serrano, and E. V. Cano. *La expansión del conocimiento abierto: los MOOC*, volume 1. Octaedro.
- [84] G. Migut, D. Koelma, C. G. M. Snoek, and N. Brouwer. Cheat me not: automated proctoring of digital exams on bring-your-own-device. In *Proceedings of the 23rd Annual ACM Conference on Innovation and Technology in Computer Science Education*, ITiCSE 2018, page 388, New York, NY, USA, July 2018. Association for Computing Machinery.
- [85] S. Miranda, G. R. Mangione, F. Orciuoli, M. Gaeta, and V. Loia. Automatic generation of assessment objects and Remedial Works for MOOCs. pages 1–8. IEEE, Oct. 2013.
- [86] A. M. Mutawa. It is time to MOOC and SPOC in the Gulf Region. *Education and Information Technologies*, pages 1–21, June 2016.
- [87] C. G. Northcutt, A. D. Ho, and I. L. Chuang. Detecting and preventing "multiple-account cheating in massive open online courses. *Computers & Education*, 100(C):71–80, Sept. 2016.
- [88] C. G. Northcutt, A. D. Ho, and I. L. Chuang. Detecting and preventing "multiple-account" cheating in massive open online courses. *Computers & Education*, 100:71–80, 2016.

- [89] A. Ortega-Arranz, E. Er, A. Martínez-Monés, M. L. Bote-Lorenzo, J. I. Asensio-Pérez, and J. A. Muñoz-Cristóbal. Understanding student behavior and perceptions toward earning badges in a gamified MOOC. *Universal Access in the Information Society*, 18(3):533–549, Aug. 2019.
- [90] L. Pickard, D. Shah, and J. Simone. Mapping Microcredentials Across MOOC Platforms. In *2018 Learning With MOOCs (LWMOOCs)*, pages 17–21, Madrid, Spain, Sept. 2018.
- [91] H. J. Pijera Díaz, J. Ruiz, J. A. Ruipérez-Valiente, P. Merino, and C. Delgado-Kloos. A Demonstration of ANALYSE: A Learning Analytics Tool for Open edX. pages 329–330, New York, NY, USA, Apr. 2016.
- [92] S. Prathish, A. N. S, and K. Bijlani. An intelligent system for online exam monitoring. In *2016 International Conference on Information Science (ICIS)*, pages 138–143, Kochi, India, Aug. 2016.
- [93] M. Pérez-Sanagustín, J. Maldonado, and N. Morales. Estado del arte de adopción de MOOCs en la Educación Superior en América Latina y Europa. *MOOC-Maker Constr. Manag. Capacit. MOOCs High. Education*, 1, 2016.
- [94] R. A. Pérez-Álvarez, J. Maldonado-Mahauad, K. Sharma, D. Sapunar-Opazo, and M. Pérez-Sanagustín. Characterizing Learners’ Engagement in MOOCs: An Observational Case Study Using the NoteMyProgress Tool for Supporting Self-Regulation. *IEEE Transactions on Learning Technologies*, 13(4):676–688, Oct. 2020.
- [95] H. Qu and Q. Chen. Visual Analytics for MOOC Data. *IEEE Computer Graphics and Applications*, 35(6):69–75, Nov. 2015.
- [96] A. Ramesh, D. Goldwasser, B. Huang, H. Daumé, III, and L. Getoor. Learning Latent Engagement Patterns of Students in Online Courses. In *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence, AAAI’14*, pages 1272–1278, Québec City, Québec, Canada, 2014. AAAI Press.
- [97] M. G. P. Rastrollo. Castaño, C. & Cabero, J. (Coords) (2013). Enseñar y Aprender en Entornos M-Learning. Madrid: Editorial Síntesis. *Píxel-Bit. Revista de Medios y Educación*, 0(45):235–236, July 2014.

- [98] J. A. Ruipérez-Valiente, G. Alexandron, Z. Chen, and D. E. Pritchard. Using multiple accounts for harvesting solutions in moocs. In *Proceedings of the Third (2016) ACM Conference on Learning@Scale*, pages 63–70. ACM, 2016.
- [99] J. A. Ruipérez-Valiente, S. Joksimović, V. Kovanović, D. Gašević, P. J. Muñoz-Merino, and C. Delgado Kloos. A Data-driven Method for the Detection of Close Submitters in Online Learning Environments. In *Proceedings of the 26th International Conference on World Wide Web Companion*, pages 361–368, 2017.
- [100] J. A. Ruipérez-Valiente, S. Joksimović, V. Kovanović, D. Gašević, P. J. Muñoz-Merino, and C. Delgado Kloos. A Data-driven Method for the Detection of Close Submitters in Online Learning Environments. In *Proceedings of the 26th International Conference on World Wide Web Companion, WWW '17 Companion*, pages 361–368, Republic and Canton of Geneva, Switzerland, 2017. International World Wide Web Conferences Steering Committee.
- [101] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, G. Alexandron, and D. E. Pritchard. Using Machine Learning to Detect 'Multiple-Account' Cheating and Analyze the Influence of Student and Problem Features (In press). *IEEE Transactions on Learning Technologies*, 2017.
- [102] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, and C. Delgado Kloos. Detecting and clustering students by their gamification behavior with badges: A case study in engineering education. *International Journal of Engineering Education*, 33(2-B):816–830, 2017.
- [103] J. A. Ruipérez. *Analyzing the behavior of students regarding learning activities, badges, and academic dishonesty in MOOC environment*. <http://purl.org/dc/dcmitype/Text>, Universidad Carlos III de Madrid, 2017.
- [104] J. A. Ruipérez-Valiente, S. Joksimović, V. Kovanović, D. Gašević, P. J. Muñoz-Merino, and C. Delgado. A Data-driven Method for the Detection of Close Submitters in Online Learning Environments. In *Proceedings of the 26th International Conference on World Wide Web Companion - WWW '17 Companion*, pages 361–368, Perth, Australia, 2017. ACM Press.
- [105] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, G. Alexandron, and D. E. Pritchard. Using Machine Learning to Detect 'Multiple-Account' Cheating and Analyze the Influence of Student and Problem Features. *IEEE Transactions on Learning Technologies*, 12:112–122, 2019.

- [106] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, and C. D. Kloos. Improving the prediction of learning outcomes in educational platforms including higher level interaction indicators. *Expert Systems*, 35(6):e12298, 2018.
- [107] R. M. Salguero and I. Aguaded. Los MOOC en la plataforma educativa MiriadaX, 2014.
- [108] C. Sandeen. Assessment's Place in the New MOOC World. *Research & Practice in Assessment*, 8:5–12, 2013.
- [109] C. Sandeen. Integrating MOOCS into Traditional Higher Education: The Emerging "MOOC 3.0" Era. *Change: The Magazine of Higher Learning*, 45(6):34–39, 2013.
- [110] L. Sanz-Martínez, A. Martínez-Monés, M. L. Bote-Lorenzo, J. A. Muñoz-Cristóbal, and Y. Dimitriadis. Automatic group formation in a mooc based on students' activity criteria. In *European Conference on Technology Enhanced Learning*, pages 179–193. Springer, 2017.
- [111] A. Saxena, M. Prasad, A. Gupta, N. Bharill, O. P. Patel, A. Tiwari, M. J. Er, W. Ding, and C.-T. Lin. A review of clustering techniques and developments. *Neurocomputing*, 267:664–681, 2017.
- [112] C. Shi, S. Fu, Q. Chen, and H. Qu. VisMOOC: Visualizing video clickstream data from massive open online courses. pages 277–278. IEEE, Oct. 2014.
- [113] C. Shi, S. Fu, Q. Chen, and H. Qu. VisMOOC: Visualizing video clickstream data from Massive Open Online Courses. In *2015 IEEE Pacific Visualization Symposium (PacificVis)*, pages 159–166, Hangzhou, Apr. 2015.
- [114] E. Sigua, B. Aguilar, P. Pesántez-Cabrera, and J. Maldonado-Mahauad. Proposal for the Design and Evaluation of a Dashboard for the Analysis of Learner Behavior and Dropout Prediction in Moodle. In *2020 XV Conferencia Latinoamericana de Tecnologías de Aprendizaje (LACLO)*, pages 1–6, Loja, Ecuador, Oct. 2020.
- [115] T. Staubitz and C. Meinel. Collaborative learning in MOOCs approaches and experiments. In *2018 IEEE Frontiers in Education Conference (FIE)*, pages 1–9, 2018. ISSN: 2377-634X.
- [116] T. Staubitz, T. Pfeiffer, J. Renz, C. Willems, and C. Meinel. Collaborative Learning in a MOOC Environment. *ICERI2015 Proceedings*, pages 8237–8246, 2015.

- [117] T. Staubitz, R. Teusner, J. Renz, and C. Meinel. An Experiment in Automated Proctoring. In *Conference Proceeding of th 4h European MOOC Summit in Graz, Austria*. Webpage: <http://emoocs2016.eu>, pages 41–53. BoD – Books on Demand, Feb. 2016.
- [118] Stephen Downes. *Connectivism and Connective Knowledge: essays on meaning and learning networks*. National Research Council, Canada, May 2012.
- [119] R. Swaray. An evaluation of a group project designed to reduce free-riding and promote active learning. *Assessment & Evaluation in Higher Education*, 37(3):285–292, 2012.
- [120] P.-N. Tan, M. Steinbach, and V. Kumar. *Introduction to Data Mining, (First Edition)*. Addison-Wesley Longman Publishing Co., Inc., Boston, MA, USA, 2005.
- [121] H. Tang, W. Xing, and B. Pei. Exploring the temporal dimension of forum participation in moocs. *Distance Education*, 39(3):353–372, 2018.
- [122] B. Thoms and E. Eryilmaz. How media choice affects learner interactions in distance learning classes. *Computers & Education*, 75:112–126, 2014.
- [123] E. Tovar, A. Dimovska, N. Piedra, and J. Chicaiza. OCW-S: Enablers for building sustainable open education evolving OCW and MOOC. pages 1262–1271. IEEE, Mar. 2013.
- [124] I. Traoré, Y. Nakkabi, S. Saad, B. Sayed, J. D. Ardigo, and P. M. de Faria Quinn. Ensuring Online Exam Integrity Through Continuous Biometric Authentication. In I. Traoré, A. Awad, and I. Woungang, editors, *Information Security Practices: Emerging Threats and Perspectives*, pages 73–81. Springer International Publishing, Cham, 2017.
- [125] V. Triglianios and C. Pautasso. ASQ: interactive web presentations for hybrid MOOCs. pages 209–210. ACM, May 2013.
- [126] A. A. Turani, J. H. Alkhateeb, and A. A. Alsewari. Students Online Exam Proctoring: A Case Study Using 360 Degree Security Cameras. In *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)*, pages 1–5, Bangladesh, Dec. 2020.

- [127] J. A. R. Valiente. *Analyzing the behavior of students regarding learning activities, badges, and academic dishonesty in MOOC environment*. <http://purl.org/dc/dcmitype/Text>, Universidad Carlos III de Madrid, 2017.
- [128] O. Viberg, A. Mavroudi, Y. Fernaeus, C. Bogdan, and J. Laaksolahti. Reducing free riding: CLASS – a system for collaborative learning assessment. In E. Popescu, A. Belén Gil, L. Lancia, L. Simona Sica, and A. Mavroudi, editors, *Methodologies and Intelligent Systems for Technology Enhanced Learning, 9th International Conference, Workshops*, pages 132–138. Springer International Publishing, 2020.
- [129] R. R. Vila. EVALUACIÓN DE LA CALIDAD PEDAGÓGICA DE LOS MOOC (Assessment the pedagogical quality of the MOOC), 2014.
- [130] T. S. Vinuesa and V. Daza. Estrategia para el seguimiento y evaluación de los aprendizajes en un MOOC de introducción al álgebra. *Digital Education Review*, (25):36–50, 2014.
- [131] A. E. Waters, C. Studer, and R. G. Baraniuk. Bayesian pairwise collaboration detection in educational datasets. In *2013 IEEE Global Conference on Signal and Information Processing*, pages 989–992, Dec. 2013. ISSN: null.
- [132] G. Witthaus, A. Inamorato, M. Childs, Anne-Christin, and Tannhäuser. Validation of Non-formal MOOC-based Learning: An Analysis of Assessment and Recognition Practices in Europe (OpenCred) - EU Science Hub - European Commission, Feb. 2016.
- [133] J.-S. Wong and X. L. Zhang. Messagelens : A visual analytics system to support multifaceted exploration of mooc forum discussions. *Visual Informatics*, 2(1):37–49, Mar. 2018.
- [134] J. Wu. Cluster Analysis and K-means Clustering: An Introduction. In J. Wu, editor, *Advances in K-means Clustering: A Data Mining Thinking*, Springer Theses, pages 1–16. Springer, Berlin, Heidelberg, 2012.
- [135] J. Yardley, M. D. R. Ph.D, S. C. Bates, and J. Nelson. True confessions?: Alumni’s retrospective reports on undergraduate cheating behaviors. *Ethics & Behavior*, 19(1):1–14, 2009.
- [136] A. M. F. Yousef, M. A. Chatti, U. Schroeder, and M. Wosnitza. What Drives a Successful MOOC? An Empirical Examination of Criteria to Assure Design Quality of MOOCs. pages 44–48. IEEE, July 2014.

-
- [137] L. Yuan and S. Powell. MOOCs and disruptive innovation: Implications for higher education, May 2013.
- [138] M. Zapata-Ros. El diseño instruccional de los mooc y el de los nuevos cursos abiertos personalizados. *Revista de Educación a Distancia (RED)*, (45), 2015.
- [139] J. Zheng, L. Huang, S. Li, S. P. Lajoie, Y. Chen, and C. E. Hmelo-Silver. Self-regulation and emotion matter: A case study of instructor interactions with a learning analytics dashboard. *Computers & Education*, 161:104061, Feb. 2021.
- [140] B. J. Zirger, E. Rutz, D. Boyd, J. Tappel, and V. Subbian. Creating pathways to higher education: A cross-disciplinary MOOC with graduate credit. In *2014 IEEE Integrated STEM Education Conference*, pages 1–5, Mar. 2014.