Estrategia de evaluación para la identificación de deshonestidad académica en cursos en linea en ambientes de masividad

Trabajo de Grado

Yehison Javier Cuchumbe Pencua
Paula Andrea Vasquez Artunduaga

Director: Mag. Daniel Alberto Jaramillo Morillo
Co-Director: PhD. Mario Fernando Solarte Sarasty

Departamento de Telemática
Facultad de Ingeniería Electrónica y Telecomunicaciones
Universidad del Cauca
Popayán, Cauca, 2021

Estrategia de evaluación para la identificación de deshonestidad académica en cursos en linea en ambientes de masividad

Yehison Javier Cuchumbe Pencua
Paula Andrea Vasquez Artunduaga

Trabajo de Grado presentado a la Facultad de Ingeniería Electrónica y Telecomunicaciones de la Universidad del Cauca para obtener el título de Ingeniero en Electrónica y Telecomunicaciones

Director: Mag. Daniel Alverto Jaramillo Morillo
Co-Director: PhD. Mario Fernando Solarte Sarasty

Departamento de Telemática
Facultad de Ingeniería Electrónica y Telecomunicaciones
Universidad del Cauca
Popayán, Cauca, 2021

Abstract

Today, there is a growing evolution of communication technologies, giving way to the rise of online education [1]. For this reason, new educational modalities have been conceived, one of them is the MOOC, whose acronym stands for Massive Open Online Course; online courses that as its name indicates have as main characteristic that they are massive and open [2], inherit the advantages of traditional e-learning. Thus, they provide the opportunity to present all types of content, and allow access to education at any time and place [3], which leads many universities and traditional education institutions to be interested in adapting it to their educational modality [4]. The MOOC model is gradually being incorporated into universities and their training programmes through strategies such as MPOC (Mass and Private Online Courses), which are a variant that has been successfully applied in various educational settings. This success and the advantage of having a more controlled environment has led these courses to seek validation as credits within training programs [4]. It is here, where the incursion of these courses as an alternative to obtain academic recognition is present. Therefore, there is a need to control aspects such as possible academic dishonesty, to which these courses are vulnerable because they are developed in online and mass environments [4]. Behaviors such as impersonation, creation of multiple accounts, use of materials or pages that are not permitted, copying or cheating during exams, were combated by implementing measures and proposing a solution that would mitigate such behaviors and take full advantage of the important benefits offered by this form of education [5], [6], [7], [8]. In this degree work, the results obtained from the application of a mechanism for the implementation of an evaluation strategy on Openedx, which allows the identification of behaviors with suspicion of fraud through Learning Analytics techniques in a case study at MPOC

level in the context of the University of Cauca, are evidenced. The mechanism is based on presenting a form with random answers, so that when reading response statistics, patterns of coincidence are sought and fraud behaviors are identified. High expectations are placed on the impact of this proposal and it is hoped that the results obtained will be very useful in this area.

Índice general

Lista de figuras viii

Lista de tablas xii

1. Introducción 1

1.1. Planteamiento del problema 1

1.2. Objetivos 4

1.2.1. Objetivo General 5

1.2.2. Objetivos Específicos 5

1.3. Metodología 5

1.3.1. Generación de la base inicial de conocimiento. 5

1.3.2. Estudio de las diferentes estrategias utilizadas para la detección de fraude. 6

1.3.3. Diseño del mecanismo para la caracterización de comportamientos de fraude aplicable en cursos en línea masivos. 6

1.3.4. Construcción de un mecanismo para la caracterización de comportamientos de fraude en estudiantes. 6

1.3.5. Validación del mecanismo. 7

1.3.6. Publicación 7

1.4. Resultados Alcanzados 7

1.5. Contenido de la monografía 9

2. Estado del Arte **10**

2.1. Contexto 10

2.1.1. MOOC 10

2.1.1.1. Variantes 13

2.1.1.2. Plataformas *MOOC* 14

2.2. Revisión Sistemática 17

2.3. Trabajos Relacionados 22

2.3.1. Fraude en *MOOC* y estrategias para disminuir la deshonestidad académica en ambientes educativos en línea 22

2.3.2. Analíticas del aprendizaje (*Learning Analytics*) aplicado en *MOOCs*. 23

2.3.3. Estrategias Evaluativas 25

2.4. Conclusiones del estado del arte 27

2.5. Brechas 28

3. Mecanismo para la detección de estudiantes con sospecha de fraude a través de estrategias evaluativas. **29**

3.1. Recolección de datos 29

3.1.1. Plataforma de aprendizaje 29

3.1.2. Características básicas de Open EdX 30

3.1.2.1. Sistema de Gestión de Aprendizaje (*LMS, Learning Management System*) 30

3.1.2.2. Sistema de Gestión de Contenidos (*CMS, Content Management System*) 32

3.1.2.3. Contenidos 32

3.1.3. Extracción de datos desde la plataforma Selene 36

3.2. Detección de estudiantes con sospecha de fraude 39

3.2.1. Criterios para la identificación de estudiantes 39

3.3. Herramienta de Apoyo a Docentes 41

3.3.1. Descripción de la Arquitectura 43

3.3.1.1. Vista de escenarios 43

3.3.1.2. Vista Lógica 47

3.3.1.3. Vista de Procesos 48

3.3.1.4. Vista de Implementación 49

3.3.1.5. Vista de Despliegue 50

4. Caso de Estudio **54**

4.1. Descripción del curso piloto 54

4.2. Estrategia Evaluativa 57

4.3. Aplicación sobre el caso de estudio 61

5. Resultados **64**

5.1. Aplicación Web: *DetectApp* 64

5.2. Uso de la herramienta sobre el caso de estudio 70

5.2.1. Unidad Temática II - Examen 1. 71

5.2.2. Unidad Temática II - Examen 2. 71

5.2.3. Unidad Temática II - Examen 3. 72

5.2.4. Unidad Temática II - Examen 4. 73

5.2.5. Unidad Temática II - Examen Final 73

5.3. Análisis de Resultados. 74

6. Conclusiones **87**

Bibliografía **90**

Anexos **97**

A. Manual de instalación de la aplicación web DetectApp **98**

B. Manual de Usuario de la aplicación DetectApp. **100**

B.1. Inicio de sesión. 100

B.2. Seleccionar curso. 101

B.3. Ver listado de exámenes. 102

B.4. Consultar informe por porcentaje de coincidencias. 103

B.4.1. Listar estudiantes por porcentaje de coincidencias. 103

B.4.2. Listar estudiantes detectados con sus porcentajes de coincidencia. 103

B.4.3. Ver el listado completo de estudiantes. 104

B.4.4. Ver el listado de estudiantes detectados que coinciden con los tres parámetros de detección. 105

B.5. Consultar informe por grupos de trabajo. 106

B.5.1. Listar estudiantes detectados en el grupo de trabajo. 106

B.5.2. Ver el listado de estudiantes detectados que coinciden con los tres parámetros de detección. 107

C. Artículo **108**

Índice de figuras

2.1. Resultados búsqueda tipo artículo con palabra clave *MOOC* 19

2.2. Resultados búsqueda tipo artículo con palabra clave *SPOC* 20

2.3. Resultados búsqueda tipo artículo con palabra clave MPOC 20

2.4. Ciclo de Sobre expectacion de *MOOC* segun la empresa Gartner [9]. . 21

3.1. Captura de pantalla del LMS de Open edX (Contenidos) 31

3.2. Captura de pantalla, interfaz para el instructor LMS de Open edX . . 31

3.3. Captura de pantalla Studio edX 32

3.4. Componentes en una unidad 33

3.5. Página de bienvenida de la Plataforma Selene 35

3.6. Cursos *MOOC* ofrecidos por la Plataforma Selene 35

3.7. Presentación del curso para el Usuario desde la Plataforma Selene . . 35

3.8. Arquitectura De Extracción de Datos 38

3.9. Arquitectura General del Mecanismo 42

3.10. Diagrama de Casos de Uso Implementación Mecanismo de Identificación 44

3.11. Diagrama de Clases para el Mecanismo de Identificación 47

3.12. Diagrama de Actividades del Sistema 48

3.13. Diagrama de componentes Modulo de Recolección de Datos 49

3.14. Diagrama de componentes Modulo de Visualización 49

3.15. Diagrama de Despliegue del Sistema 50

3.16. Interfaz de exámenes DetectApp 53

4.1. Primera pregunta para cada versión del examen según la estrategía evaluativa 59

4.2. Respuesta a una pregunta en un grupo de trabajo con un examen y distitnas versiones 60

5.1. Interfaz Inicio de Sesión de la aplicación *DetectApp* 65

5.2. Interfaz Cursos de la aplicación *DetectApp* 65

5.3. Interfaz Exámenes del caso de estudio *DetectApp* 66

5.4. Interfaz listado de estudiantes con coincidencias en respuestas superior al 90 % *DetectApp* 67

5.5. Interfaz listado de estudiantes con coincidencias en respuestas superior al 90 % detallada en *DetectApp* 67

5.6. Interfaz listado completo de estudiantes con sus respuestas en *DetectApp* 68

5.7. Listado de estudiantes que cumplen los tres parámetros de detección en *DetectApp* 69

5.8. Listado de grupos de trabajo por IP de detección en *DetectApp* 69

5.9. Listado de grupos de trabajo por IP de detección en *DetectApp* 70

5.10. Captura e Identificación de estudiantes sospechosos en la aplicación *DetectApp* 74

5.11. Cantidades de estudiantes sospechosos con un 90 % o mas de coincidencias en la aplicación *DetectApp* a lo largo del caso de estudio. . . . 77

5.12. Cantidad de grupos de trabajo sospechosos en la aplicación *DetectApp* a lo largo del curso. 78

5.13. Gráfica Comparativa de la cantidad detectados VS la cantidad de estudiantes en los grupos de trabajo por *DetectApp* 79

5.14. Cantidad de estudiantes que reinciden en formar grupos de trabajo y son detectados por *DetectApp* 80

5.15. Gráfica de interacciones a través de la duración del curso. 83

5.16. Cantidad de estudiantes que reinciden en formar grupos de trabajo y son detectados por *DetectApp* 84

5.17. Gráfico de participación en foros en el curso 85

5.18. Cantidad de tiempo dedicado a visualizaciones de un estudiante no detectado (azul) VS uno reincidente (rojo) en formar grupos de trabajo y son detectados por *DetectApp* 85

B.1. Interfaz Inicio de Sesión de la aplicación *DetectApp* 101

B.2. Interfaz Listado de cursos disponibles en *DetectApp* 102

B.3. Interfaz Listado de exámenes en *DetectApp* 102

B.4. Interfaz Listado de estudiantes con coincidencias superior al 90 % en *DetectApp* 103

B.5. Interfaz Listado con porcentajes *DetectApp* 104

B.6. Interfaz Listado completo en *DetectApp* 104

B.7. Interfaz Estudiantes detectados por *DetectApp* 105

B.8. Interfaz Informe por IP *DetectApp* 106

B.9. Interfaz Grupos de trabajo en *DetectApp* 106

B.10.Interfaz Estudiantes detectados *DetectApp* 107

Índice de cuadros

2.1. Resultados búsqueda *MOOC* 18

2.2. Resultados búsqueda *SPOC* 19

2.3. Resultados búsqueda MPOC 20

Capítulo 1

Introducción

1.1. Planteamiento del problema

La educación abierta y a distancia ha permitido ampliar el alcance de la educación tradicional, acortando barreras geográficas y temporales. Además, ha contribuido a solucionar problemas como el límite de oferta de cupos en los cursos presenciales, dándole a las personas la facilidad y la comodidad de estudiar desde casa sin tener que trasladarse al sitio donde se imparten [10], contando también con ventajas dentro de las que se pueden mencionar, la personalización completa en el plan de estudios, fomentar la autonomía, promover una elevada interacción y favorecer el equilibrio perfecto entre la vida profesional y personal [11].

El alcance de este modelo educativo se vio ampliado en gran medida con la llegada del Internet [1]. Con el cual surgieron los cursos en línea ó lo que hoy se conoce como el aprendizaje en línea (*e-learning*), donde el estudiante puede inscribirse y obtener materiales como lecturas, presentaciones, *E-books*, entre otros; todo esto para que al final se responda a ciertas pruebas para evaluar si el estudiante habrá adquirido las competencias que el curso quería impartir, al igual que en los cursos presenciales [12]. Con la nueva tendencia de cursos en línea, el MIT en el año 1999 lanzó su proyecto *OpenCourseWare* [13], cuyo objetivo era poner en la web el contenido de algunas de sus asignaturas impartidas de manera presencial, lo que dió paso a una

nueva modalidad de educación: los Cursos en Línea Masivos y Abiertos (*Massive Open Online Course - MOOC*). Cursos en abierto, que abarcan una gran cantidad de estudiantes y que además, son ofrecidos por las más prestigiosas universidades. Por lo que rápidamente se hicieron muy populares [14].

En el 2008 Georges Siemens y Stephen Downes crean en la universidad de Manitoba, el primer curso en línea que recibe el apelativo de MOOC. El curso contó con una duración de 12 semanas y una participación de 2.300 estudiantes de diferentes partes del mundo [15]. En el 2011 Sebastian Thrun y Peter Norvig crearon el curso "*Introduction to Artificial Intelligence*" con un total de 160.000 personas inscritas. Este fue el detonante para que este modelo educativo empezará su auge [16]. Con el éxito de esta modalidad en línea comenzaron a surgir plataformas dedicadas a impartir este tipo de cursos, contando con una gran participación de estudiantes alrededor del mundo [17].

Los MOOCs presentan muchas ventajas para los estudiantes, como lo son: el ser gratuitos, lo que permite que estudiantes de recursos económicos limitados puedan acceder a ellos; una gran variedad de temas, ya que hay muchos cursos de diferentes asignaturas impartidos semanalmente; son flexibles, pues no tienen un horario fijo, se pueden aprovechar desde cualquier sitio que cuente con una conexión web; permiten un mayor acceso a contenidos de calidad, dada su característica de masivos; favorecen y apoyan el acercamiento a nuevas áreas del conocimiento [17]; permiten acceder a nuevo conocimiento que investigadores o profesores han desarrollado; son de fácil admisión por su característica de abiertos, se puede pertenecer a muchos cursos a la vez y algunos presentan la oportunidad de ser validados por créditos de cursos presenciales.

Por otra parte, también existen un gran número de ventajas para las instituciones que los imparten, éstas tienen la oportunidad de ampliar la cobertura, ampliar su oferta académica, dar a conocer su nombre a nivel mundial y generar una posible nueva forma de ingreso económico [18].

Estos beneficios han despertado el interés de las universidades que desean poder articular este modelo educativo a su modelo tradicional y beneficiarse de las ventajas que presentan. Sin embargo, existen muchos desafíos que se deben superar, como:

las plataformas con soporte masivo, la evaluación a un número elevado de personas, el acompañamiento al estudiante en el desarrollo del curso, entre otros. Actividades que debido a la masividad, aumentan su complejidad, y aún más cuando los cursos son reconocidos académicamente [18].

Por otra parte, desde el año 2012, en donde el auge de los MOOC creció, el desafío fue garantizar que el estudiante no presente comportamientos fraudulentos. Si bien, en la modalidad tradicional donde el alumno presenta sus evaluaciones de manera presencial y con supervisión pueden presentarse fraudes, con la modalidad en línea las probabilidades aumentan y con éstas la necesidad de métodos para identificarlos y evitarlos o disminuirlos [3].

Al incorporar los MOOC al ámbito universitario y usarlo como medio para otorgar reconocimiento y/o créditos académicos, el ambiente académico tradicional obtiene las principales ventajas de los MOOC [4], pero con ellos también sus principales problemas. Si en los MOOC que no ofrecen reconocimiento académico pero si una certificación, una parte de los alumnos estaban dispuestos a hacer trampa [19], con el reconocimiento académico y/o créditos académicos, el número de intentos de fraude aumenta [4]. Con el incremento de la recompensa existe un incremento en el número de estudiantes con comportamientos deshonestos [20].

Mas estudiantes con predisposición a tener comportamientos deshonestos, conlleva a que los MOOCs deban implementar nuevas técnicas para confirmar la identidad del estudiante [21], erradicar el plagio [19] y evitar la colaboración ilícita. Pero la implementación de estos métodos también conlleva al incremento en los costos de sostenibilidad del curso [4]. Por lo tanto los comportamientos fraudulentos detienen o dificultan la incorporación de los MOOC al ámbito universitario.

Aunque en la actualidad existan maneras de identificar el fraude, estas aún no presentan una buena fiabilidad y garantía de que la persona no la cometa. Por ejemplo en la India se utilizó como método anti-fraude el llamado “Código de Honor”, el cual es un texto que se lee antes de presentar el examen en línea con el fin de que el evaluado se comprometa a no cometer fraude [22]. O el uso de advertencias persuasivas que amenazaban al estudiante en ser expulsado del curso al detectarse comportamiento fraudulento [5].

El desafío de garantizar la honestidad académica en los MOOCs es tal, que algunas de las plataformas que ofrecen cursos en línea han comenzado a contratar con proveedores de servicios que brindan la oportunidad de presentar exámenes presenciales supervisados [4], pero aunque los proveedores tengan presencia en gran parte del mundo, dicha medida solo limita la naturaleza de los MOOCs, ser abiertos, masivos y en línea.

Como se puede ver, algunas estrategias para la disminución de fraude, no están encaminadas a identificar comportamientos con sospecha de fraude sino a persuadir al estudiante a que no lo cometa, sin entender el trasfondo de este comportamiento. Otras estrategias antifraude basan su efectividad en la buena fe del evaluado o en métodos de detección del fraude como los supervisados por cámaras o personalmente, pero no contemplan los costos asociados tanto para las instituciones que imparten los cursos, como para los estudiantes que dependen de recursos computacionales y gran ancho de banda de conexión a internet. Pero no hemos trabajos que hagan uso de métodos evaluativos con el fin de identificar y dificultar las actividades fraudulentas.

Con esto en mente, surge la siguiente pregunta: ¿Cómo identificar comportamientos de fraude mediante estrategias evaluativas aplicadas a los MOOC?.

Creemos que las estrategias evaluativas pueden tanto ayudar a identificar el fraude en MOOC, como también evitarlo. Con el fin de dar respuesta a la pregunta, nos basaremos en métodos de creación de evaluaciones para cursos en línea y en técnicas de *Learning Analytics* que ayuden a identificar conductas o comportamientos deshonestos de los estudiantes.

1.2. Objetivos

A continuación se describen los objetivos planteados para el desarrollo del trabajo de grado:

1.2.1. Objetivo General

- Proponer una estrategia de evaluación que permita la identificación de conductas con sospecha de fraude en cursos en línea en ambientes de masividad.

1.2.2. Objetivos Específicos

- Identificar prácticas de evaluación relacionadas con la deshonestidad académica en MOOC.
- Desarrollar un mecanismo para la implementación de la estrategia evaluativa sobre *Open edX* para la identificación de conductas con sospecha de fraude.
- Verificar las estrategias y el mecanismo propuesto en un caso de estudio a nivel de MPOC en el contexto de la Universidad del Cauca mediante una prueba de concepto.

1.3. Metodología

La estructura básica de las actividades propuestas para el desarrollo del presente trabajo, toma como referencia la descomposición jerárquica WBS (*Work Breakdown Structure*) sugerida en la metodología PMBOK (*Project Management Base of Knowledge*) por el PMI (*Project Managament Institute*), específicamente en el área de gestión del alcance (*Scope Management*) [23].

A continuación se realiza un resumen de actividades para cada uno de los paquetes de trabajo propuestos.

1.3.1. Generación de la base inicial de conocimiento.

- Definición del objetivo de la vigilancia (planeación).
- Construcción del *corpus* (búsqueda y capacitación).

- Análisis del *corpus* (análisis).

1.3.2. Estudio de las diferentes estrategias utilizadas para la detección de fraude.

- Búsqueda de técnicas existentes para la detección de fraude.
- Búsqueda de estrategias evaluativas para prevenir el fraude en entornos educativos.
- Análisis de las técnicas y estrategias encontradas.
- Selección de las técnicas y estrategias más adecuadas para su adaptación en el ámbito MOOC.

1.3.3. Diseño del mecanismo para la caracterización de comportamientos de fraude aplicable en cursos en línea masivos.

- Estudio de las técnicas y estrategias más adecuadas de identificación de fraude para la adaptación en el entorno *MOOC*.
- Selección de las estrategias evaluativas que permitan la identificación de comportamientos con sospecha de fraude.
- Diseño de mecanismo para la caracterización de los comportamientos de los estudiantes.

1.3.4. Construcción de un mecanismo para la caracterización de comportamientos de fraude en estudiantes.

- Análisis de las técnicas que permitan la construcción del mecanismo.
- Elección de la técnica apropiada.

- Implementación de la técnica con base en datos de una plataforma *MOOC*.

1.3.5. Validación del mecanismo.

- Diseño del caso de estudio.
- Estructuración del plan de evaluación.
- Selección e implementación de las métricas de evaluación.
- Pruebas.
- Análisis de datos de la evaluación.

1.3.6. 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:

- Creación de un marco teórico en el que se muestra la terminología empleada en el desarrollo del proyecto.
- Resultados del estado del conocimiento, donde se muestran las estrategias de detección de fraude identificadas, y la bibliografía relacionada y empleada con el desarrollo del proyecto.
- Aproximación arquitectónica del mecanismo para la identificación de comportamientos de fraude en *MOOCs* a partir del modelo de 4+1 vistas.

- Implementación de un prototipo del mecanismo con base en la arquitectura propuesta.
- Resultados de pruebas del mecanismo sobre el *MPOC* Curso virtual "Introducción al emprendimiento con *Lean StartUp*"de la Universidad del Cauca.
- Borrador de artículo con los resultados obtenidos del caso de estudio.

1.5. Contenido de la monografía

Este documento se divide en los capítulos que se describen a continuación.

- Capítulo 1. INTRODUCCIÓN.
  Definición del problema y la estructura general del desarrollo del proyecto de investigación.

- Capítulo 2. ESTADO DEL ARTE.
  En el cual se presenta un resumen de los trabajos relacionados existentes y experiencias previas de otros investigadores acerca del seguimiento de las actividades de aprendizaje en *MOOC*.

- Capítulo 3. DISEÑO Y CONSTRUCCIÓN DE UN MECANISMO.
  Se ilustra el proceso de diseño e implementación de un prototipo basado en una aproximación arquitectónica que soporte el seguimiento de las actividades de aprendizaje de los estudiantes en *MOOC* con base en el conjunto de identificadores encontrados.

- Capítulo 4. CASO DE ESTUDIO.
  En donde se describe el curso piloto, la estrategia evaluativa propuesta y la aplicación del mecanismo y estrategia evaluativa sobre el caso de estudio.

- Capítulo 5. PRUEBAS Y RESULTADOS.
  En donde se expone la adecuación de la plataforma de aprendizaje, curso piloto y las herramientas necesarias que dan soporte a las pruebas realizadas al prototipo diseñado y creado en el capítulo anterior.

- Capitulo 6. CONCLUSIONES Y TRABAJO FUTURO.
  Por último, se analizan los resultados del trabajo realizado, se detallan las principales contribuciones obtenidas durante el ciclo del proyecto y se expone un conjunto de recomendaciones importantes para el desarrollo de trabajos futuros.

Capítulo 2

Estado del Arte

Este capítulo presenta los resultados obtenidos de la aplicación de una revisión sistémica para la generación y síntesis del estado actual de conocimiento, donde se muestran los principales trabajos relacionados y empleados con el desarrollo del proyecto. El estado del arte ha permitido encontrar brechas existentes en cuanto a seguimiento en *MOOC* y además, ha permitido la identificación de las estrategias evaluativas presentadas en este documento.

2.1. Contexto

A continuación se realiza un recorrido por los conceptos que fueron estudiados y empleados para la realización del proyecto.

2.1.1. MOOC

Para que un curso en línea sea considerado un *MOOC* debe cumplir con dos requisitos principales: Debe ser masivo; es decir que pueda aceptar a cientos o miles de estudiantes sin que eso afecte su funcionamiento. Deben ser abierto; esta característica implica que el acceso a los contenidos debe ser gratuito y que no existan

restricciones de acceso para un usuario con acceso a Internet, brindándole la oportunidad de acceder a educación en cualquier momento y lugar [14]

Los *MOOC* están diseñados para la participación de un gran número de estudiantes dispersos geográficamente. Es un curso que integra redes sociales y recursos en línea accesibles que son facilitados por profesionales líderes en el campo de estudio. Los *MOOC* basan su resultado en el compromiso de los alumnos que se auto organizan de acuerdo con los objetivos de aprendizaje, los conocimientos y habilidades previos y los intereses comunes [24].

Según los modelos de organización, la metodología y los objetivos planteados, los *MOOC* pueden dividirse en distintos tipos, dos de los más populares son los *cMOOC* y los *xMOOC* [25]. Los *cMOOC* son cursos masivos, abiertos y en línea cuya filosofía es el aprendizaje conectivista [26], desarrollado en una plataforma en línea donde un grupo de personas socializa, comparte y construye conocimiento. El enfoque en este tipo de *MOOC* es el estudiante, que a partir de material compartido por el profesor, usualmente con el fin de motivar al alumno a crear sus propios aportes, hacen participaciones, discusiones, o entradas en foros, con el propósito de comparar y retro alimentarse del conocimiento colectivo del grupo de personas participantes [26].

Los *xMOOC* son cursos universitarios tradicionales llevados a una plataforma en línea, el actor principal es el profesor quien es el encargado de seleccionar el material impartido en el curso así como también las fechas y modos de evaluación con el fin de verificar la correcta interpretación y aprendizaje por parte de los estudiantes. El profesor generalmente sube un vídeo explicativo acerca de la temática impartida y el deber de los alumnos es visualizarlo y seguirlo para al final presentar actividades con el fin de validar los objetivos del curso. El objetivo de los *xMOOC* es que el alumno aprenda un conocimiento impartido por un educador profesional y al final poder certificar que el conocimiento fue adquirido [25].

Sin embargo, es pertinente aclarar cuándo un curso es o no es un *MOOC*, teniendo en cuenta una serie de características que normalmente necesitan para que un curso sea considerado un *MOOC*. Estas características son presentadas a continuación:

Curso: Debe tener objetivos de aprendizaje que deben alcanzar los estudiantes después de ciertas actividades de aprendizaje dentro de un período de tiempo determinado (por lo tanto, debe tener un principio y un final). Debe tener, recursos, guías, ejercicios, pruebas y exámenes, que permitan desarrollar un procesos formativos en donde al final los conocimientos adquiridos por los estudiantes sean evaluados [27].

Abierto: El hecho de ser abierto, tiene varios significados en los *MOOC*. Por un lado, el curso debe estar disponible para todos y no debe solicitar requisitos previos, como la posesión de una calificación o un nivel de rendimiento en estudios anteriores. Por otro lado, el acceso a los recursos educativos (vídeos, apuntes de clase) debe ser gratuito (pero otras cosas, como poder hacer preguntas directas al profesor, corregir las actividades u obtener un certificado al final del curso pueden tener un costo económico). Finalmente, "abierto.ª menudo también refiera a el curso que hace un uso extensivo de su contenido, en otras palabras, el contenido generado se publica abiertamente para que otros puedan reutilizarlo. Esta última interpretación es la que menos se cumple, ya que en la actualidad, los *MOOC* más exitosos están organizados por empresas, como Coursera o Udacity, que tienen poco interés en compartir de manera pública el contenido de sus cursos [27].

En línea: El curso es tomado de forma remota a través de Internet y no requiere asistencia física en un aula. Esta función es esencial para que cualquier persona del mundo con conexión a Internet pueda participar en estos cursos [27].

Masivo: Debe permitir el acceso a un gran número de estudiantes, mucho más grande que una clase presencial o un curso en línea tradicional. Además, el curso debe estar preparado para aceptar cambios en el número de estudiantes en varios órdenes de magnitud [27].

La idea básica de *MOOC* es ser un curso en línea que no tiene restricción de acceso, lo que lo hace de naturaleza abierta, tener un gran número de participantes lo que lo hace masivo, tener objetivos pedagógicos lo que lo hace un curso y servir como medio para el compartir y adquirir conocimiento [26].

2.1.1.1. Variantes

A continuación se presentan dos de las variantes que tienen los *MOOC*, las cuales han surgido como solución al deseo de ser incorporados en el ámbito universitario como alternativa a los cursos tradicionales.

SPOC

Las dos primeras letras de *SPOC* (Small Private Online Course) son intencionalmente opuestas a las dos primeras letras de *MOOC* (Massive Open Online Course). Los *MOOC* son masivos y abiertos, enseñan a miles de estudiantes a la vez, mientras que los *SPOC* son privados, lo cual hace que limiten el tamaño, son pequeños, pero no hay un número específico de estudiantes que permitan caracterizar los *SPOC*. Ambos funcionan en la modalidad de cursos en línea [28].

Pero más allá de las siglas, un *SPOC* es un curso en línea que ha sido de gran impacto y ofrecido en diferentes países. Los *SPOC* son una continuación y expansión de lo que ya esta implementado en el aprendizaje en línea, son una alternativa para integrar cada vez más los *MOOC* hacia el ámbito universitario [28].

Mientras que los *MOOC* deben escalar sin importar el número de participantes y sin contar con recursos adicionales, los *SPOC* generalmente limitan su número de inscripciones, implementando procesos de solicitud competitivos o mediante el cobro de una tarifa que representan más recursos a la plataforma que los dicta, lo que significa que el trabajo individual puede ser analizado y acreditado con más confianza [28].

Los estudiantes de cursos *SPOC* pueden participar en diálogos de realimentación con sus maestros; pueden hacer preguntas y esperar una respuesta de calidad; además de participar en actividades que son facilitadas por personas reales, siendo esto, una diferencia con los *MOOC*, ya que en los últimos dificulta el hecho de prestar atención personalizada, debido al gran número de estudiantes que los conforman [28].

Los *SPOC* por su modalidad, dan paso a modelos pedagógicos como la clase invertida; en la que los estudiantes completan el aprendizaje normalmente cubierto en el aula, en su propio tiempo (viendo videos y/o accediendo a recursos), mientras que

el tiempo en clases se dedica a actividades prácticas y aprendizaje personalizado e interactivo, lo que los lleva a una comprensión más profunda de los contenidos [29].

Se concluirá con la siguiente analogía: "Si un *MOOC* es como una bicicleta de ejercicio, el SPOC agrega un entrenador personal" [30].

MPOC

Los MPOC de sus siglas en inglés Massive Private Online Course o Cursos en Línea Masivos y Privados, fueron cursos en línea propuestos en la universidad de Beijing en el año 2006 a partir de la idea de *SPOC* [31]. Nacen con el fin de capacitar a profesores como facilitadores virtuales en K-12 [32], los MPOC tienen la misma naturaleza virtual de los *SPOC* Y *MOOC*, pueden compartir el mismo plan de estudios y objetivos de curso. Al igual que los *SPOC*, los MPOC presentan un acceso restringido y controlado a su contenido lo que les confiere una característica privada, pero difieren en la cantidad de estudiantes que admiten [33], lo cual representa una desventaja para los MPOC, ya que en ellos el número de admitidos es mucho más alto y el docente ya no puede tener una relación de uno a uno con sus estudiantes, lo que dificulta el seguimiento a estos y por tanto poco acompañamiento en el desarrollo del curso [34]. Por ende, lo que diferencia a los MPOC de sus símiles son los límites de clases (gratuitas o pagas) y las proporciones entre maestros y alumnos.

Los MPOC tienden a tener más facilitadores en línea que los *MOOC*, con el fin de poder hacer un seguimiento al proceso de aprendizaje de los alumnos, actualmente es el modelo más usado en cuanto a implementación en ambientes universitarios [31].

En conclusión, los MPOC son cursos *MOOC* de menor alcance donde la aceptación a estos dependen del criterio de aceptación académica en el curso, pero su número de alumnos admitidos es lo suficientemente alto en comparación con un curso en línea tradicional lo que le permite conservar su característica de masivo [31].

2.1.1.2. Plataformas *MOOC*

Dado el importante auge que los cursos *MOOC* han tomado y la acogida que han tenido en muchas universidades, el número de plataformas web que ofrecen este tipo

de cursos ha crecido en la misma medida. A continuación, mencionan las plataformas más utilizadas, teniendo como referencia el informe Scopeo 2013 [35]:

- **Coursera**
  Es una plataforma de cursos *MOOC* que nació 2011, de la mano de docentes de la Universidad de Stanford, con el objeto de proporcionar cursos gratuitos en todo el mundo. Forman parte de Coursera más de 148 instituciones educativas de todo el mundo [36].

  La plataforma Coursera esta disponible en diferentes idiomas (inglés, español, portugués, mandarín, francés y ruso). El registro en la página es gratuito, una vez que la persona se ha registrado, puede acceder a un listado con los cursos disponibles los cuales pueden ser gratuitos o de paga y realizar aquél que desee. La plataforma cuenta además con diversos foros y artículos de ayuda en los que puede solicitar asistencia ante determinados problemas y encontrar respuestas para su solución [36].

- **MiriadaX**
  En un intento de la Comisión Europea en el año 2012 por utilizar eficientemente las TIC y los recursos educativos abiertos para el aprendizaje y efectos pedagógicos, surgen distintas iniciativas *MOOC*, entre las que destaca el proyecto MiriadaX, que es creado con el objetivo de promover el acceso al conocimiento de forma libre y sin restricciones [37].

  La plataforma MiriadaX es una iniciativa libre y gratuita, que cuenta con la participación de universidades reconocidas de Iberoamérica. Es implementada en dos idiomas (portugués y español), presentando una interfaz del curso sencilla, clara e intuitiva. Cuenta con una sección de soporte en el que puede solicitarse ayuda o acceder a las preguntas y respuestas frecuentes del alumnado. Cada módulo finaliza con un sistema de evaluación y al terminar el curso MiriadaX ofrece dos tipos de certificaciones; la certificación de participación y la certificación de superación [37].

- **EdX**
  Es una conocida plataforma de cursos *MOOC* creada en 2012 de forma conjunta por el MIT (Instituto Tecnológico de Massachusett) y la universidad de

Harvard. Se trata de un proyecto sin ánimo de lucro por lo que el acceso a los cursos es gratuito, no obstante, si desea recibir algún tipo de certificación hay que realizar pagos, es además, un proyecto "open-source" al que durante estos años, se han unido más de 90 instituciones educativas de todo el mundo [36].

En su catálogo formativo EdX ofrece información muy detallada respecto a las características de sus cursos *MOOC*. La mayoría de esta está en inglés, siendo el español la segunda lengua con mayor presencia. Los cursos están organizados por niveles de complejidad, los cuales son; iniciación, intermedio y avanzado. La plataforma cuenta además con la posibilidad de acceso a través de dispositivos móviles utilizando una aplicación disponible para sistema operativo Android y iOS [36].

EdX en su proyecto de código abierto presenta una plataforma en línea, la cual puede descargarse e implementar en un servidor local, dicha plataforma es llamada OpenEdX [38]. Esta plataforma permite crear, implementar, probar y analizar cursos en línea en cualquier parte del mundo, de ahí que sea usada por numerosas universidades.

La adopción de este modelo representa para las universidades un desafío tecnológico y pedagógico [31], el cual sortean por medio de la implementación de plataformas que necesiten bajos recursos tecnológicos y económicos. Un ejemplo de esta implementación lo muestra la Universidad del Cauca, con la adaptación de uno de sus cursos presenciales con reconocimiento académico al modelo MPOC en el año 2016, basado en una instancia Open EdX.

Para esta adaptación la Universidad del Cauca, en el primer periodo del 2016, lanzó inicialmente el curso "Astronomía Cotidiana", perteneciente al componente de Formación Integral Social y Humana FISH, con una participación de 400 estudiantes. El curso fue implementado en una instancia OpenEdX, la cual en la Universidad del Cauca fue nombrada como SELENE [31]. Debido al éxito de su implementación [33], esta plataforma cuenta actualmente con varios cursos de tipo MPOC y *SPOC*, sirviendo como solución para el problema de cupos en el componente FISH.

2.2. Revisión Sistemática

Para la generación de la base de conocimiento se adoptó una perspectiva de Revisión Sistemática, la cual se define como "metodología que implica un resumen crítico y reproducible de los resultados de las publicaciones disponibles sobre un mismo tema o pregunta concreta. Con el fin de mejorar la escritura científica, se expone de una forma estructurada la metodología para la realización de una revisión sistemática: la búsqueda, detección, análisis y comunicación de informaciones orientadas a la toma de decisiones sobre amenazas y oportunidades externas en el ámbito de la ciencia y tecnología" (E.Linares Espinósa, 2018) [39]. Hacen parte de la Revisión Sistemática las siguientes fases a tener en cuenta:

- **Adquisición de la Evidencia**
  El proceso de revisión debe estar bien desarrollado y planificado de antemano para reducir sesgos y eliminar estudios irrelevantes o de baja calidad. Los pasos a seguir para la realización de una revisión sistemática incluyen: (i) formular correctamente la pregunta a responder (PICO), (ii) desarrollo de un protocolo (criterios de inclusión y exclusión), (iii) realizar una búsqueda bibliográfica detallada y amplia, (iv) cribar los resúmenes de los trabajos identificados en la búsqueda y posteriormente de los textos completos seleccionados (PRISMA).

- **Síntesis de la Evidencia**
  Una vez seleccionados los estudios se debe: (v) extraer en un formulario diseñado en el protocolo los datos necesarios para resumir los estudios incluidos, (vi) evaluar los sesgos de cada estudio pudiendo identificar la calidad de la evidencia disponible y, por último, (vii) desarrollar las tablas y el texto que sinteticen la evidencia.

Siguiendo los lineamientos de dicha metodología, se identificaron las fuentes de información más relevantes y las palabras clave que orientaron el proceso de búsqueda de información:

- **Fuentes:** *Dialnet, Google Scholar, IEEE, Redalyc, Scielo.*

- **Palabras Clave:** *MOOC*, *SPOC*, MPOC.

Se realizo una búsqueda general con la palabra clave *MOOC*, visualizando sólo resultados de tipo artículo en los buscadores mencionados, se organiza por año de publicación y los resultados son los mostrados en la tabla 2.1. Para visualizar la tendencia en los números de artículos publicados que incluyan la palabra *MOOC*, se realizó una gráfica con los resultados de la tabla 1 la cual es mostrada en la figura 2.1. Siguiendo la misma metodología para sus variantes *SPOC* y MPOC.

La figura 2.1 permite observar una disminución en los artículos publicados entre los años 2016-2018 que coinciden con un incremento visto en la figura 2.2 en la publicación de artículos sobre *SPOC*s en los mismos años. Esto concuerda con el ciclo de sobre expectación o "hype cycle" (figura 2.4) que realiza la compañía Gartner con el fin de medir el lanzamiento, sobre expectación, madurez, adopción y aplicación comercial de una nueva tecnología, metodología o producto. Donde puede observarse que las investigaciones acerca de MOOCs alcanzan la madurez después del año 2015 [9], para luego dar paso a la implementación de dicho tema y sus variantes en la educación superior, de ahí el incremento en artículos publicados (figura 2.2 y 2.3).

MOOC							
	2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
DIALNET	22	75	114	75	98	25	44
GOOGLE SCHOLAR	4510	9140	14400	16200	14700	13000	15700
IEEE	20	12	7	8	9	12	22
REDALYC	40	102	125	107	106	11	5
SCIELO	50	40	33	22	25	12	9
TOTAL	**4642**	**9369**	**14679**	**16412**	**14938**	**13060**	**15780**

Cuadro 2.1: Resultados búsqueda *MOOC*

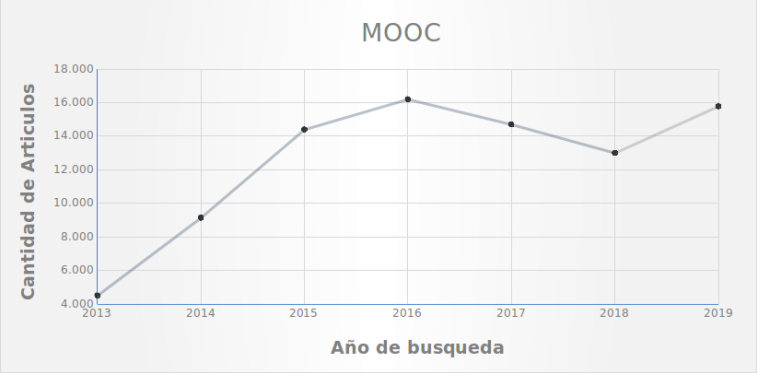


Figura 2.1: Resultados búsqueda tipo artículo con palabra clave *MOOC*

SPOC							
	2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
DIALNET	0	2	2	1	3	3	7
GOOGLE SCHOLAR	734	951	1240	2200	2190	2310	3150
IEEE	3	10	13	18	29	6	7
REDALYC	1	2	4	5	6	1	3
SCIELO	31	39	38	9	10	1	9
TOTAL	**4642**	**9369**	**14679**	**16412**	**14938**	**13060**	**3176**

Cuadro 2.2: Resultados búsqueda *SPOC*

Figura 2.2: Resultados búsqueda tipo artículo con palabra clave *SPOC*

MPOC							
	2013	**2014**	**2015**	**2016**	**2017**	**2018**	**2019**
DIALNET	0	1	1	3	0	0	1
GOOGLE SCHOLAR	362	380	404	392	377	352	447
IEEE	3	2	0	1	1	0	1
REDALYC	2	0	0	2	0	0	0
SCIELO	0	0	0	0	0	0	0
TOTAL	**367**	**383**	**405**	**398**	**378**	**352**	**449**

Cuadro 2.3: Resultados búsqueda MPOC

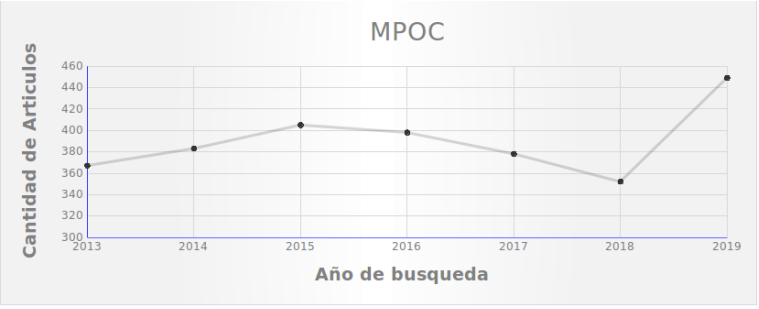


Figura 2.3: Resultados búsqueda tipo artículo con palabra clave MPOC

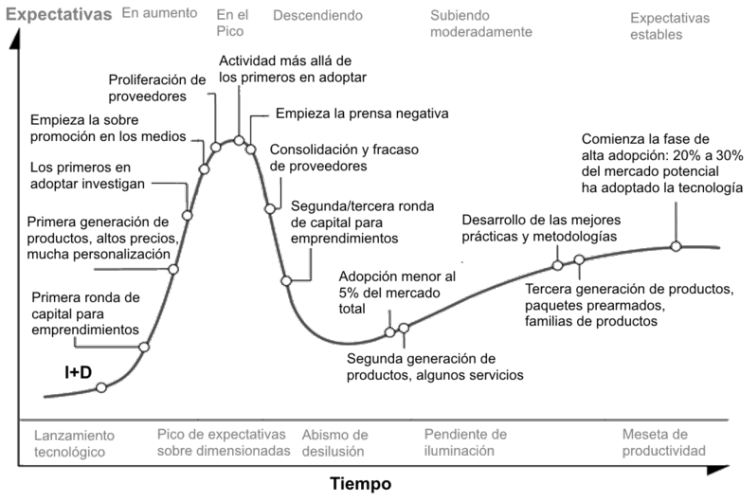


Figura 2.4: Ciclo de Sobre expectacion de *MOOC* segun la empresa Gartner [9].

Después de filtrar los resultados de las búsquedas, se obtuvieron en total 41 trabajos relevantes para la presente propuesta y a partir de estos documentos fueron identificados tres temas principales:

- Fraude en *MOOC* y Estrategias para Disminuir la Deshonestidad Académica en Ambientes Educativos en Línea.
- Analíticas del aprendizaje (Learning Analytics) aplicado en *MOOCs*.
- Estrategias Evaluativas en ambientes *e-learning*

Los trabajos relacionados son descritos a continuación:

2.3. Trabajos Relacionados

2.3.1. Fraude en *MOOC* y estrategias para disminuir la deshonestidad académica en ambientes educativos en línea

Los *MOOC* fueron concebidos con la idea de poder llevar la oportunidad de educación más allá de las aulas de un instituto y poder convertir la web en una gran aula educativa que permitiera acceder a la educación a cualquier persona independientemente de su ubicación geográfica [24]. Este nuevo modelo pedagógico tuvo tal aceptación, que el año 2012 fue considerado por la revista NEW YORK TIMES como el año del *MOOC* [21], este hecho hizo que universidades catedráticas fueran atraídas a este novedoso modelo educativo y decidieran adaptarlo al suyo [18]. Con esto las universidades buscan que por medio de los cursos masivos puedan certificarse créditos académicos en algunos de sus programas [3]. Este cambio hizo que conductas como el fraude aumentaran en los ambientes en línea. A medida que la homologación de los contenidos del *MOOC* por créditos académicos eran más populares, igualmente eran más variados los métodos de cometer fraude [4], en los que destaca la técnica de CAMEO [7] "Copying Answers using Multiple Existences Online" la cual consiste en que un usuario crea múltiples cuentas con el objetivo de extraer la respuestas correctas para luego desde una cuenta denominada "maestra" introducirlas y certificarse sin fallar. Aunque en [7] hacen uso de Learning Analytics para detectar CAMEO, este tipo de técnica de fraude no es implementable en ambientes MPOC. Debido a su naturaleza privada los MPOC realizan un control de ingreso y registro por lo que ningún estudiante puede tener más de dos cuentas.

Sin embargo, por ser un ambiente en línea se facilita la suplantación y el plagio [4]. Aquí hacen aparición técnicas como el uso indebido de recursos web para buscar las respuestas, o que alguien del curso envíe las respuestas correctas a los demás antes de la presentación del examen. Debido al aumento de estas tendencias fraudulentas, han surgido métodos de prevención y corrección. El código de honor es un método traído de la enseñanza tradicional [6] y aplicado en los entornos en línea, pero su efectividad depende del nivel ético de la persona. Por tanto surge una mejora conocida como el

"método de las advertencias" [5], el cual consiste en desplegar advertencias previas a la presentación de los exámenes, para que persuada al estudiante de no cometer fraude, con el fin de no tener represalias en la calificación o la permanencia en el curso.

Para hacer frente a las suplantaciones y plagios han puesto en marcha sistemas de vigilancia que usan las cámaras web de los equipos, con el objeto de identificar y monitorear al estudiante [40],[41]. O el uso de elementos software como los "honeypots" [22] que consisten en un sitio web que contiene las respuestas del examen con el propósito de identificar a las personas que lo consultan en el instante de presentación de la prueba.

Según [42] estos métodos conllevan a una transformación de la filosofía de aprendizaje de los *MOOC* y un posible incremento en el mantenimiento y sostenimiento de la plataforma, haciendo que los cursos sean poco escalables, disminuyendo así la capacidad de masivo y abierto de estos.

Como expone en [4],[42],[43] la supervisión personal, la validación por examen presencial, el uso de terceros como garantes [41], hacen que sus características y la filosofía con la que nace el concepto de *MOOC* [26] se vea alterada dejando de ser abierto, en línea, conectivista o escalable.

En conclusión ninguno de los métodos expuestos anteriormente, son totalmente efectivos y algunos presentan una limitación en el concepto de masivo y abierto, principales caracterisiticas de los *MOOC* [42], por eso la brecha está en encontrar estrategias efectivas en la presentación de las pruebas, que no afecten directamente el concepto de *MOOC* y que permitan la caracterización de los estudiantes fraudulentos y la integración con la academia.

2.3.2. Analíticas del aprendizaje (*Learning Analytics*) aplicado en *MOOCs*.

Learning Analytics o analíticas de aprendizaje es una técnica de análisis de datos en procesos de aprendizaje, estos sirven para dar pautas en el desarrollo de los procesos

mencionados, además de optimizar los entornos en los que ellos se producen. Las analíticas de aprendizaje tienen el potencial de ayudar a estudiantes, profesores y organizaciones a tomar mejores decisiones para conducir a mejores resultados en la formación [44].

En la educación tradicional las analíticas son complejas y pierden valiosos datos sobre los procesos de aprendizaje, pero hoy, las mejores plataformas en linea son capaces de obtener importantes datos sobre lo que ocurre mientras aprende. En los *MOOC*, debido a la gran cantidad de usuarios, les dificulta a los maestros enfocar sus instrucciones de manera personalizada [43]. El *Learning Analytics* en aplicaciones informáticas ha surgido como una solución. En la actualidad, las plataformas *MOOC* ofrecen un bajo soporte para visualizaciones de esta técnica [44], por lo que el desafío es proporcionar aplicaciones de visualización útiles y efectivas sobre el proceso de aprendizaje.

Actualmente, en los cursos en línea de la Universidad del Cauca [45], mediante la herramienta Google Analytics, se han aplicado técnicas para conocer los hábitos de ingreso de los estudiantes y las posibles correlaciones de los resultados de aprendizaje con variables típicas de la educación en línea masiva. Siendo esto de gran aporte, ya que permite a los docentes encargados conocer datos específicos, que son información valiosa a la hora de ajustar el diseño y el desarrollo de dichos cursos para la obtención de mejores resultados educativos.

En los cursos en línea con reconocimiento académico, un factor importante es el poder evitar que los estudiantes hagan fraude en sus exámenes, es aquí donde mediante algoritmos y métodos basados en patrones pueden identificar y detectar comportamientos de fraude como el uso de cuentas múltiples para copiar respuestas, en donde el alumno usa una o más cuentas denominadas “de extracción” para obtener la respuesta correcta, y luego la envía a la cuenta principal llamada “maestra”, esta cuenta es con la que se busca obtener la certificación del curso. Este método de fraude comúnmente utilizado, es conocido como *CAMEO* [7],[8], y la forma de detectarlo para poder combatirlo es mediante dos patrones; el inmediato, en donde cuentan con un margen de 15 minutos para pasar las respuestas de la cuenta de extracción a la cuenta maestra, y el de lotes retrasados, en donde obtienen un grupo de

respuestas en la cuenta de extracción para pasarlas en lote a la cuenta maestra para esto tienen un umbral de 10 respuestas correctas consecutivas. Gracias a él esfuerzo en el desarrollo e investigación de estos métodos de detección cada vez consideran más a los cursos en línea, como metodología de educación, para obtener títulos y certificados importantes.

La implementación de *Learning Analytics* en los *MOOC*, permite además, establecer un factor determinante como lo es la calidad de dichos cursos, considerando criterios que evalúen aspectos pedagógicos, didácticos y técnicos de los mismos [46]. Lo cual da una clara idea de la relevancia que tienen las diferentes técnicas de análisis, cuando son aplicadas al contexto de los cursos en línea.

2.3.3. Estrategias Evaluativas

El proceso de aprendizaje incluye e implica la revisión constante y continúa de lo que se aprende y cómo se aprende, es decir, un proceso de evaluación [47]. Dicho proceso, toma en cuenta registros, observaciones de conducta y trabajo del alumno. Cada uno de estos puntos son acotados mediante diferentes estrategias evaluativas.

En el proceso de evaluación toma importancia la calificación obtenida, por lo que es frecuente que los evaluados recurran durante el examen a métodos no permitidos para aumentar su calificación, incurriendo en conductas que son castigadas o penalizadas en su mayoría con graves consecuencias [47].

Es por esto, que es necesario implementar técnicas o estrategias a la hora de evaluar, que permitan evitar o detectar el fraude. En la educación tradicional han optado por utilizar algunos métodos evaluativos con este propósito, algunos de ellos aplicables tanto a clases presenciales como a cursos en línea [48], dentro de los que mencionan:

- Vigilar a los evaluados. Este método refuerza el acompañamiento presencial poniendo varios vigilantes, o paseando por el lugar del examen (con el fin de evitar “puntos ciegos”). Esto aplicable solo para clases presenciales.
- Elaborar varios modelos de examen, de tal forma que dos estudiantes próximos no tengan nunca las mismas preguntas, o las tengan en distinto orden.

- Comparar exámenes de gente que estuviera sentada en lugares próximos durante el examen, de tal forma poder detectar similitudes excesivas, que hagan sospechar comportamientos fraudulentos. En el caso de exámenes tipo test, puede incluso recurrir a programas estadísticos para este fin.
- Sentar a los evaluados según un orden establecido por el evaluador (que puede ser aleatorio), de tal forma que los evaluados no puedan "trabajar en equipo". Esto solo aplicable para clases presenciales.
- Diseñar el examen relajando o eliminando ciertas normas. Por ejemplo, si el objetivo del examen es determinar la capacidad del evaluado de utilizar sus conocimientos o aplicar procedimientos (como suele ocurrir en exámenes de matemáticas) permitirse el libre uso de apuntes y libros de texto, o navegación en diferentes páginas (para el caso de cursos en línea), aumentando a cambio su dificultad.

La combinación de estos métodos con el uso de distintos modelos de examen es lo que hace realmente difícil que se cometa trampa.

La extensión del uso de Internet, ha proporcionado a los estudiantes la posibilidad de acceder a fuentes de información no disponibles en el pasado. Este hecho ha incrementado el uso inadecuado de la información obtenida, generando más opciones de fraude [49]. Por lo tanto hace necesario también mejorar cada día las técnicas para combatirlo; actualmente en instituciones de educación superior implementan una de las herramientas disponibles comercialmente, dentro de la categoría de sistemas anti copia, llamada *"Turnitin"*, que nos permite conocer si el trabajo de un alumno presenta puntos de coincidencia con textos de autoría distinta a la del propio alumno, además de su localización y acceso a las fuentes sobre las que presentan los indicios de plagio. Lo que representa un gran avance al poder realizar, prácticamente de forma automática, una comprobación de similitud de los trabajos que el docente recibe para su calificación.

En lo que respecta a cursos en línea, de acuerdo a lo definido en el modelo ELQ (*E-Learning Quality*) [50], los métodos de evaluación deben implementar estrategias para lidiar con el plagio, la seguridad y la autenticación de los estudiantes. Para

lograrlo, deben institucionalizar estrategias a través de la creación de normativas, de tal forma que permita orientar al estudiante sobre la manera correcta de consultar y citar sus fuentes de consulta.

2.4. Conclusiones del estado del arte

- En algunos trabajos se evalúan los beneficios que la modalidad en línea aporta a la educación como la masividad y superar las brechas geográficas, pero no se tiene en cuenta que estos también traen problemas como suplantación de identidades, y otros comportamientos fraudulentos.
- La mayoría de los trabajos hacen énfasis en cómo detectar el comportamiento fraudulento por medios electrónicos como el uso de cámaras web, detectores de emociones, señuelos y demás, mas no hay trabajos donde se propongan estrategias evaluativas para identificar el fraude.
- Hay trabajos en los que describen posibles escenarios en los que las detecciones electrónicas que se han planteado pueden ser burladas. No proponen solución para dichos escenarios.
- En los trabajos actuales no tienen en cuenta el componente evaluativo como un elemento que permita la caracterización de comportamientos fraudulentos en *MOOC* y sus variaciones.
- Los pocos trabajos en donde se proponen estrategias evaluativas no son validados, sólo muestran los problemas que el fraude trae en el desarrollo del curso, mas no plantean soluciones a partir de las estrategias evaluativas.
- En la mayoría de los trabajos la tendencia es hacer que el estudiante tome el compromiso de no cometer fraude, mas no tienen en cuenta los escenarios donde los *MOOC* presentan una recompensa académica como reconocimiento de créditos en programas universitarios.
- Son pocos los trabajos con reconocimiento académico que detecten y caractericen comportamientos con sospecha de fraude utilizando métodos que no

atenten contra la filosofía de los *MOOC*.

- En los trabajos actuales se evidencia la falta de soluciones que permitan caracterizar el comportamiento fraudulento en ambientes *MOOC* por medio de estrategias evaluativas.

2.5. Brechas

Se puede ver que son muchas las brechas encontradas a partir del procedimiento de revisión sistemática que se realizó, de donde se resaltan tres brechas principales a las cuales se pretende contribuir con este trabajo:

- Los *MOOC* no estudian los métodos de evaluación como factor que facilite o disminuya los comportamientos fraudulentos.
- Son pocos los trabajos con reconocimiento académico que detecten y caractericen comportamientos con sospecha de fraude en cursos con reconocimiento académico.
- La falta de soluciones que permitan caracterizar el comportamiento fraudulento en ambientes *MOOC* por medio de estrategias evaluativas.

Capítulo 3

Mecanismo para la detección de estudiantes con sospecha de fraude a través de estrategias evaluativas.

3.1. Recolección de datos

En esta sección se presentan las características principales de la plataforma de aprendizaje con la cual se trabajó (Selene Unicauca) y La manera cómo se extrajo los datos desde la misma.

3.1.1. Plataforma de aprendizaje

La plataforma de aprendizaje utilizada es una instancia de Open edX la cual en la Universidad del Cauca lleva el nombre de Selene. La plataforma fue implementada en el año 2016 como parte del trabajo de grado de un estudiante de maestría (Jaramillo-Morillo, 2017). La plataforma hasta el momento se encuentra en uso y en ella se imparten cursos tipo MPOC que son reconocidos académicamente. La mayoría de estos cursos son electivos y transversales para todas las carreras de la universidad,

algunos de estos son ofertados en modalidad cursos FISH (Componente de Formación Integral Social y Humana).

3.1.2. Características básicas de Open EdX

Open edX es una plataforma de código abierto. Esta basada en un marco de programación que utiliza lenguajes como *python, javascript, php* y *framerworks* como *Django*, que está disponible gratuitamente para la comunidad en general. Cualquier institución puede descargarla y ejecutar su propia instancia, permitiendo a los educadores gestionar los recursos o contenidos de aprendizaje que satisfagan sus necesidades y a los desarrolladores aportar nuevas funcionalidades a la plataforma.

Open edX cuenta con los siguientes componentes:

- LMS.
- CMS.
- Módulo XBlock.
- ORA2 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 vídeos, textos y problemas organizados bajo una estructura jerarquica que los presenta en secciones, subsecciones y unidades. En la Figura 3.1 se muestra un ejemplo de cómo los contenidos son presentados a los estudiantes.

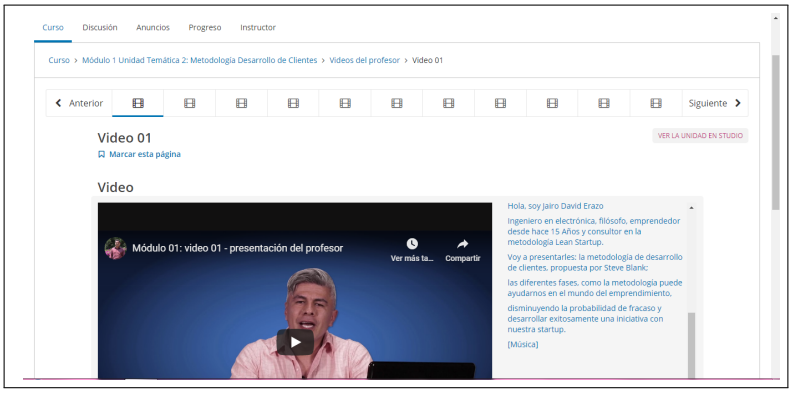


Figura 3.1: Captura de pantalla del LMS de Open edX (Contenidos)

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 3.2 se muestra la interfaz que es visualizada por el docente.

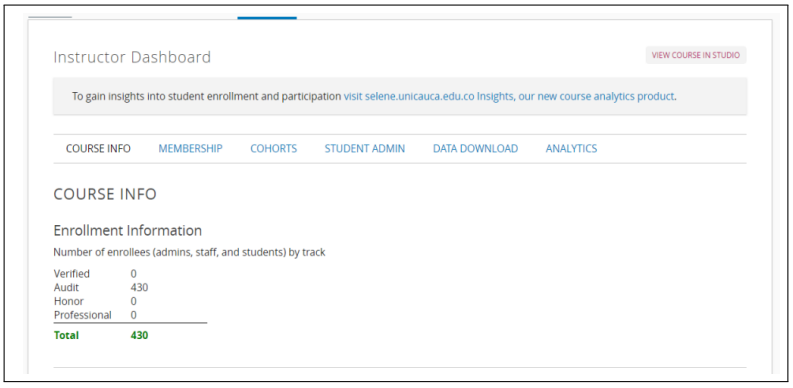


Figura 3.2: Captura de pantalla, interfaz para el instructor LMS de Open edX

Aquí se puede encontrar la información de los inscritos en el curso, inscribir o dar de baja a estudiantes, generar grupos de trabajo para mostrar contenidos, administrar los foros, generar y descargar reporte sobre las evaluaciones del curso entre otras funciones. 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 conocido como *Studio* es la herramienta que se utiliza para crear cursos. En *Studio* se puede gestionar la estructura del curso y agregar el contenido del mismo, incluidos los diferentes tipos de evaluaciones que van desde cuestionarios de opción múltiple hasta problemas matemáticos con uso de ecuaciones aritméticas, vídeos, lecturas, *Blogs* y otros recursos para los alumnos. En la Figura 3.3 se muestra parte de la interfaz en la cual los docentes pueden crear sus cursos.

Para que un usuario pueda crear cursos es necesario que la cuenta cuente con permisos especiales, esto solo los puede otorgar el administrador de la instancia Open edX instalada.

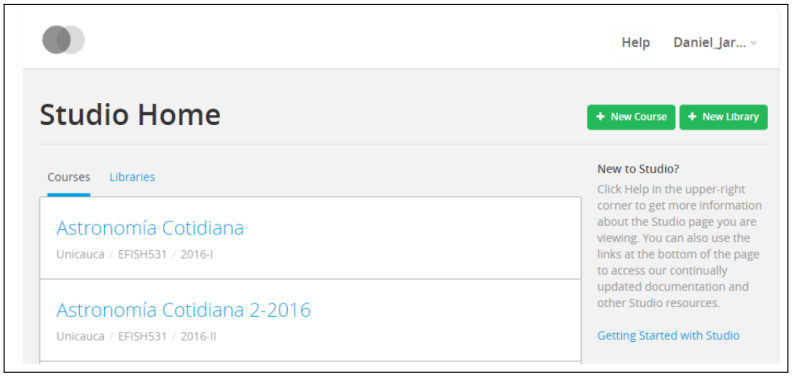


Figura 3.3: Captura de pantalla Studio edX

3.1.2.3. Contenidos

Existen cuatro tipos de componentes que se pueden crear como contenido en una unidad, estos son: foros, ejercicios, vídeos y componentes Html. En la Figura 3.4 se muestra la interfaz para adicionar componentes a una unidad..

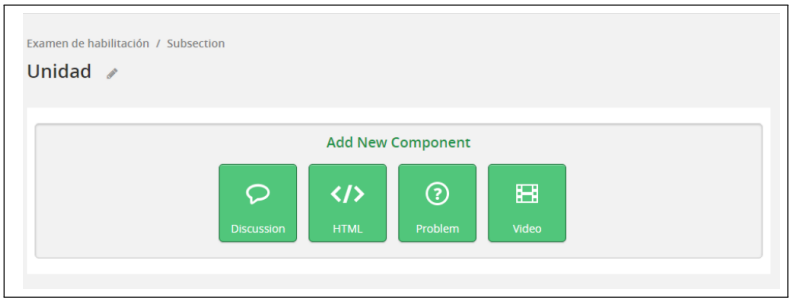


Figura 3.4: Componentes en una unidad

- **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 discusión en el LMS.

- **Componentes HTML**

Este tipo de contenido permite agregar gran variedad de contenido multimedia que se pueda soportar en HTML como lo son textos, listados, enlaces a otros sitios web, presentaciones, archivos en formato pdf, etc. brindando así versatilidad al contenido de un curso.

- **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, por defecto, los ejercicios no cuentan en la nota del estudiante a menos que sea configurado en la subsección que contiene el ejercicio. Los ejercicios más comunes que están disponibles en *Studio* son:

- Casillas de verificación: Son ejercicios cuya respuesta debe ser seleccionada y marcada con un clic en una casilla.
- Lista desplegable: Son ejercicios en los que se muestra una lista de opciones y el estudiante debe validar la respuesta correcta.
- Opción múltiple: Son ejercicios donde se muestra por cada pregunta múltiples opciones de respuesta las cuales pueden ser una o mas correctas.
- Entrada numérica.: Son ejercicios donde existe una caja de texto en la cual el estudiante ingresa una respuesta de tipo numérico la cual debe ser exactamente igual a la definida por el profesor.
- Entrada de texto corta: Similar a la entrada numérica, la entrada de texto corta son ejercicios donde existe una caja de texto donde el estudiante ingresa una respuesta de tipo texto la cual debe ser exactamente igual a la definida por el docente.

- **Componentes de Vídeo**

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 visualizar un vídeo en el curso, se buscará primero el mismo en YouTube. Si no estuviera disponible o el vídeo no comenzar a visualizarse, 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.

Selene Unicauca es una plataforma de ofrecimiento de cursos en línea masivos y abiertos (*MOOC*) de la Universidad del Cauca. Esta iniciativa nace desde el Departamento de Telemática y ha sido soportada por los proyectos *MOOC-Maker*, *MOOC*-MenTES y *MOOC*-Unicauca. Surge gracias a trabajos de grado de estudiantes del Doctorado y Maestría en Ingeniería Telemática. Estrategia digital de innovaciones educativas basadas en *MOOC*.

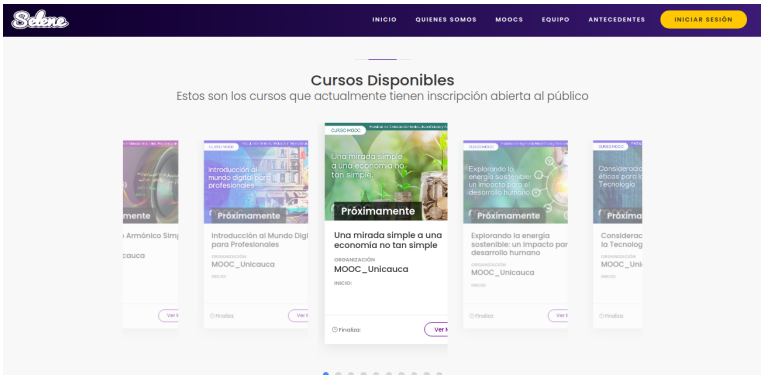


Figura 3.5: Página de bienvenida de la Plataforma Selene

Los 11 cursos que se presentan en la pestaña cursos son desarrollados por *MOOC*-Unicauca, actualmente se están completando para que próximamente se liberen.

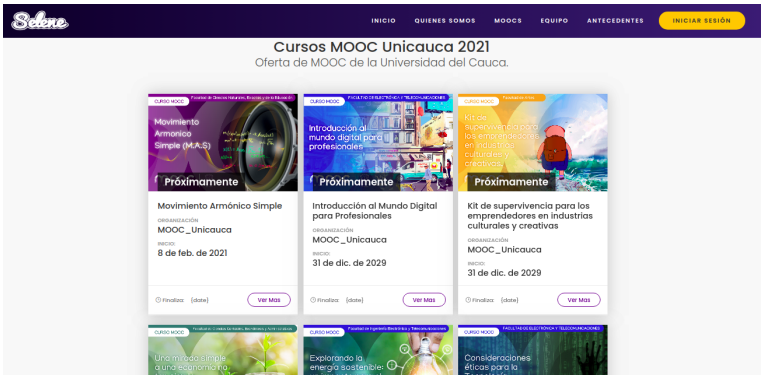


Figura 3.6: Cursos *MOOC* ofrecidos por la Plataforma Selene

El estudiante cuenta con acceso dentro de la plataforma a cada uno de los cursos en los que esta inscrito.

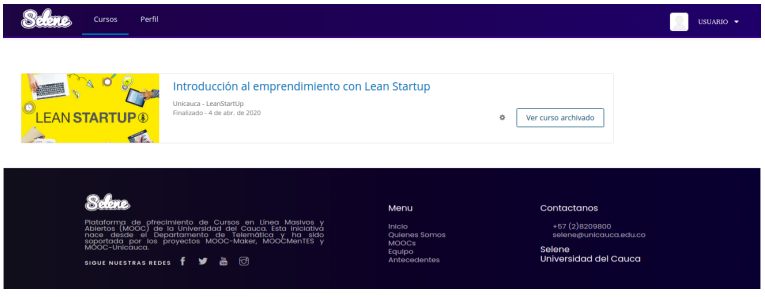


Figura 3.7: Presentación del curso para el Usuario desde la Plataforma Selene

3.1.3. Extracción de datos desde la plataforma Selene

Todas las interacciones de los estudiantes se registran en un archivo con formato JSON llamado *tracking.log*, el cual se encuentra dentro del servidor que aloja la instancia de Open edX. De aquí, la extracción de datos comienza con la consulta realizada por una API de sus siglas en ingles *Application Programming Interface* la cual se encarga de realizar una conexión al servidor de Selene para ingresar al componente *tracking.log* y recolectar los datos, este componente contiene todos los eventos realizados en los cursos de la plataforma Selene, comprenden desde el inicio de sesión, visualización de contenidos multimedia, matriculas de un curso hasta las direcciones IP publicas desde donde se realizo cada uno.

Una vez establecida la conexión el recolector de datos, por medio de una petición GET obtiene el contenido del archivo *tracking.log*, como el contenido de este archivo esta en formato texto lineal, el recolector de datos tiene un componente que se encarga de convertir ese texto a formato *JSON* el cual es el formato que utiliza la base de datos noSQL llamada cursosmooc donde se alojara la información.

Esta base de datos esta implementada en el sistema de bases de datos no relacionales documentales llamada *MONGODB*, al estar implementada en una base de datos no relacional y de tipo documental la persistencia y almacenamiento de los datos debe estar definida por un esquema de datos representados por etiquetas el cual se muestra a continuación:

- username: Campo de tipo *String* comprende caracteres alfanuméricos e identifica el usuario único de un estudiante registrado en Selene.
- dir_ip: Campo de tipo *String* comprende caracteres alfanuméricos que representan la dirección de ip publica unica, en la que el estudiante realiza distintos eventos en Selene.
- answers: Campo de tipo *String* comprende caracteres alfanuméricos que muestran las respuestas a los cuestionarios planteados en los cursos de Selene presenta la siguiente estructura **input_eee7918aec_2_1=choice_1** donde la primera cadena de caracteres corresponde al identificador único de la pregunta

asignado por la plataforma Selene y que lo relaciona a una actividad única en la plataforma, los caracteres choice_1 indican la respuesta que el usuario eligió siendo choice_0 análogo a la respuesta A o a la respuesta de primer orden y choice_1 a la respuesta B o segunda respuesta y así consecutivamente.

- course: Campo de tipo *String* comprende caracteres alfanuméricos que identifica el mpoc al cual pertenece el evento asociado al usuario.
- session: Campo de tipo *String* comprende caracteres alfanuméricos que identifican la sesión del usuario, este es único e irrepetible.
- date: Campo de tipo *String* comprende caracteres alfanuméricos que representan la fecha en la cual sucedió el evento o interacción.
- unit: Campo de tipo *String* comprende caracteres alfanuméricos que identifican la unidad en la cual esta almacenado el examen o evento que se ha realizado.
- name: Campo de tipo *String* comprende caracteres alfanuméricos que representan el tipo de evento realizado en el mpoc, para este caso de estudio nos interesa unicamente el name = "problem_check"
- section: Campo de tipo *String* comprende caracteres alfanuméricos que identifica la sección de una Unidad a la cual el evento realizado pertenece.
- subsection: Campo de tipo *String* comprende caracteres alfanuméricos que representa la subsección de una sección de una Unidad a la cual el evento realizado pertenece.
- time: Campo de tipo *String* comprende caracteres alfanuméricos que muestra la hora en Colombia a la cual el evento en el mpoc fue realizada.
- page: Campo de tipo *String* comprende caracteres alfanuméricos que almacena la *URL* del ingles uniform resource location, de la pagina donde se realizo el evento en el mpoc.

La arquitectura de extracción de datos se presenta a continuación:

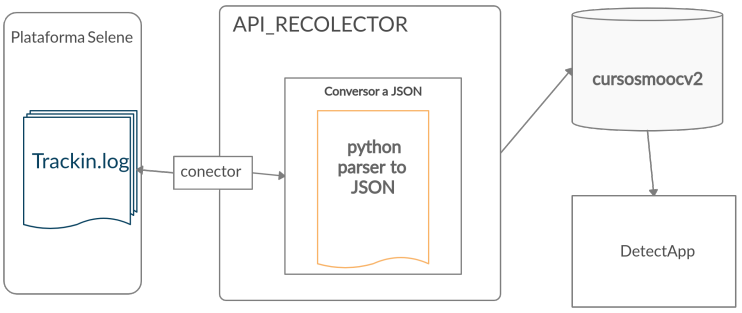


Figura 3.8: Arquitectura De Extracción de Datos

- ***Plataforma Selene:*** La plataforma Selene es una instancia de Open edX que permite la oferta de MPOC para la Universidad del Cauca y se instala en un servidor físico de la universidad, la plataforma comenzó a funcionar en el primer período de 2016.

- ***Tracking.log:*** Cada actividad que los estudiantes realizan a través de la plataforma se registra en un archivo llamado tracking.log. De este archivo se obtienen tanto las respuestas de los estudiantes como las direcciones IP para enviar los exámenes. Además de obtener otra información que ayuda a entender el comportamiento de los estudiantes como: contenido visto, interacciones de video, publicaciones en foros, etc. Es decir todo lo relacionado con la interacción de los estudiantes con la plataforma de aprendizaje.

- ***API_RECOLECTOR:*** Este modulo esta compuestos por dos componentes: el primero es el **Conector** que esta encargado de establecer una conexión con el servidor de la plataforma Selene. El segundo es el **Conversor a JSON** el cual es responsable de buscar el archivo Tracking.log, toma todos los registros que se generan a partir de las interacciones de los estudiantes con la plataforma de aprendizaje, extrae la información y la procesa en un archivo *JSON* el cual es entregado a la aplicación DetectApp que se encarga de su procesamiento y visualización.

- ***DetectApp:*** Es la aplicación desarrollada para filtrar, analizar y visualizar los resultados obtenidos con la implementación de la estrategia evaluativa di-

señada en este trabajo, con el fin identificar a los estudiantes con sospecha de fraude.

3.2. Detección de estudiantes con sospecha de fraude

En esta sección se describe el método utilizado para la identificación de estudiantes con sospecha de fraude.

3.2.1. Criterios para la identificación de estudiantes

Para la identificación de sospecha de fraude nos basamos en el trabajo previo[51], en donde identifican colaboraciones o asociaciones de estudiantes para trabajar en conjunto en las evaluaciones, utilizando el tiempo de envío de las respuestas de los estudiantes para su detección. Luego clasifican a los estudiantes como sospechosos de fraude dependiendo de un análisis de su interacción con la plataforma de aprendizaje. Por tanto, con base en este trabajo previo, diseñamos una estrategia evaluativa cuyos resultados nos permita aplicar dos indicadores más que el trabajo anterior para definir si un estudiante es o no sospechoso de fraude, estos parámetros están descritos a continuación:

- Número de coincidencias.
- Dirección IP desde donde se envío las respuestas.
- Tiempo de envío de respuesta.

Número de coincidencias.

Hace referencia a la cantidad de respuestas exactas que dos estudiantes tienen cuando realizan un examen en línea. Al haber una gran cantidad de estudiantes y por ende una gran cantidad de respuestas a comparar, se hizo uso de una estructura de datos

tipo matricial donde filas y columnas serán el porcentaje de coincidencias entre las respuestas de cada alumno en el examen de cada unidad. Este método está basado en [52];

Para la creación se utilizaron los datos organizados por el componente recolector de datos, este componente genera un archivo que es persistido en la base de datos noSql, dicho archivo contiene entre otra información las respuestas enviadas por los estudiantes. La aplicación web DetectApp hace una clasificación de las respuestas basada en el examen y a la unidad que corresponda. Por cada examen en una unidad, la matriz se construye con el cálculo de coincidencias comparando cada una de las respuestas de un estudiante con si mismo y con el resto de alumnos. Obteniendo así la matriz DS.

$$DS = \begin{pmatrix} ds_{1,1} & ds_{1,2} & ds_{1,3} & \cdots & ds_{1,M} \\ ds_{2,1} & ds_{2,2} & ds_{2,3} & \cdots & ds_{2,M} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ ds_{N,1} & ds_{N,2} & ds_{N,3} & \cdots & ds_{N,M} \end{pmatrix} \tag{3.1}$$

Dónde cada entrada ds sub i,j es el porcentaje de coincidencia de respuestas. Cada elemento de la matriz está calculado por una coincidencia exacta. Por ejemplo, si un estudiante coincide con otro en todas las respuestas tendrá un valor de 100, caso contrario un valor de 0. Basados en [52] los estudiantes con sospecha de fraude son aquellos que tienen más del 90 % de similitud en sus respuestas.

Parte de la matriz resultante para el caso de estudio es la siguiente:

$$\begin{pmatrix} & Est,1 & Est,2 & Est,3 & Est,4 & Est,5 & Est,6 & Est,7 \\ Estudiante1 & 100 & 30 & 40 & 40 & 50 & 50 & 10 \\ Estudiante2 & 30 & 100 & 20 & 10 & 10 & 10 & 20 \\ Estudiante3 & 40 & 20 & 100 & 60 & 70 & 90 & 100 \\ Estudiante4 & 40 & 10 & 60 & 100 & 20 & 40 & 20 \\ Estudiante5 & 50 & 10 & 70 & 20 & 100 & 20 & 70 \\ Estudiante6 & 50 & 10 & 90 & 40 & 20 & 100 & 20 \\ Estudiante7 & 10 & 20 & 100 & 20 & 70 & 20 & 100 \end{pmatrix} \tag{3.2}$$

Dirección IP desde donde se envió las respuestas.

Según [8], [21], [22] y [41] la creación de grupos de trabajo para responder un examen es un método de fraude muy común en entornos en linea como los *MOOC* o sus variantes, para la detección de estos se diseño el criterio de la dirección IP. DetectApp hace una consulta por cada examen realizado en una unidad, trae todas las respuestas de los alumnos que presentaron dicho examen junto con la hora de envío y la dirección IP desde donde se enviaron. Con esa información agrupa a los estudiantes que presenten direcciones IP idénticas y los clasifica en grupos de trabajo, estos grupos de trabajo son mostrados al docente en una vista donde se despliegan enlaces con las direcciones IP coincidentes, al dar clic en ella se muestra la lista de estudiantes que presentaron el examen en la misma dirección IP junto con sus respuestas y hora de envío, permitiendo así tener los tres parámetros de criterio para la identificación de estudiantes con sospecha de fraude.

Tiempo de envío de respuesta.

Este parámetro ayuda a aumentar el nivel de sospecha de fraude en situaciones donde se presente que dos o mas alumnos tengan la misma cantidad de respuestas coincidentes pero no la misma dirección IP, así pues se puede presumir que han sido compartidas las respuestas entre alumnos que presenten un porcentaje entre el 90 y 100 por ciento de coincidencia y la diferencia de tiempo de envíos sea mínima.

3.3. Herramienta de Apoyo a Docentes

El diseño del mecanismo para la detección de estudiantes con sospecha de fraude aplicable en cursos en linea masivos, se realizó teniendo en cuenta la estructura y funcionamiento de la plataforma Open edX y la evaluación del cumplimiento de los parámetros de la estrategia evaluativa definida en la sección 4.2. En la Figura 3.9 se presenta la arquitectura general del mecanismo construido.

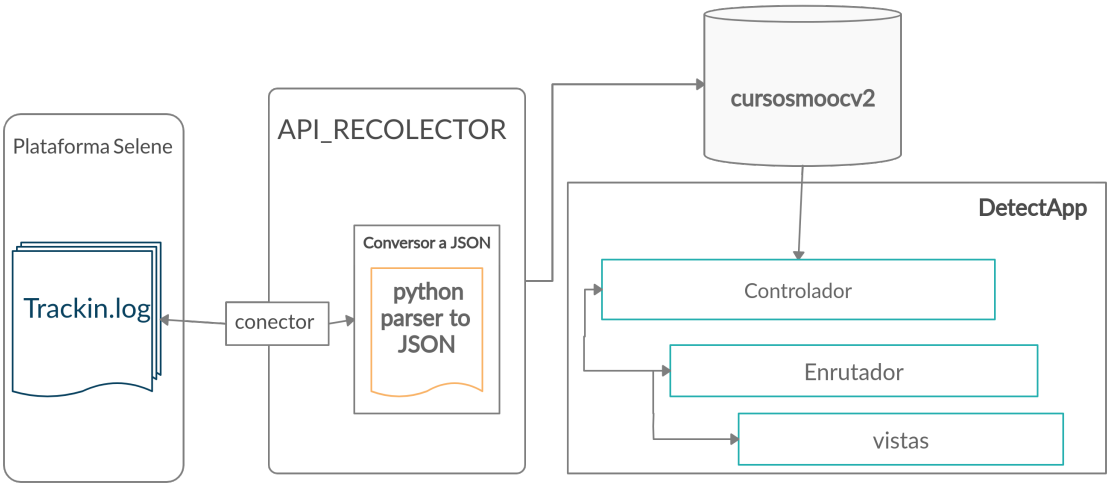


Figura 3.9: Arquitectura General del Mecanismo

En esta arquitectura, el archivo tracking.log es donde se registran las interacciones de los estudiantes con la plataforma de aprendizaje Selene. Su ubicación dentro del servidor es: /edx/var/log/tracking/. API_RECOLECTOR es el componente recolector de datos, este está encargado de obtener los datos del archivo tracking.log, leer su contenido, extraer la información relevante para la identificación de comportamientos de fraude y guardarla en la base de datos cursosmoocV2.

Por su parte, la aplicación web llamada DetectApp, permite realizar consultas a través del componente vistas, éste genera una petición web la cual es identificada por el componente Enrutador.

Desarrollado en el *framework NodeJS*, el Enrutador busca en la lista de enrutamiento el proceso que esta relacionado con la petición web, una vez encontrado, éste consulta y clasifica la información relevante por medio de su componente Controlador, el cual contiene la lógica de negocio. Los dos principales procesos a destacar son, el contador de ocurrencias iguales, encargado de buscar y contar cuantas respuestas coinciden entre estudiantes, y el identificador de IPs coincidentes, también diseñado en el mismo *framework*. Dicha información es entregada al Enrutador que define en que vista se va a mostrar en el navegador web.

Estas vistas están desarrolladas en un motor de plantilla para *NodeJS* llamado *EJS*,

el cual permite mostrar a los usuarios en este caso a los docentes del curso, la información relevante según los criterios definidos para la detección de estudiantes con sospecha de fraude.

3.3.1. Descripción de la Arquitectura

Aunque existen diferentes aproximaciones para describir arquitecturas de software, se optó por utilizar el modelo de 4+1 vistas, ya que permite representar de forma estándar la arquitectura a través de diagramas UML. El modelo "4+1" de Kruchten, es un modelo de vistas diseñado por el profesor Philippe Kruchten y encaja con el estándar "IEEE 1471-2000" (*Recommended Practice for Architecture Description of Software-Intensive Systems*) que se utiliza para describir la arquitectura de un sistema software.

3.3.1.1. Vista de escenarios

La descripción de la arquitectura en esta vista se hace mediante diagramas de casos de uso, a partir de aquí se comienza a unir y relacionar las otras 4 vistas. Esta vista es obligatoria cuando se utiliza el modelo 4+1 vistas, ya que todos los elementos de la arquitectura se derivan de los requerimientos que aquí se presentan. A continuación, se listan los casos de uso del sistema:

- Iniciar Sesión.
- Escoger Curso.
- Escoger Examen.
- Consultar Coincidencias en Respuestas.
- Consultar Direcciones IP.
- Seleccionar Informe por IP.

En la Figura 3.10, se presenta el Diagrama de Casos de Uso que orienta el diseño arquitectónico del presente trabajo.

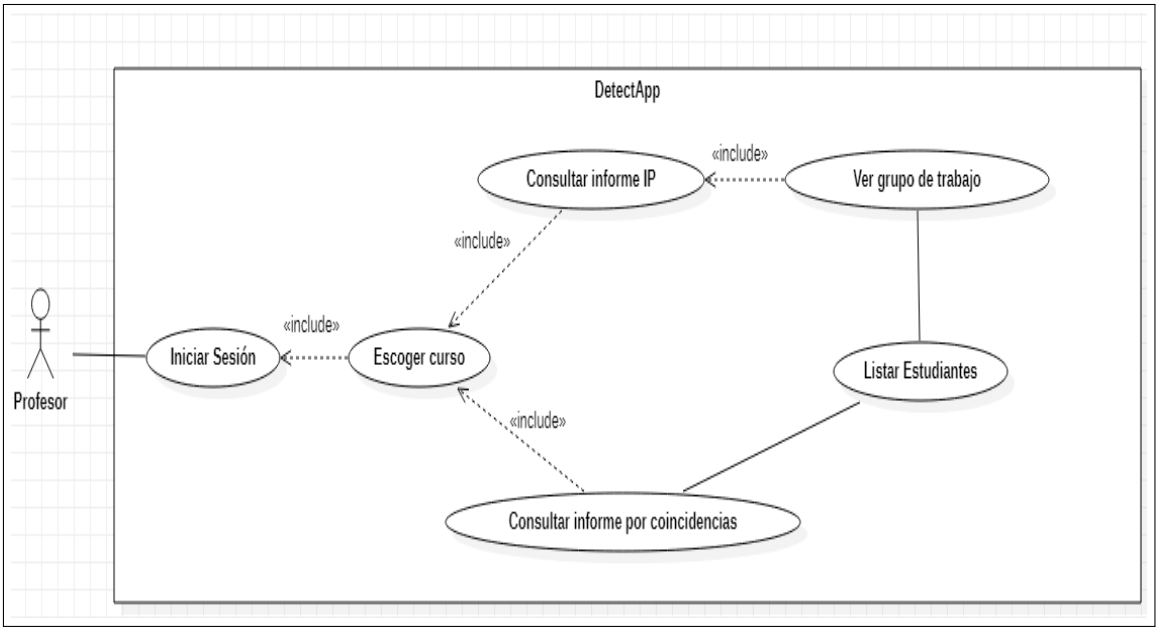


Figura 3.10: Diagrama de Casos de Uso Implementación Mecanismo de Identificación

La herramienta debe permitir a los docentes realizar consultas sobre los indicadores de sospecha de fraude de los estudiantes en los exámenes de la plataforma de aprendizaje Selene. Los principales datos se relacionan con las respuestas de los exámenes presentados por los estudiantes del curso, hora de presentación, dirección IP, e identificación de cada uno de los cuestionarios. Para las consultas los usuarios registrados deben poder escoger el curso, escoger el examen, seleccionar la consulta que desea realizar, y de esta forma realizar un seguimiento al comportamiento de los estudiantes, es decir, ver cuáles son los datos específicos de las respuestas a cada examen realizado por los estudiantes en la plataforma de aprendizaje.

A continuación se describen los casos de uso del sistema:

- **Iniciar Sesión**
  Actores: Profesor.

Escenario de Interacción: Interfaz de inicio de sesión.
Requisitos: Tener credenciales registradas en la aplicación.
Flujo de eventos: Paso 1. El usuario ingresa los datos de sus credenciales, Paso 2. El sistema valida las credenciales y da acceso.
Pos condición: Muestra la página con los cursos disponibles.
Excepciones: Dirección de correo no registrada, Contraseña o Correo incorrectos.

- **Escoger curso**
  Actores: Profesor.
  Escenario de Interacción: Interfaz cursos.
  Requisitos: Previa validación de credenciales.
  Flujo de eventos: Paso 1. El sistema despliega la lista de cursos disponibles, Paso 2. El usuario selecciona el curso al cual desea analizar.
  Pos condición: Muestra la página EXÁMENES con los exámenes disponibles.
  Excepciones: Ninguna.

- **Consultar informe por IP**
  Actores: Profesor.
  Escenario de Interacción: Interfaz Exámenes del curso.
  Requisitos: El usuario debe haber elegido el examen al cual quiere hacer la consulta.
  Flujo de eventos: Previa elección del examen a analizar realizada por el usuario. El sistema mostrará los grupos de trabajo que corresponden a estudiantes que tengan la misma dirección IP con sus respectivos datos.
  Pos condición: Muestra el listado de los grupos de trabajo.
  Excepciones: Ninguna.

- **Ver grupo de trabajo**
  Actores: Profesor.
  Escenario de Interacción: Interfaz listado de direcciones IP o grupos de trabajo.
  Requisitos: El usuario debe haber elegido el curso al cual quiere hacer la consulta y la dirección IP del grupo de trabajo.
  Flujo de eventos: Previa selección del examen a analizar realizada por el

usuario, y escoger la dirección IP del grupo de trabajo el cual se quiere analizar, se selecciona la opción ver grupo de trabajo. El sistema mostrara una interfaz con las opciones de listado según los parámetros de la estrategia evaluativa.
Pos condición: Despliega una interfaz con los distintos listados relacionados con la cantidad de estudiantes detectados.
Excepciones: Ninguna.

- **Consultar Informe por Coincidencias**
  Actores: Profesor.
  Escenario de Interacción: Interfaz Exámenes del curso.
  Requisitos: El usuario debe haber elegido el examen al cual quiere hacer la consulta.
  Flujo de eventos: Previa selección del examen a analizar realizada por el usuario, en el cual se desea ver a los estudiantes con un 90 % a un 100 % de coincidencias en respuestas en el mismo orden, el sistema mostrará los estudiantes que cumplan con ese requisito junto a su porcentaje de respuestas coincidentes.
  Pos condición: Despliega una interfaz con los distintos listados relacionados con la cantidad de estudiantes detectados.
  Excepciones: Puede no existir coincidencias del 90 % al 100 % entre respuestas en el mismo orden, por lo que el sistema mostrará un mensaje.

- **Listar estudiantes**
  Actores: Profesor.
  Escenario de Interacción: Interfaz grupo de trabajo o Interfaz informes por coincidencias.
  Requisitos: El usuario debe haber elegido el curso al cual quiere hacer la consulta y seleccionar el tipo de informe.
  Flujo de eventos: Previa selección del examen a analizar realizada por el usuario, el sistema despliega una interfaz con opciones que dependen de si es una consulta por informe de coincidencias o por informe IP, el usuario selecciona una de estas opciones y el sistema lista los estudiantes de acuerdo a la opción seleccionada.
  Pos condición: Despliega una interfaz con los distintos listados relacionados

con la cantidad de estudiantes detectados.
Excepciones: Ninguna.

3.3.1.2. Vista Lógica

En esta vista se representa la funcionalidad que el sistema proporcionará a los usuarios finales. La representación se realiza mediante diagramas de clases. El diagrama que se presenta a continuación es un diagrama general para el mecanismo de identificación, de manera que permita entender cómo funciona el mecanismo de una manera más simple.

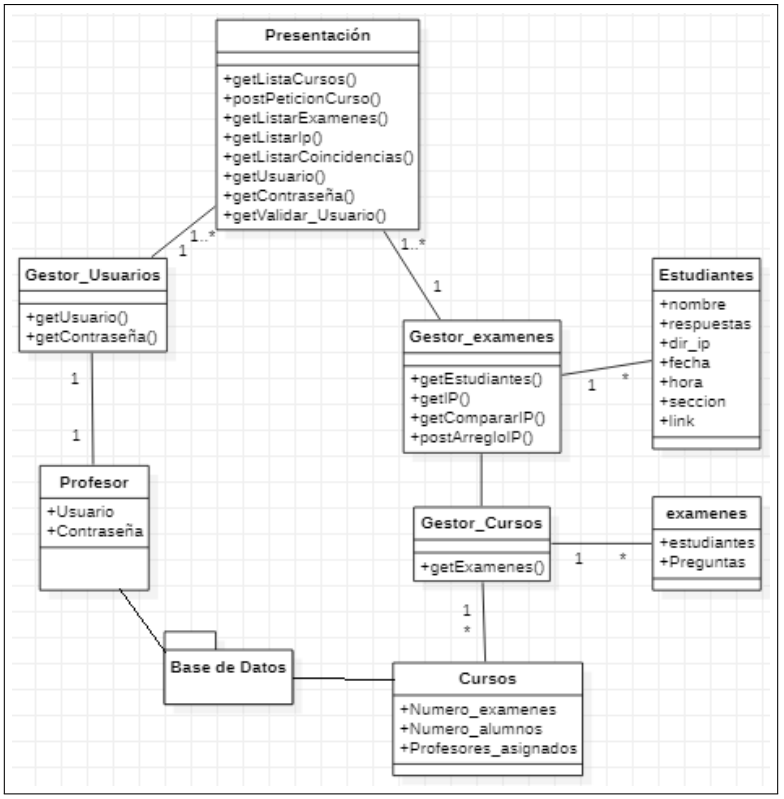


Figura 3.11: Diagrama de Clases para el Mecanismo de Identificación

3.3.1.3. 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. Se describen por medio de un diagrama de actividades directamente relacionado con los casos de uso definidos anteriormente.

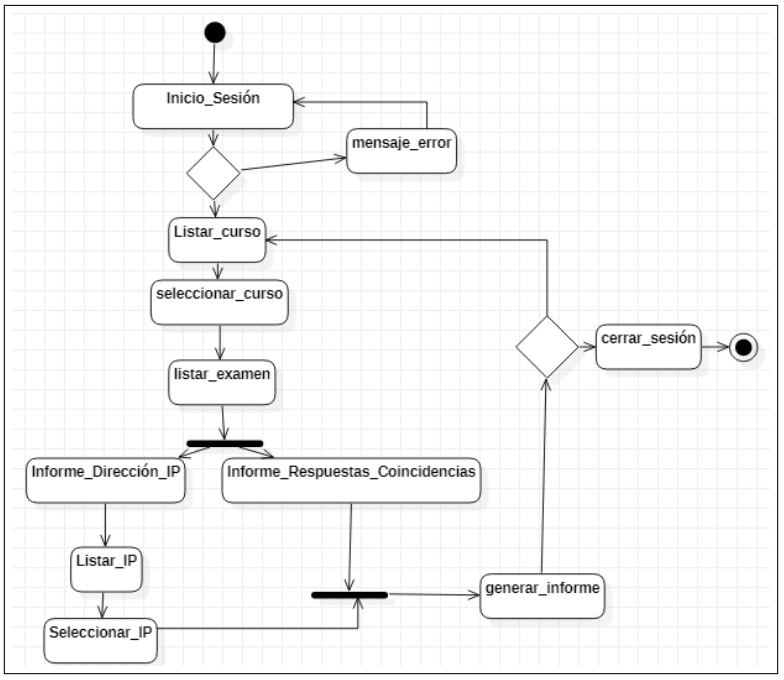


Figura 3.12: Diagrama de Actividades del Sistema

En la Figura 3.12, se presenta el diagrama de actividades asociado a la visualización de la información previamente tratada en la lógica de programación, ésta obtenida de los registros que se generan en la plataforma de aprendizaje Selene (OpenedX). La plataforma Selene actualiza constantemente un archivo con extensión JSON en el cual se registran las interacciones del estudiante con la plataforma de aprendizaje. De aquí, mediante el adecuado manejo del archivo es posible filtrar la información con el fin de obtener solo los datos relacionados con los exámenes.

Una vez filtrados los datos, es necesario tratarlos bajo los criterios de direcciones IP idénticas con esta información útil se despliega a el usuario Profesor. Para esto, se propone la construcción de una aplicación Web que permite al docente ver los

resultados del análisis de la información en una manera usable y útil.

3.3.1.4. 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.

El mecanismo se encuentra dividido en dos módulos importantes, el módulo de recolección y trata de datos y el módulo de visualización (Aplicación web). En las Figuras 3.13 y 3.14 se muestran los diagramas de componentes para estos módulos.

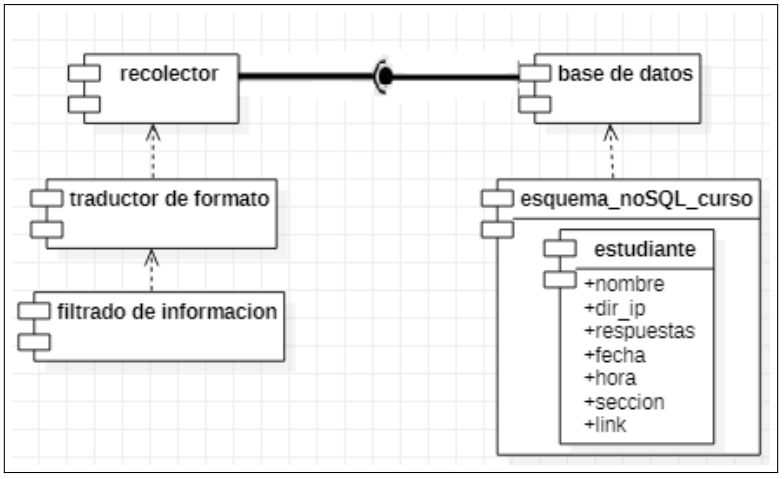


Figura 3.13: Diagrama de componentes Modulo de Recolección de Datos

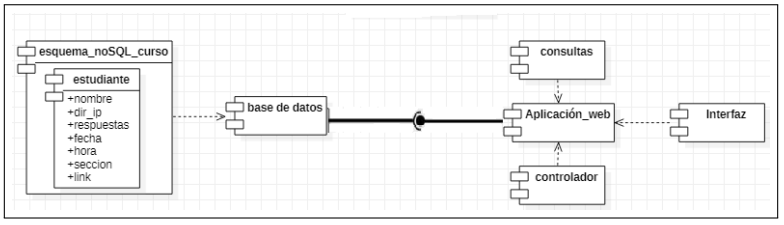


Figura 3.14: Diagrama de componentes Modulo de Visualización

3.3.1.5. Vista de Despliegue

En esta vista se muestran todos los componentes físicos del sistema, así como las conexiones que conforman la solución (incluyendo los servicios). En la Figura 3.15, se muestra el diagrama de despliegue del mecanismo para el seguimiento.

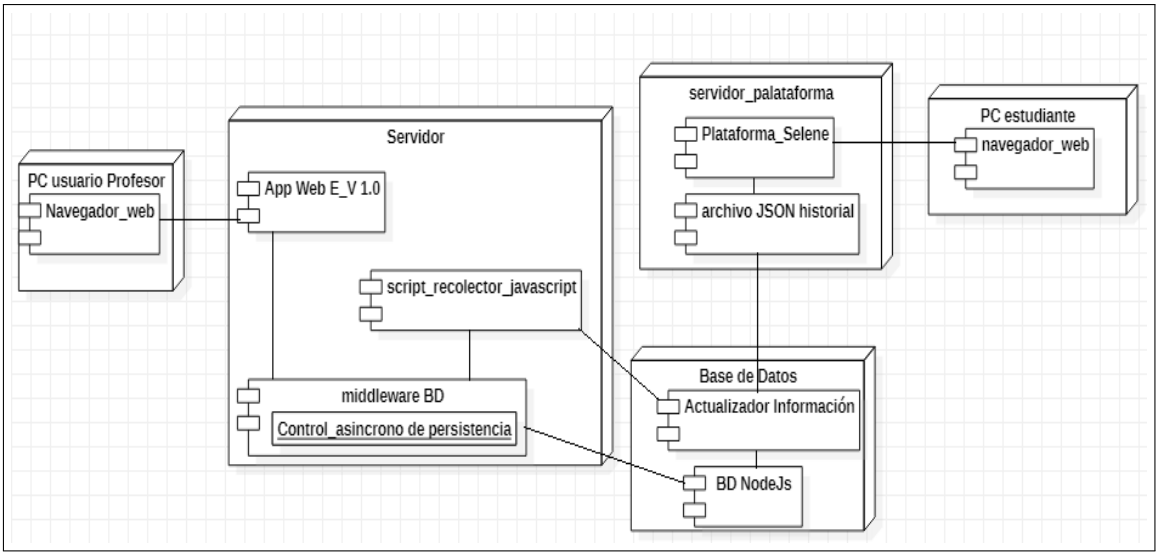


Figura 3.15: Diagrama de Despliegue del Sistema

Navegador Web Estudiante

El estudiante desde su navegador web podrá interactuar con la plataforma Selene mediante una conexión web, dicha conexión web es conducida a través de una dirección IP publica la cual es única y la misma para todos los dispositivos que estén conectados a ese nodo de red.

Plataforma Selene

La plataforma Selene gestiona todas las interacciones que los estudiantes, docentes y administrativos generan en ella, estas interacciones son guardadas en un archivo de ocurrencias conocido como *Tracking.log*. dicho archivo es el que el recolector de datos API_RECOLECTOR consulta y procesa.

Recolector de Datos

El recolector de datos genera una conexión al servidor de alojamiento de la Plataforma Selene por medio de un socket gestor de conexiones basado en un *framework* de *python* llamado *Django*, este abre un túnel el cual permite que el API_RECOLECTOR busque en la ruta predefinida (/edx/var/log/tracking/) el archivo tracking.log y lo copie para cambiarlo de formato de texto lineal a *JSON*. Una vez dado formato, el API_RECOLECTOR lleva este nuevo archivo y lo almacena en la base de datos cursosmoocV2.

Base de Datos

La base de datos es una entidad no relacional llamada cursosmoocV2, esta implementada en el sistema de base de datos *MONGODB*, este sistema solo admite documentos en formato *JSON*, de ahí que el recolector de datos le de formato anteriormente.

DetectApp

La aplicación DetectApp es un conjunto de componentes cuyo fin es organizar la información almacenada en la base de datos cursosmoocV2 y mostrarla de una manera legible y útil a el docente, para esto hace uso de:

- **Vistas:** las vistas estan construidas en el motor de plantilla *EJS*, éste esta basado en plantillas que usan *javascript* para potenciar las existentes y dotar de nuevas características a el tradicional *HTML*, este motor de plantillas recibe objetos y los procesa para mostrarlos de manera útil al usuario final.
- **Enrutador:** El enrutador es el componente que se encarga de dirigir las peticiones realizadas en el navegador del usuario hacia el proceso que ejecuta la acción que va ligada a la anterior petición, este componente esta desarrollado en el *framework NODE JS*, cada ruta es única y están basados en los métodos de solicitud de *HTML*.
- **Controlador:** Es el componente que contiene la lógica de negocio, en el se hacen los procesos que dan utilidad a la información. para la evaluación de los tres criterios definidos en la estrategia evaluativa se usan:

- **Contador de coincidencias en respuestas:** Cuando el docente por medio del navegador lanza la petición para visualizar las coincidencias entre alumnos de un examen en una unidad, el contador de coincidencia usa el identificador de unidad para buscar en la base de datos únicamente los registros de respuesta al examen de esa unidad y traerlos. Para clasificarlos, el contador de coincidencias guarda todas las respuestas del examen del primer alumno en una estructura de datos tipo vector y por medio de un ciclo iterativo las compara con el resto de respuestas de los alumnos restantes. Al identificar una respuesta coincidente en valor y posición, se incrementa un contador el cual al final es divido por el número de respuestas totales del examen en la unidad, generando el dato de porcentaje de coincidencias que llenara la posición en la matriz, repitiendo ese ciclo con todos los alumnos. De la Matriz resultante se clasifican como sospechosos de fraude a todos os estudiantes que tengan más del 90 % de coincidencias en sus respuestas. Este porcentaje se eligió con base en trabajos previos [51, 52].
- **Identificador de IPs coincidentes:** Al invocarse este proceso, el controlador busca entre la información de todos los alumnos que presentaron el examen de una unidad y guarda todas las direcciones IP de cada alumno, luego por un ciclo iterativo este compara cada una de esas IP consigo mismas, mientras que un contador se incrementa cuando exista una coincidencia exacta, al final se crea un vector con las IP que presenten una número de coincidencias superior a 1 y con cada una de ellas se agrupan en un listado que es mostrado al usuario. Éste esta compuesto de enlaces que llevan a visualizar los estudiantes que componen cada grupo de trabajo, es decir que coinciden en la misma IP, junto con las respuestas, el número de coincidencias entre ellos y el tiempo de envío.
- **Tiempo de envío:** Cuando se realiza cualquiera de los dos anteriores procesos, el sistema siempre incluye en su esquema de datos a consultar el tiempo de envío el cual es mostrado en una casilla homónima.

Por ultimo se muestra en la figura 3.16 la aplicación desplegada en tiempo real y en funcionamiento.

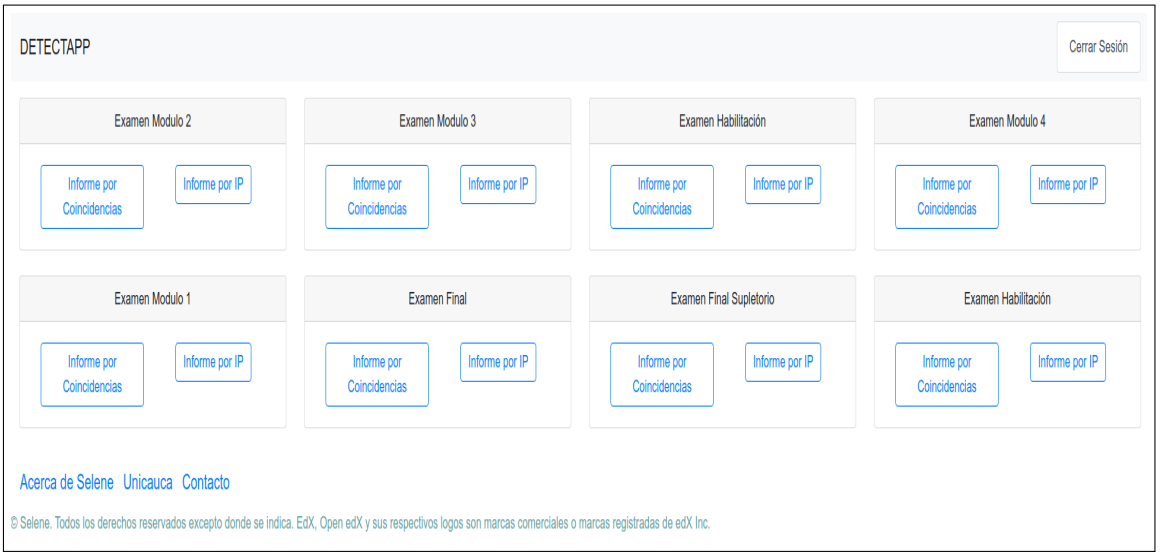


Figura 3.16: Interfaz de exámenes DetectApp

Capítulo 4

Caso de Estudio

Para realizar las pruebas de uso del mecanismo, se utilizó el curso piloto tipo MPOC "Introducción al emprendimiento con Lean Startup", ofrecido a estudiantes de la Universidad del Cauca y mediante el cual obtienen créditos dentro de los programas de formación de la universidad.

4.1. Descripción del curso piloto

El curso virtual Introducción al emprendimiento con *Lean Startup*, se propuso para ser desarrollado como una materia electiva FISH (Componente de Formación Social Integral y Humana) de la Universidad del Cauca y 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*.

El concepto de *Lean Startup* fue desarrollado en el año 2008 por Eric Ries, convirtiéndose en una metodología ampliamente utilizada como medio para validar las ideas de negocio antes de lanzarlas al mercado. Otros autores han aportado al conjunto de herramientas y métodos que pueden ser instanciados en el marco de la propuesta de Eric Ries.

La idea básica en el *Lean Startup*, es a través de la experimentación, determinar si nuestras hipótesis sobre una posibilidad de negocio son erradas y deben ser ajustadas. Es fundamentalmente una forma de aprender, poniendo a prueba nuestro entendimiento de los problemas y retro alimentando nuestro conocimiento.

Lean Startup es una herramienta que provee a emprendedores, constructores de negocios o empresas en sus etapas tempranas de operación, de un medio para captar clientes, aprender sobre sus necesidades y generar valor.

El curso fue diseñado y ejecutado por el equipo docente conformado por Mario Solarte Sarasty, Ingeniero en Electrónica y Telecomunicaciones de la Universidad del Cauca, Magíster en Ingeniería y Doctor en Ingeniería Telemática; Cristhian Figueroa, Ingeniero en Electrónica y Telecomunicaciones de la Universidad del Cauca, Magíster en Ingeniería Telemática y Doctor en Ingeniería Telemática y Doctor en Informática e Ingeniería de Sistemas (Universidad de Torino); Daniel Jaramillo, estudiante de Doctorado en Ingeniería Telemática, Administrador de Selene; Fabián Anacona, estudiante de Maestría en Ingeniería Telemática; Luís Alejandro Cruz, estudiante de Maestría en Ingeniería Telemática. Tuvo una duración de 14 semanas, comenzando el 16 de septiembre y terminando el 18 de diciembre de 2019, contando con la inscripción de estudiantes de diferentes programas de la Universidad del Cauca.

La temática del curso se dividió en tres temas principales:

- Unidad temática I: Emprendimiento de alto impacto innovador.
- Unidad temática II: Introducción a la Metodología Lean Startup.
- Evaluación Final.

Al ser el curso diseñado desde una perspectiva no presencial, se basa en el desarrollo de actividades de aprendizaje semanales soportadas en las tecnologías de Internet, donde el estudiante tiene acceso a diversos recursos didácticos construidos y seleccionados por el profesor independiente de otros que bien puede conseguir y consultar.

Además cuenta con actividades como la participación en foros de discusión, la elaboración de mapas conceptuales, uso de simuladores en línea, la presentación de pruebas en línea y el desarrollo de talleres no presenciales.

La plataforma Selene, permite el uso de servicios como preguntas y respuestas ademas de foros, para facilitar la interacción entre los distintos actores del proceso. En todo caso, se promoverá el desarrollo de estrategias discursivas argumentativas.

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: Encuesta de caracterización y test de estilos de aprendizaje en línea: 20 % de la nota total con la realización de 1 encuesta.
- Evaluación de la Unidad temática II: Examen en línea: 40 % de la nota total dividida en cuatro módulos con la realización de 4 exámenes.
- Evaluación Final: Examen en línea y encuestas de apreciaciones finales: 40 % de la nota total con la realización de 1 examen.

Durante la primera semana de clase se permite la interacción con los servicios de la plataforma tecnológica que soporta el curso, para que los estudiantes se familiaricen con las funciones y actividades que van a desarrollar durante el curso. Posteriormente, semana a semana, los estudiantes tienen acceso a un conjunto de recursos digitales que deben consultar previo al desarrollo de actividades de aprendizaje, algunas de ellas pueden ser calificables y tenerse en cuenta en la obtención de la evaluación del estudiante a lo largo del curso.

Los exámenes están compuestos por preguntas de opción múltiple que tienen una o mas respuestas correcta. Se plantearon de 12 a 14 preguntas con la posibilidad de presentar el examen en dos fechas diferentes.

Para la presentación de cada examen el estudiante tenía un límite de tiempo de 60 minutos para reducir al mínimo las posibilidades de fraude o de respuestas compartidas y por normativa de la universidad. Además, el curso cuenta con una ventaja,

ya que el acceso al curso está controlado por el administrador de la plataforma, esto hace imposible crear cuentas adicionales y realizar CAMEO [8].

Sin embargo se pueden presentar otros tipos de métodos de fraude [22], [6], [21], [41], como el compartirse respuestas por otros medios o el reunirse en grupos de estudiantes y resolver los exámenes. Por lo cual planteamos la creación de una estrategia evaluativa que permita generar condiciones para la identificación de estudiantes con sospecha de fraude.

4.2. Estrategia Evaluativa

Para el desarrollo de esta estrategia evaluativa se inicio haciendo una revisión sistemática de la literatura relacionada con los ambientes de aprendizaje en linea [4] [20], el fraude en ellos [22] [53] y los métodos evaluativos [54] [47], para conocer los avances y las falencias en el proceso evaluativo en un entorno *MOOC*.

Como se evidencia en [4], la virtualidad de un curso sumado a la posibilidad de obtener créditos académicos con su aprobación [20] aumentan las probabilidades de cometer fraude. Según [5] [35] [7] [41] los métodos evaluativos de la educación tradicional parecen no ser suficiente para mitigar o identificar estos comportamientos y los nuevos métodos de control [40] [4] [22] limitan los alcances de la educación masiva.

Por tanto para el diseño de esta estrategia evaluativa se buscó que esta sea usable, medible y no limite los alcances de la educación masiva. Para el caso de estudio la usabilidad se mide dependiendo de que pueda ser implementada sin alterar la distribución evaluativa, es decir el tipo de evaluación, la cual es de exámenes de opción múltiple. Por otra parte en que no altere los porcentajes de evaluación. El que la estrategia evaluativa sea medible se logra con la implementación de parámetros cuantificables que permitan emitir un criterio de sospecha de fraude en uno o mas estudiantes y el que no limite los alcances de la educación masiva implica que la estrategia no conlleve a la presencialidad o el uso de dispositivos electrónicos especializados.

Usabilidad.

Como se mencionó, para que la estrategia sea usable debe coincidir con la distribución evaluativa definida para el curso, por eso se partió de la idea de que el examen debe ser de tipo respuestas de opción múltiple. Una vez definido el tipo de examen se hizo una revisión sistemática de los tipos de fraude en estos exámenes [48] [29] y se encuentra que el fraude mas común es compartirse las respuestas entre alumnos.

En [55] se estudia la existencia de la posibilidad de que un estudiante pueda aprobar un examen de respuestas con opción múltiple basado únicamente en el azar a la hora de seleccionar las respuestas, aunque los resultados no son concluyentes se puede especular que la probabilidad es baja y se puede disminuir aún mas si se escogen métodos aleatorios para la creación de los exámenes.

Siendo que anteriormente el crear distintos exámenes y entregarlos de manera aleatoria a los estudiantes ya fue probado [54], con el objetivo de lograr detectar a la mayor cantidad de estudiantes que comparten sus respuestas, se desarrolló una estrategia evaluativa para que los estudiantes fueran detectados. La estrategia consistió en un método de aleatorización de 5 cuestionarios diseñados, en donde se varia el orden en que aparecen las opciones de respuesta para cada pregunta. Esto se logra mediante la creación de un banco de 5 cuestionarios previamente definidos en la plataforma Selene, en donde a cada estudiante del caso de estudio le apareció aleatoriamente uno de ellos. De esta manera si los estudiantes compartían sus respuestas sin percatarse del cambio sutil en el orden de las opciones de respuesta, iban a compartir las respuestas, pero tener calificaciones diferentes.

Se pensó así para no alterar la distribución evaluativa establecida, y además, no levantar sospechas en los estudiantes del caso de estudio que se pudieran alertar de la implementación de la estrategia. A su vez basados en el concepto de porcentaje de coincidencias en las respuestas, el cual define que si un estudiante coincide en el en más del 90 % de sus respuestas con otro alumno, éste es sospechoso de fraude. Es aquí donde nuestra estrategia aumenta estas sospechas, ya que si dos alumnos tienen la mismas respuestas pero ambos tienen cuestionarios diferentes, esto es un indicativo de peso para la sospecha de fraude en la presentación del examen de estos alumnos.

Medible.

Una vez definido el modo de evaluación de la estrategia evaluativa, debemos poder cuantificar los resultados obtenidos, para esto diseñamos una aplicación web que con base en los resultados y a unos parámetros que se definen a continuación muestre al docente los estudiantes con sospecha de fraude.

Examen con la estrategia evaluativa aplicada.				
Versión Original	**Versión 1**	**Versión 2**	**Versión 3**	**Versión 4**
Enunciado de la pregunta 1	**Enunciado de la pregunta 1**	**Enunciado de la pregunta 1**	**Enunciado de la pregunta 1**	**Enunciado de la pregunta 1**
¿Cuál es el objetivo principal de realizar entrevistas de validación de problema a potenciales clientes?	¿Cuál es el objetivo principal de realizar entrevistas de validación de problema a potenciales clientes?	¿Cuál es el objetivo principal de realizar entrevistas de validación de problema a potenciales clientes?	¿Cuál es el objetivo principal de realizar entrevistas de validación de problema a potenciales clientes?	¿Cuál es el objetivo principal de realizar entrevistas de validación de problema a potenciales clientes?
Respuesta correcta:	**Respuesta correcta:**	**Respuesta correcta:**	**Respuesta correcta:**	**Respuesta correcta:**
Identificar los problemas que frustran al cliente.	Identificar los problemas que frustran al cliente.	Identificar los problemas que frustran al cliente.	Identificar los problemas que frustran al cliente.	Identificar los problemas que frustran al cliente.
Opciones de respuesta.	**Opciones de respuesta.**	**Opciones de respuesta.**	**Opciones de respuesta.**	**Opciones de respuesta.**
A- Saber si el cliente está en nuestro interesado producto. B- Identificar los problemas que frustran al cliente. C- Establecer un precio para mi producto . D- Enseñarle al cliente los beneficios de mi producto.	A- Saber si el cliente está en nuestro interesado producto. B- Establecer un precio para mi producto. C- Identificar los problemas que frustran al cliente. D- Enseñarle al cliente los beneficios de mi producto.	A- Saber si el cliente está en nuestro interesado producto B- Establecer un precio para mi producto C- Enseñarle al cliente los beneficios de mi producto. D- Identificar los problemas que frustran al cliente.	A- Identificar los problemas que frustran al cliente. B- Establecer un precio para mi producto. C- Enseñarle al cliente los beneficios de mi producto. D- Saber si el cliente está en nuestro interesado producto.	A- Enseñarle al cliente los beneficios de mi producto. B- Establecer un precio para mi producto. C- Identificar los problemas que frustran al cliente. D- Saber si el cliente está en nuestro interesado producto.

Figura 4.1: Primera pregunta para cada versión del examen según la estrategía evaluativa

Dado que el objetivo de la estrategia evaluativa es identificar a los estudiantes con sospecha de fraude por medio del cambio de orden en las opciones de respuestas, pensamos que organizar 5 cuestionarios para un mismo examen (ver figura 4.1) de tal manera que el estudiante no se percatara que las repuestas están en un orden diferente, podría permitir identificar de manera más clara a aquellos estudiantes que son sospechosos de fraude. Pues al compartir sus respuestas pero presentar exámenes diferentes obtendrían una calificación diferente. Las calificaciones del curso van de 0 a 5 según normatividad del la Universidad del Cauca.

Respuestas en un grupo de trabajo al examen con distintas versiones.							
Estudiante 1		**Estudiante 2**		**Estudiante 3**		**Estudiante 4**	
Letra y opciones de respuesta correcta según la versión de examen.	**Letra y opciones que el Estudiante 1 respondió.**	**Letra y opciones de respuesta correcta según la versión de examen.**	**Letra y opciones que el Estudiante 2 respondió.**	**Letra y opciones de respuesta correcta según la versión de examen.**	**Letra y opciones que el Estudiante 3 respondió.**	**Letra y opciones de respuesta correcta según la versión de examen.**	**Letra y opciones que el Estudiante 4 respondió.**
1- **C** Identificar los problemas que frustran al cliente. 2- **D** Familiares y amigos, porque ayudarán a mejorar paulatinamente el guion de la entrevista. 3- **B** Una cafetería o zona neutral, porque entrevistado y entrevistador saldrán de su zona de confort y la entrevista será más natural. 4- **D** Para que el entrevistado se sienta identificado con la historia 5- **C** Cara a cara para poder advertir los gestos que haga el cliente	**1- C** Identificar los problemas que frustran al cliente. **2- B** Profesores, porque me ayudarán a mejorar los diálogos con los futuros entrevistados **3- B** Una cafetería o zona neutral, porque entrevistado y entrevistador saldrán de su zona de confort y la entrevista será más natural. **4- D** Para que el entrevistado se sienta identificado con la historia **5- C** Cara a cara para poder advertir los gestos que haga el cliente	1- **D** Identificar los problemas que frustran al cliente. 2- **B** Familiares y amigos, porque ayudarán a mejorar paulatinamente el guion de la entrevista. 3- **D** Una cafetería o zona neutral, porque entrevistado y entrevistador saldrán de su zona de confort y la entrevista será más natural. 4- **C** Para que el entrevistado se sienta identificado con la historia. 5- **D** Cara a cara para poder advertir los gestos que haga el cliente.	1- **C** Para que el entrevistado tenga una idea del nivel de dominio del problema del entrevistador 2- **D** Profesores, porque me ayudarán a mejorar los diálogos con los futuros entrevistados. 3- **B** La casa del entrevistado, porque estando en su zona de confort estará más cómodo. 4- **D** Para que el entrevistado tenga una idea del nivel de dominio del problema del entrevistador. 5- **C** Mediante videollamada para evitar los nervios y hacer que la entrevista salga mejor	1- **A** Identificar los problemas que frustran al cliente. 2- **C** Familiares y amigos, porque ayudarán a mejorar paulatinamente el guion de la entrevista. 3- **C** Una cafetería o zona neutral, porque entrevistado y entrevistador saldrán de su zona de confort y la entrevista será más natural. 4- **A** Para que el entrevistado se sienta identificado con la historia **5- B** Cara a cara para poder advertir los gestos que haga el cliente	1- **C** Enseñarle al cliente los beneficios de mi producto. 2- **D** Profesores, porque me ayudarán a mejorar los diálogos con los futuros entrevistados. 3- **B** La casa del entrevistado, porque estando en su zona de confort estará más cómodo. 4- **D** Para que el entrevistado tenga una idea del nivel de dominio del problema del entrevistador. 5- **C** Mediante videollamada para evitar los nervios y hacer que la entrevista salga mejor.	1- **C** Identificar los problemas que frustran al cliente. 2- **B** Familiares y amigos, porque ayudarán a mejorar paulatinamente el guion de la entrevista. 3- **D** Una cafetería o zona neutral, porque entrevistado y entrevistador saldrán de su zona de confort y la entrevista será más natural. 4- **A** Para que el entrevistado se sienta identificado con la historia 5- **D** Cara a cara para poder advertir los gestos que haga el cliente	1- **C** Identificar los problemas que frustran al cliente. 2- **D** Profesores, porque me ayudarán a mejorar los diálogos con los futuros entrevistados. 3- **B** La casa del entrevistado, porque estando en su zona de confort estará más cómodo 4- **D** Para que el entrevistado tenga una idea del nivel de dominio del problema del entrevistador. 5- **C** Mediante videollamada para evitar los nervios y hacer que la entrevista salga mejor.
Nota Estudiante 1		**Nota Estudiante 2**		**Nota Estudiante 3**		**Nota Estudiante 4**	
4		**0**		**0**		**1**	

Figura 4.2: Respuesta a una pregunta en un grupo de trabajo con un examen y distitnas versiones

Al suceder esto el primer parámetro medible es la diferencia de nota del alumno que transmitió las respuestas con respecto a los que las recibieron, esto se muestra en la figura 4.2. Se puede ver que para el estudiante 1 que compartió las respuestas con los otros 3 estudiantes, la primera pregunta tiene como opción correcta la opción **C Identificar los problemas que frustran al cliente** mientras que para los otros tres estudiantes las opciones correcta corresponden para el estudiante 2 la **D Identificar los problemas que frustran al cliente**, para el estudiante 3 la **A Identificar los problemas que frustran al cliente** y para el estudiante 4 la **C Identificar los problemas que frustran al cliente**, Sin embargo los estudiantes 2, 3 y 4 respondieron para la pregunta 1 la opción C debido a que el estudiante 1 así lo compartió generando que se equivoquen.

Al haber una gran cantidad de alumnos, hacer esta revisión de manera manual es tediosa y muy compleja, aquí es donde más impacto tiene la aplicación DetectApp, la cual por medio de procedimientos lógicos e iterativos compara todas las respuestas de los estudiantes que presentan un examen y cuenta el número de coincidencias que tienen entre si y con base en esto calcula un porcentaje de coincidencias. Si el porcentaje de coincidencias es superior o igual a 90 % *DetectApp* lo clasifica como

sospechosos de fraude, este es el primer parámetro para la sospecha de fraude. Luego, *DetectApp* organiza los sospechosos los muestra en un listado al docente. Para poder identificar quien es el estudiante que comparte las respuestas, se muestra el tiempo de envío de cada alumno, siendo este el segundo parámetro de clasificación.

Como tercer parámetro se tienen en cuenta los grupos de trabajo como conducta fraudulenta, ya que por la naturaleza del curso y como se explica al comienzo del mismo, los exámenes deben presentarse individualmente y sin ayuda de medios exteriores. Para identificar los estudiantes que conforman grupos de trabajo se hace uso de la información obtenida de las interacciones de los alumnos con la plataforma Selene, en este caso específicamente de la dirección IP de cada uno, así al compararlas con un proceso lógico e iterativo, se encuentra el listado de direcciones IP que presentan coincidencia en mas de un alumno, estas se almacenan y luego se consulta los estudiantes que estén asociados a estas direcciones. Al finalizar la consulta se muestra al profesor un listado con las direcciones IP que se repiten, cada dirección muestra al docente el listado de estudiantes relacionados a ella con sus respectivas respuestas, tiempo de envío y número de coincidencias.

Por ultimo esta estrategia evaluativa no limita las características de masividad del curso MPOC ni cambia el transcurso del mismo.

4.3. Aplicación sobre el caso de estudio

Definida la estrategia de evaluación, se aplicó sobre el curso piloto tipo MPOC "Introducción al emprendimiento con Lean Startup", la cantidad y forma de los exámenes fue definida por el docente de esta manera:

- Evaluación de la Unidad temática I: Examen, encuesta de caracterización y test de estilos de aprendizaje en línea: 20 % de la nota total con la realización de 1 encuesta.
- Evaluación de la Unidad temática II: Examen en línea: 40 % de la nota total dividida en cuatro módulos con la realización de 4 exámenes.

- Evaluación Final: Examen en línea y encuestas de apreciaciones finales: 40 % de la nota total con la realización de 1 examen.

Como en la Unidad Temática I, no se realizan exámenes sino encuestas de caracterización y test de apreciación, en esta unidad no se aplicó la estrategia evaluativa. En la Unidad Temática II se definieron cuatro exámenes, fue en ésta donde inició la aplicación de la estrategia.

Unidad Temática II:

- Examen 1: Este examen inicial se uso como referencia, en este no se aplico la estrategia evaluativa, pues se quizo observar el comportamiento de los estudiantes frente a un típico examen en un ambiente MPOC, sin embargo si se hicieron las mediciones de respuestas coincidentes y grupos de trabajo por IP, para comparar con los resultados de la siguiente evaluación.
- Examen 2: En este examen comienza la implementación de la estrategia evaluativa, el docente diseña un examen de 12 preguntas, con el cual se hizo una aleatorización en el orden de las respuestas de cada pregunta generando así cinco versiones del mismo, usando el componente de banco de preguntas de la plataforma Selene, se guardaron las cinco versiones del examen y este mismo se encarga de entregarlas aleatoriamente a los estudiantes del curso.

  Se comparo los resultados del segundo examen con el primero para identificar que tantos grupos de trabajo como estudiantes reincidentes en la sospecha de fraude ubieron.
- Examen 3: Para el examen 3, se decide no aplicar la estrategia evalutiva, para analizar el comportamiento de los estudiantes y grupos de trabajo reincidentes, se quiere saber si dicha sospecha de fraude se mantiene aún con un examen típico de ambiente *MOOC*.
- Examen 4: Nuevamente se aplica la estrategia evaluativa, esta vez se desea confirmar los grupos de trabajo que se han consolidado a través del curso y verificar si varia el número de estudiantes con sospecha de fraude de una examen sin estrategia a otro con la aplicación de estrategia.

Examen Final:

Para el examen final se aplica la estrategia evaluativa, el objetivo es observar si los grupos de trabajo detectados se mantienen, y ver la variación en el número de estudiantes con sospecha de fraude entre dos exámenes con la aplicación de la estrategia.

Existe la probabilidad de que un alumno coincida en respuestas con otro debido a que por azar la plataforma Selene les asigna el mismo examen, para descartar esta probabilidad se hace un seguimiento a los estudiantes con sospecha de fraude identificados en el primer examen y se hace alternancia en la aplicación de la estrategia evaluativa entre el examen 1 y el examen 4, dicho método de aplicación arroja resultados positivos que se muestran en el siguiente capítulo.

Por otra parte, aunque en la aplicación de la estrategia evaluativa exista la posibilidad que los estudiantes se percaten del cambio en el orden de las respuestas entre distintos cuestionarios, dicha posibilidad se presenta si existe algún tipo de comunicación mientras presentan el examen, esta comunicación se tipifica como fraude. Es ahí donde la identificación de grupos de trabajo en un mismo sitio sirve como indicador de sospecha de fraude. Por lo tanto, la herramienta le muestra al profesor, por cada examen, una lista de direcciones IP en las que al seleccionar una de ellas despliega un listado de alumnos con sus nombres, fecha de entrega, hora de entrega y sus respuestas. Esto se refleja en los resultados que se muestran en el capítulo 5.

Capítulo 5

Resultados

Esta sección presenta los resultados del diseño, programación y ejecución de la estrategia evaluativa, la cual fue aplicada en los exámenes correspondiente al Modulo 2, modulo 4 y examen final de la Unidad Temática 2 en el caso de estudio. La herramienta fue creada con el fin de visualizar los datos obtenidos de la aplicación de la estrategia evaluativa y en general el comportamiento en los exámenes del caso de estudio se construyo la herramienta DetectApp.

5.1. Aplicación Web: *DetectApp*

La aplicación web *DetectApp* ordena, compara y clasifica la información relacionada con los exámenes de los estudiantes del caso de estudio, para luego visualizar en listas o grupos a los detectados como sospechosos de fraude.

Para hacer uso de la aplicación, el docente debe iniciar sesión con sus credenciales, las cuales se componen de un nombre de usuario que debe ser en formato correo electrónico y una contraseña con un mínimo de 8 caracteres como se muestra en la figura 5.1.

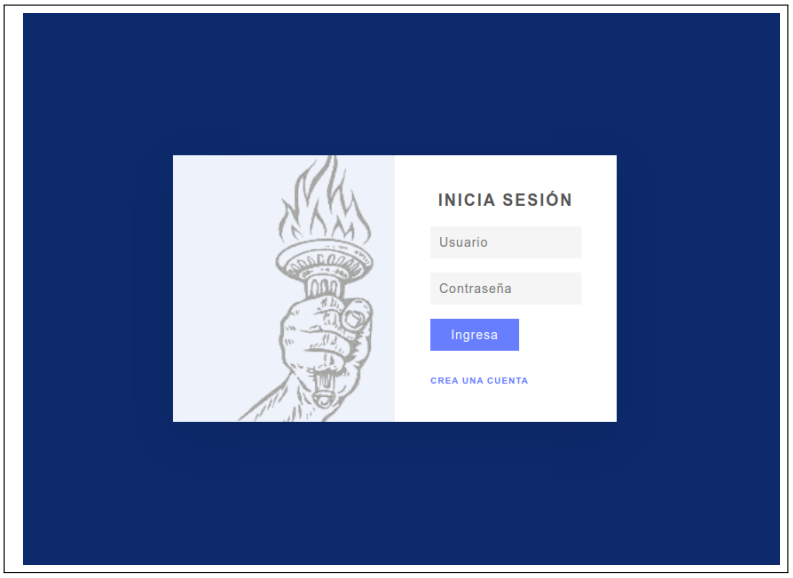


Figura 5.1: Interfaz Inicio de Sesión de la aplicación *DetectApp*

La figura 5.2 muestra la vista del administrador de la herramienta donde se muestran los seis cursos activos para el periodo del caso de estudio, en donde cada curso tiene su correspondiente informe.

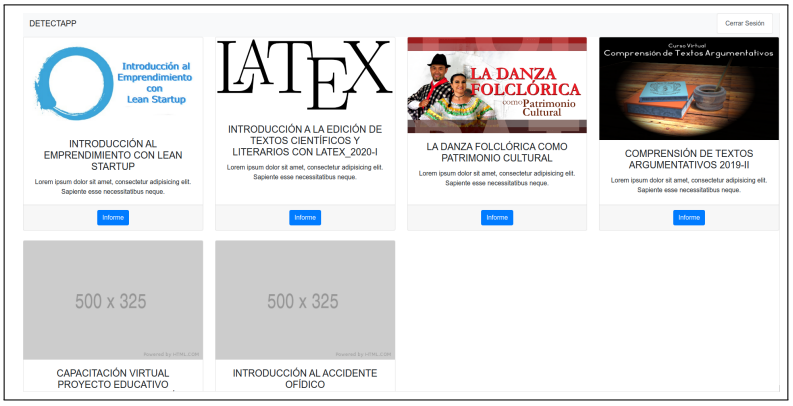


Figura 5.2: Interfaz Cursos de la aplicación *DetectApp*

Desde la aplicación, el docente puede seleccionar un curso para ver su información, siendo "INTRODUCCIÓN AL EMPRENDIMIENTO CON LEAN STARTUP.[el] caso de estudio, cada curso se identifica con la imagen de presentación que tiene en la plataforma Selene, junto al nombre exacto que presenta en la misma, y una breve

descripción tomada de la pagina de presentación del curso. En la parte inferior de la tarjeta del curso se muestra un botón azul que es el encargado de conducir al docente al listado de exámenes en el caso de estudio.

La figura 5.3 muestra todos los exámenes que contiene el curso seleccionado por el profesor. Con la posibilidad de observar dos tipos de informe según la estrategia evaluativa.

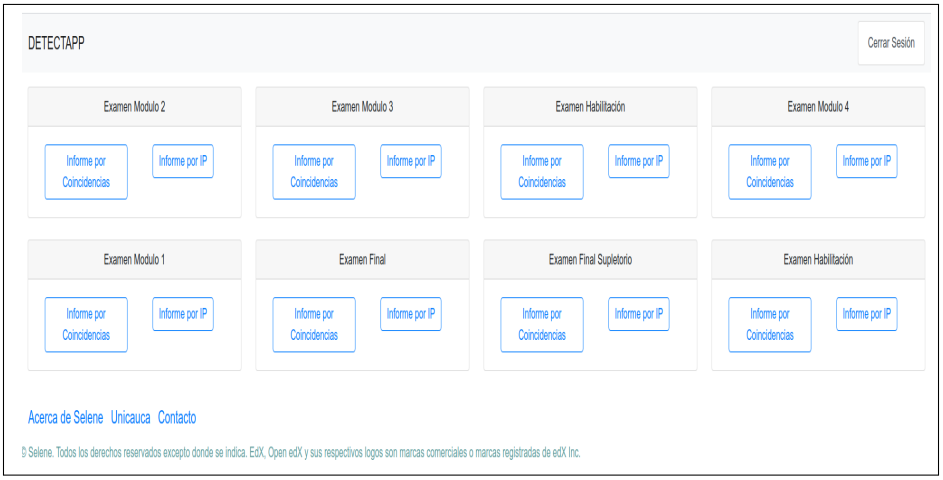


Figura 5.3: Interfaz Exámenes del caso de estudio *DetectApp*

En la interfaz de la figura 5.4 se muestra el informe de estudiantes que coinciden en un 90 % o mas en sus respuesta. Este listado tiene cuatro opciones para visualizar la información, la primera identificada mediante el botón “Listado de estudiantes por coincidencias”, esta muestra todos los estudiantes que fueron detectados con el parámetro de coincidencia. Para la escritura del presente trabajo, los datos de los estudiantes fueron anonimizados. El listado de estudiantes en la aplicación, aparece con su correspondiente nombre.

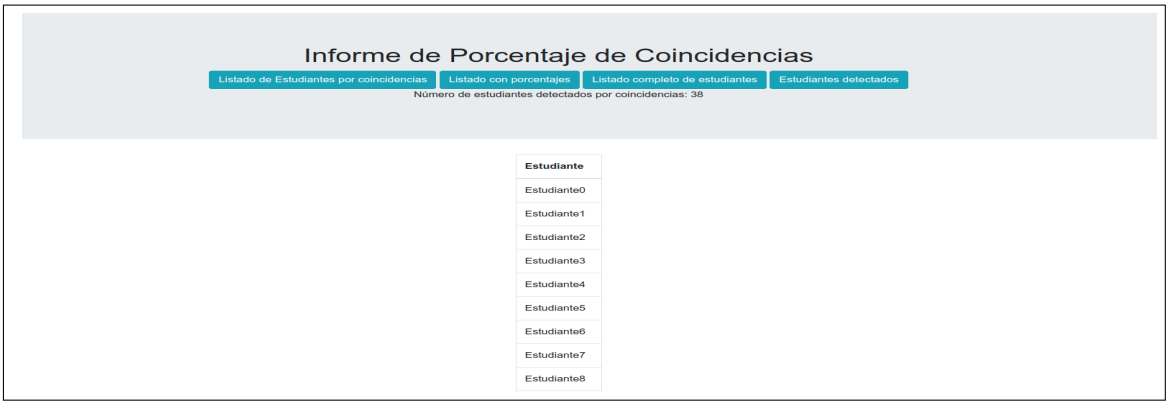
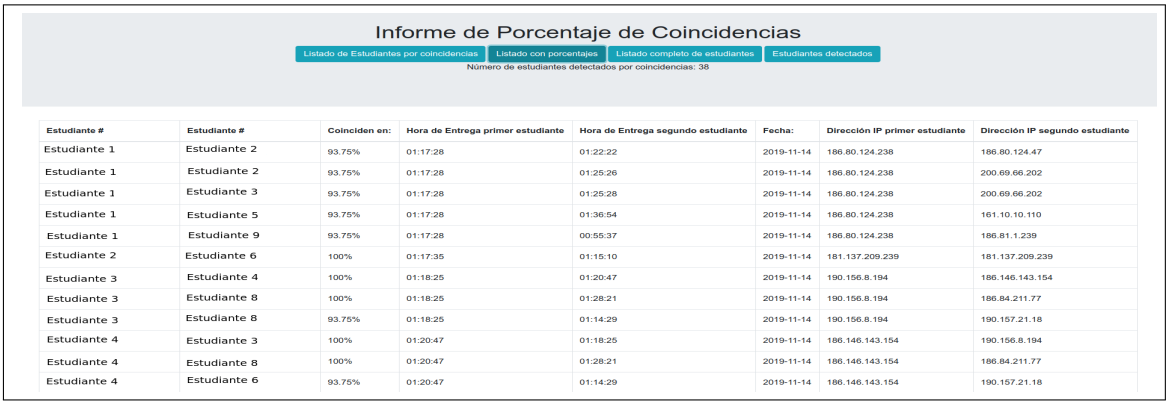


Figura 5.4: Interfaz listado de estudiantes con coincidencias en respuestas superior al 90 % *DetectApp*

La segunda opción, identificada por el botón "Listado con porcentajes"muestra el listado anterior con mas detalles, como el porcentaje de coincidencia exacto entre estudiantes, las horas, las direcciones IP y la fecha del envío de respuestas, esto se puede evidenciar en la figura 5.5 .



Estudiante #	Estudiante #	Coinciden en:	Hora de Entrega primer estudiante	Hora de Entrega segundo estudiante	Fecha:	Dirección IP primer estudiante	Dirección IP segundo estudiante
Estudiante 1	Estudiante 2	93.75%	01:17:28	01:22:22	2019-11-14	186.80.124.238	186.80.124.47
Estudiante 1	Estudiante 2	93.75%	01:17:28	01:25:26	2019-11-14	186.80.124.238	200.69.66.202
Estudiante 1	Estudiante 3	93.75%	01:17:28	01:25:28	2019-11-14	186.80.124.238	200.69.66.202
Estudiante 1	Estudiante 5	93.75%	01:17:28	01:36:54	2019-11-14	186.80.124.238	161.10.10.110
Estudiante 1	Estudiante 9	93.75%	01:17:28	00:55:37	2019-11-14	186.80.124.238	186.81.1.239
Estudiante 2	Estudiante 6	100%	01:17:35	01:15:10	2019-11-14	181.137.209.239	181.137.209.239
Estudiante 3	Estudiante 4	100%	01:18:25	01:20:47	2019-11-14	190.156.8.194	186.146.143.154
Estudiante 3	Estudiante 8	100%	01:18:25	01:28:21	2019-11-14	190.156.8.194	186.84.211.77
Estudiante 3	Estudiante 8	93.75%	01:18:25	01:14:29	2019-11-14	190.156.8.194	190.157.21.18
Estudiante 4	Estudiante 3	100%	01:20:47	01:18:25	2019-11-14	186.146.143.154	190.156.8.194
Estudiante 4	Estudiante 8	100%	01:20:47	01:28:21	2019-11-14	186.146.143.154	186.84.211.77
Estudiante 4	Estudiante 6	93.75%	01:20:47	01:14:29	2019-11-14	186.146.143.154	190.157.21.18

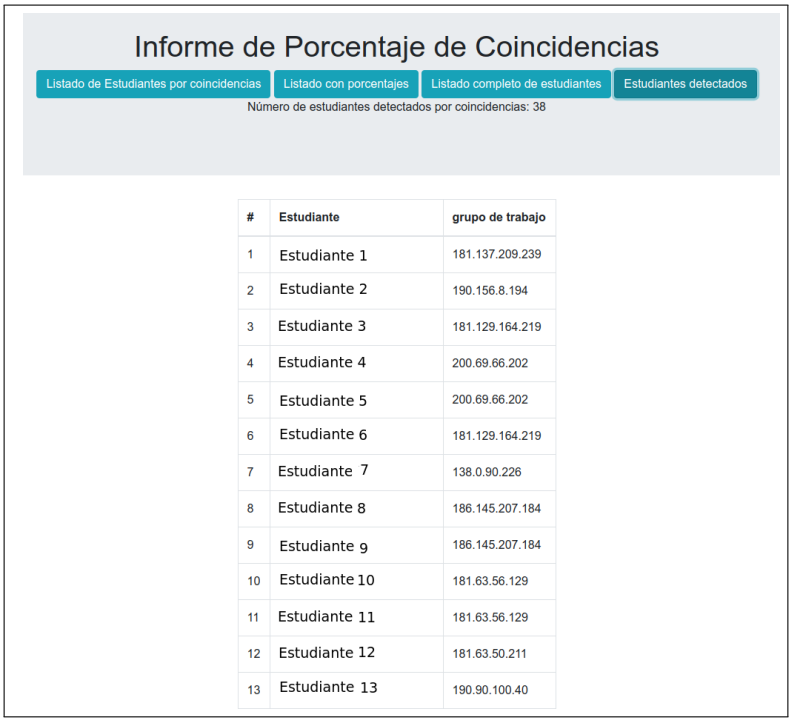
Figura 5.5: Interfaz listado de estudiantes con coincidencias en respuestas superior al 90 % detallada en *DetectApp*

La tercera opción, como se muestra en la figurar 5.6 muestra un listado completo de los estudiantes que presentaron el examen junto con las respuestas y dirección de la pagina web donde se aloja el examen en la plataforma Selene.

N°	Estudiante	Dir IP	Respuesta1	Respuesta2	Respuesta3	Respuesta4	Respuesta5	Respuesta6	Respuesta7	Respuesta8	Respuesta9	Respuesta10	Respuesta11
1	Estudiante 1	186.80.124.238	choice_3	choice_0	choice_1	choice_2	choice_1	choice_3	choice_2	choice_0	choice_2	choice_2	choice_3
2	Estudiante 2	181.137.209.239	choice_3	choice_0	choice_1	choice_1	choice_1	choice_2	choice_2	choice_0	choice_2	choice_2	choice_3
3	Estudiante 3	190.156.8.194	choice_3	choice_0	choice_3	choice_3	choice_1	choice_2	choice_2	choice_0	choice_2	choice_2	choice_3
4	Estudiante 4	67.73.224.4	choice_3	choice_0	choice_3	choice_3	choice_1	choice_2	choice_2	choice_0	choice_2	choice_3	choice_2
5	Estudiante 5	186.146.143.154	choice_3	choice_0	choice_3	choice_3	choice_1	choice_2	choice_2	choice_0	choice_2	choice_2	choice_3

Figura 5.6: Interfaz listado completo de estudiantes con sus respuestas en *DetectApp*

La ultima opción identificada con el botón “Estudiantes detectados”muestra el listado de estudiantes en esa unidad que coinciden con los tres parámetros de la estrategia evaluativa, es decir los estudiantes que presentan un 90 % o mas de coincidencias entre sus respuestas, presentan la misma dirección IP y diferencia en tiempos de envío cercanos a un minuto o menos como se ve en la figura 5.7.



#	Estudiante	grupo de trabajo
1	Estudiante 1	181.137.209.239
2	Estudiante 2	190.156.8.194
3	Estudiante 3	181.129.164.219
4	Estudiante 4	200.69.66.202
5	Estudiante 5	200.69.66.202
6	Estudiante 6	181.129.164.219
7	Estudiante 7	138.0.90.226
8	Estudiante 8	186.145.207.184
9	Estudiante 9	186.145.207.184
10	Estudiante 10	181.63.56.129
11	Estudiante 11	181.63.56.129
12	Estudiante 12	181.63.50.211
13	Estudiante 13	190.90.100.40

Figura 5.7: Listado de estudiantes que cumplen los tres parámetros de detección en *DetectApp*

En la figura 5.8 se muestra la interfaz de informe por grupos de trabajo basado en el parámetro de detección de IPs coincidentes, en este se muestra inicialmente la cantidad de grupos de trabajos detectado en un examen y se agrupan por direcciones IP, para visualizar cada grupo el docente debe seleccionar uno y dar clic en el botón "ver grupo de trabajo".

Figura 5.8: Listado de grupos de trabajo por IP de detección en *DetectApp*

Cuando se selecciona la opción “ver grupo de trabajo”, la aplicación *DetectApp* nos muestra un listado con los estudiantes que coinciden en hora de envío y en dirección IP de entrega, junto con su porcentaje de coincidencia y la fecha, adicional a eso se muestra una tabla con los estudiantes que además de coincidir con el parámetro grupo de trabajo, coinciden con los otros dos parámetros de detección de la estrategia evaluativa como se muestra en la figura 5.9.

Figura 5.9: Listado de grupos de trabajo por IP de detección en *DetectApp*

Estas son las interfaces que la aplicación *DetectApp* ofrece para visualizar la información que ha sido clasificada según los tres parámetros de la estrategia evaluativa, a continuación se muestra los resultados por cada uno de los exámenes en el caso de estudio.

5.2. Uso de la herramienta sobre el caso de estudio

Para mostrar los resultados obtenidos al aplicar la estrategia evaluativa hemos seguido el procedimiento definido en la sección 4.3 de este documento, por tanto se muestra los resultados obtenidos por cada uno de los exámenes.

5.2.1. Unidad Temática II - Examen 1.

Este examen es usado como referencia para analizar el comportamiento de los estudiantes sin la aplicación de una estrategia evaluativa, sin embargo se uso los tres parámetros definidos en la sección 4.2 para obtener estadísticas e identificar estudiantes con comportamiento de fraude.

La herramienta permite visualizar dos tipos de informe, el primero es el informe por coincidencias que como se define en la sección 4.2 y que esta basado en el número de coincidencias que tienen los estudiantes entre si, al hacerlo la aplicación *DetectApp* encuentra que: de 92 estudiantes que presentaron el examen, 38 tienen más del 90 % de coincidencias de respuestas entre ellos. Dado que en este caso para todos los estudiantes el examen es el mismo, este dato por si solo no es diciente, pero sirve como dato de referencia para comparar el número de estudiantes con porcentaje de coincidencias en un examen donde se aplique la estrategia evaluativa frente a un examen de opción múltiple tradicional.

El informe por grupos de trabajo nos proporciona información mas diciente en este escenario ya que realizar el examen en grupo si es una clara conducta de fraude, al consultar este informe *DetectApp* nos muestra que existe 15 grupos de trabajo, de estos grupos de trabajo solo 3 grupos tienen un porcentaje de coincidencia inferior al 90 %, 8 grupos sus respuestas coinciden en un 100 % y 4 grupos están entre 90 % y menos de 100 %.

De los 38 estudiantes detectados por el informe de coincidencia, 22 estudiantes además de coincidir entre si en mas de 90 % sus respuestas, también son detectados por el parámetro tiempo de envío y grupo de trabajo definido en 4.2. Estos 22 alumnos constituyen el grupo encontrado con sospecha de fraude por *DetectApp* para el examen número 1 de la unidad temática 2, reinciden en su comportamiento.

5.2.2. Unidad Temática II - Examen 2.

La estrategia evaluativa definida en 4.2 comienza a ser aplicada en este examen. En el se encuentran datos y comportamientos de gran valor que serán descritos a lo

largo de este capítulo.

En el informe de coincidencias se encuentra un leve aumento comparado con el examen I donde no se aplica la estrategia evaluativa, en el examen número 2 se detecto 39 estudiantes que superan el 90 % de coincidencias de respuestas entre si frente a 38 estudiantes del anterior examen.

en el informe de grupos de trabajo hay una disminución de estos al haber solo 12 grupos comparado con los 15 detectados en el primer examen, de estos grupos de trabajo 6 coinciden en un 100 % de sus respuestas, 4 en menos de 90 % y 2 entre un 90 % y menos de 100 %. Como era de esperarse, pues los exámenes son diferentes para cada estudiante.

Solo 16 estudiantes de los 39 detectados por el informe de coincidencia constituyen el grupo encontrado con sospecha de fraude por *DetectApp* para el examen número 2 de la unidad temática 2.

5.2.3. Unidad Temática II - Examen 3.

Para el examen número 3 no se aplico la estrategia evaluativa para ver si se repiten el comportamiento mostrado en el examen 1 , los informes de coincidencia y grupos de trabajo reflejan datos similares al primer examen.

37 estudiantes fueron detectados en el informe de coincidencias frente a 38 encontrados en el primer examen, de esos 37 detectados 14 han sido detectados tanto en el primer examen como en el segundo, los restantes 24 no han sido detectados en ninguno de los dos exámenes previos.

DetectApp encontró 12 grupos de trabajo para el examen 3, de esos 12 grupos de trabajo 4 son reincidentes en esta conducta respecto a los anteriores exámenes, cabe resaltar que los 4 reincidentes coinciden en un 100 % sus respuestas entre si, 7 grupos están entre un 90 % y 99 % de coincidencias y solo uno presenta un porcentaje de coincidencias menor al 90 %.

Para este examen, *DetectApp* encontró 19 estudiantes que coinciden con los tres

parámetros de detección de la estrategia evaluativa, siendo 14 estudiantes reincidentes respecto al primer examen y 15 al segundo examen.

5.2.4. Unidad Temática II - Examen 4.

En este examen nuevamente se aplica la estrategia evaluativa. Este es el último examen de la unidad temática II. Para este tiempo ya ha transcurrido mas del 50 % de la duración del semestre y los resultados arrojan que ya hay grupos de trabajo consolidados y el número de reincidentes tiende a ser el mismo.

EL informe de coincidencia muestra que 23 alumnos coinciden con mas del 90 % de sus respuestas entre si, de esos 23, 13 reinciden con la misma conducta en los examen 2, 17 con el examen 1 y 16 con el examen 3.

11 son los grupos de trabajo detectados en este examen, 10 de los cuales coinciden en sus respuestas en menos del 90 %, solo un grupo coincide en el 100 % de sus respuestas.

Para este examen *DetectApp* encontró 10 estudiantes con sospecha de fraude según sus parámetros de decisión, de esos 10, 9 son reincidentes en cuanto a los anteriores exámenes.

5.2.5. Unidad Temática II - Examen Final

Para el examen final se aplica la estrategia evaluativa, se quiere observar el comportamiento de los estudiantes cuando la estrategia evaluativa es usada de manera continua.

El listado de estudiantes detectados en el informe de coincidencias es de 10 estudiantes, de los 10 detectados solo 4 son reincidentes en este comportamiento con respecto a los anteriores exámenes. Se podría deber a que al aplicar la estrategia evaluativa, más estudiantes se percatan de la misma.

En cuanto al número de grupos de trabajo se mantienen igual al del examen 4, el comportamiento es igual con tan solo un grupo con coincidencias en respuesta superior al 100 %.

DetectApp encuentra 5 estudiantes que cumplen con los tres parámetros de detección definidos en la sección 4.2, el total de detectados son reincidentes en todos los exámenes anteriores, estos 5 componen el grupo de sospecha de comportamiento de fraude del caso de estudio.

5.3. Análisis de Resultados.

Para el análisis de resultados comparamos las cifras obtenidas en la sección 5.2. Con base en estas, se realizan gráficos con el objetivo de evidenciar el comportamiento del caso de estudio.

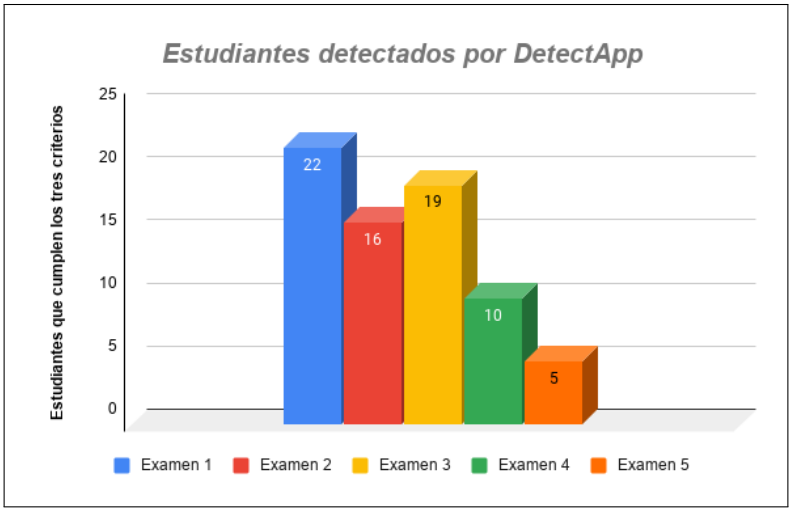


Figura 5.10: Captura e Identificación de estudiantes sospechosos en la aplicación *DetectApp*

La figura 5.10 muestra la cantidad de estudiantes detectados por la aplicación *DetectApp*, estos son los estudiantes que cumplen con todos los tres criterios definidos en la sección 4.2. En el primer examen se puede observar que 22 estudiantes presentan sospecha de fraude. Para este examen, el orden de los enunciados de las respuestas

no ha sido alterado como lo plantea la estrategia evaluativa. Se clasifican a los 22 estudiantes por grupos de trabajo que presenten la misma IP encontrando 10 grupos de trabajo y en cada uno de estos las notas de cada miembro son menores a cinco, lo que supone que se equivocan en contestar en alguna pregunta y los porcentajes de coincidencia en respuesta difieren en menos de un 10 %. Además, los tiempos de envío son muy cercanos, lo que puede corresponder a un comportamiento con sospecha de fraude.

Para el segundo examen donde se aplica la estrategia evaluativa, se pudo observar que hay 16 estudiantes detectados. Como en este examen el orden de los enunciados de las respuestas cambian dando origen a cinco versiones diferentes de un mismo examen, el compartir las respuestas entre estudiantes conlleva a que al menos uno de ellos tenga menor nota, prueba de esto es el caso donde *DetectApp* identifica a tres estudiantes con comportamiento con sospecha de fraude, por privacidad los estudiantes serán llamados así:

- Estudiante 1
- Estudiante 2
- Estudiante 3

Estos estudiantes presentan la misma dirección IP, en la misma fecha de presentación del examen, con diferencia de menos de un minuto en el tiempo de envío y un porcentaje de coincidencia del 100 % en las respuestas, al revisar las notas de los mismos se encontró lo mostrado en la siguiente tabla.

Notas Examen 2			
Estudiante	versión de examen	Notas	Hora de envío
Estudiante 1	1	2.2	22:21:33
Estudiante 2	3	1.1	22:21:35
Estudiante 3	5	1.1	22:21:38

Las notas confirman las sospechas de fraude al haber dos de los tres estudiantes detectados que presentan notas bajas con respecto al estudiante que comparte las respuestas, lo que supone que dichos estudiantes se equivocaron en más del 50 % de sus respuestas de manera idéntica, pues tienen un porcentaje de coincidencia del 100 %. Esto es lo que buscábamos encontrar con la estrategía, pues la idea es que los estudiantes cayeran en la trampa de compartir sus respuestas sin que se dieran cuenta que son exámenes diferentes, lo que muestra la efectividad de la estrategia evaluativa a la hora de permitir la identificación de conductas con sospecha de fraude. Otro caso similar es el de 2 estudiantes que serán llamados:

- Estudiante 3
- Estudiante 4

A estos estudiantes en particular se le asigno la misma versión de exámenes, pero *DetectApp* los clasifica como sospechosos de fraude al identificar que presentan la misma IP, el mismo tiempo de envío y el 100 % de coincidencia en las respuestas, dado que ambos tienen la misma versión del examen se puede pensar que es normal que ambos tengan las mismas respuestas si todas fuesen correctas, pero al hacer una revisión de sus notas se encuentran que son menor a 5 por ende presentan las mismas respuestas erradas. Ellos se equivocan de manera exacta en un 20 % del examen.

Notas Examen 2					
Estudiante	versión de examen	Cantidad de Aciertos	Cantidad de Errores	Notas	Hora de envío
Estudiante 1	5	8	2	4.0	22:25:33
Estudiante 2	5	8	2	4.0	22:25:05

Como este ultimo caso existe 4 más, con tres grupos de tres estudiantes y uno de dos, donde en cada grupo a los estudiantes por azar se les asigno la misma versión de examen pero *DetectApp* los clasifica como sospechosos de comportamiento fraudulento, y al revisar a fondo sus notas y respuestas se encuentra que presentan los mismos errores. Esto da indicios que aún teniendo la misma versión de los exámenes, *DetectApp* permite la identificación de conducta con sospecha de fraude.

En el tercer examen se encuentra un comportamiento similar al presentado en el primer examen, mientras que en el examen 4 y examen final se observa un cambio en el comportamiento de los estudiantes el cual se ve reflejado en la disminución de un 37 % y un 68 % respectivamente en la cantidad de estudiantes detectados respecto al segundo examen.

Para entender este comportamiento se analizo a los estudiantes en cada examen, basados en los parámetros de detección explicados en la sección 4.2. El primer parámetro que se analizo fue el parámetro de coincidencia, los resultados de este se reflejan en la figura 5.11

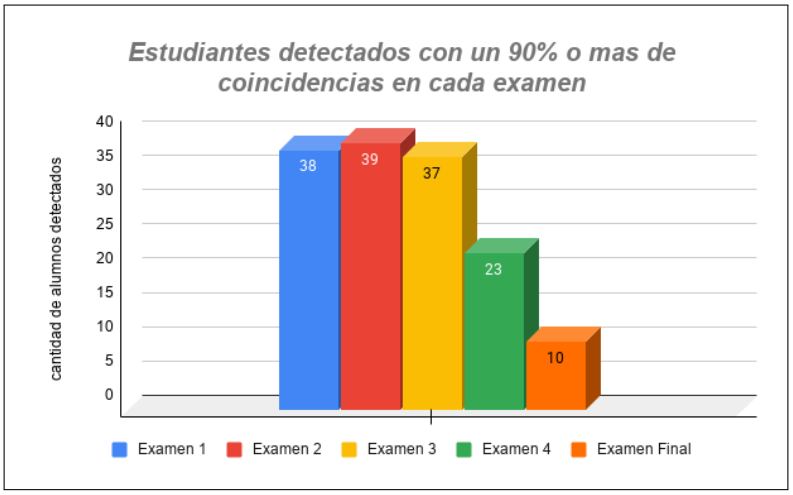


Figura 5.11: Cantidades de estudiantes sospechosos con un 90 % o mas de coincidencias en la aplicación *DetectApp* a lo largo del caso de estudio.

La figura 5.11 muestra la cantidad de estudiantes detectados con un 90 % o mas de coincidencias en cada examen, en los tres primeros exámenes se puede ver un comportamiento consistente. En el examen 4 y final se puede evidenciar una disminución considerable en la detección de estudiantes que presentan coincidencias en las respuestas de sus exámenes. Esto podría deberse a que los estudiantes cada vez más se percatan de estrategia aplicada. Y el hecho que la cantidad de estudiantes detectados con base en el parámetro de coincidencia haya disminuido, no es indicativo que el comportamiento de fraude también lo haga. Prueba de esto se muestra en la gráfica 5.12, donde la cantidad de grupos de trabajo se mantiene en cifras cercanas.

Es decir, la colaboración entre estudiantes continua, pero ahora son conscientes de las diferencias en los exámenes.

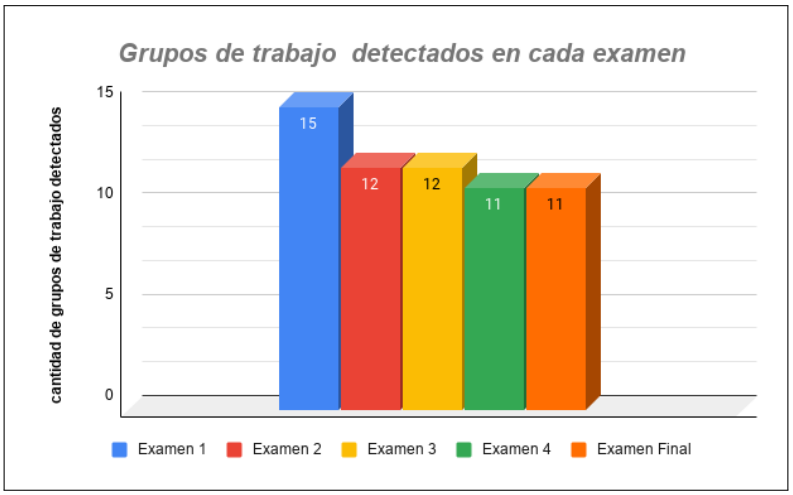


Figura 5.12: Cantidad de grupos de trabajo sospechosos en la aplicación *DetectApp* a lo largo del curso.

Si comparamos la cantidad de estudiantes detectados por *DetectApp* que cumple con los tres parámetros y los detectados sólo por el grupo de trabajo obtenemos la gráfica 5.13.

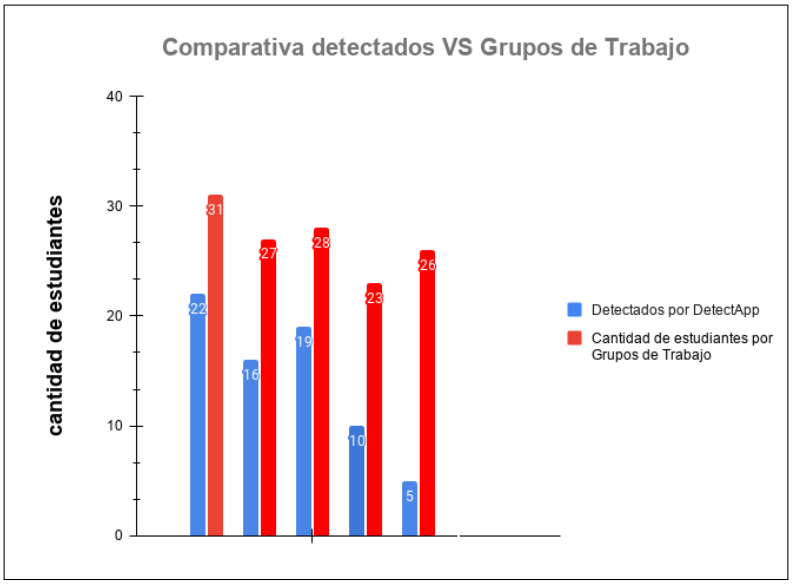


Figura 5.13: Gráfica Comparativa de la cantidad detectados VS la cantidad de estudiantes en los grupos de trabajo por *DetectApp*

Esta gráfica muestra que el número de estudiantes que forman grupos de trabajo es mayor y algo consistente, la cantidad de grupos casi se mantiene durante todo el curso, a diferencia de la cantidad de estudiantes detectados que cumplen con los tres parámetros de identificación. La hipótesis es que los estudiantes al estar en el mismo lugar presentando el examen evidencian la estrategia evaluativa y la sortean, para analizar dicha hipótesis se realiza la gráfica 5.14, que muestra la cantidad de estudiantes que reinciden en formar grupos de trabajo a lo largo de los 5 exámenes.

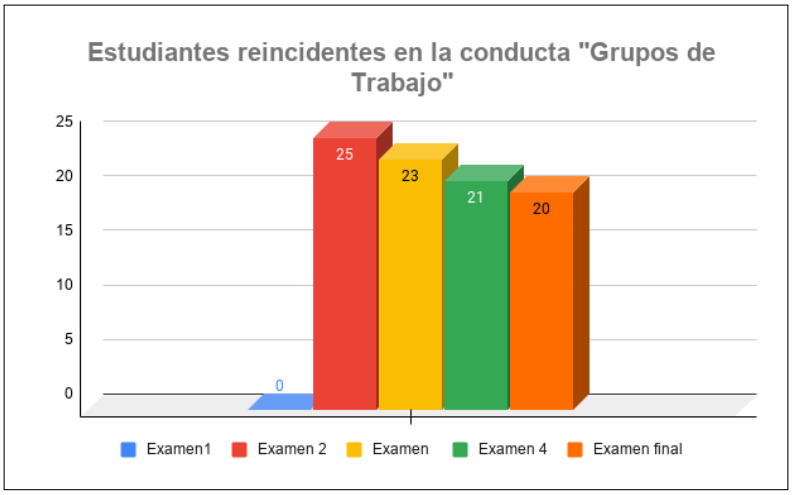


Figura 5.14: Cantidad de estudiantes que reinciden en formar grupos de trabajo y son detectados por *DetectApp*

Por otro lado, tomamos cuatro estudiantes que dejaron de ser detectados por los tres parámetros descritos en la sección 4.2 en el examen 4 y final, pero que son reincidentes en formar grupos de trabajo. Se analiza su comportamiento y se encuentra lo siguiente:

Para este caso, los estudiantes serán llamados así:

- Estudiante reincidente 1
- Estudiante reincidente 2
- Estudiante reincidente 3
- Estudiante reincidente 4

De los cuatro estudiantes mencionados anteriormente, todos dejan de ser detectados por que el porcentaje de coincidencia es menor a 90 %. Los 4 estudiantes son detectados en el informe de grupo de trabajó del examen 1, 2 y 3. Cada grupo esta formado por 2 estudiantes, estando **Estudiante reincidente 1** y **Estudiante reincidente 3** en un grupo y los restantes en otro.

Al revisar las respuestas se encuentra que **Estudiante reincidente 1** y **Estudiante reincidente 3** presentan diferentes versiones de exámenes pero se equivocan en las mismas preguntas y tienen la misma cantidad de aciertos, de igual manera con **Estudiante reincidente 2** y **Estudiante reincidente 4**. Esta revisión fue realizada manualmente, con el fin de detectar a aquellos estudiantes que se dan cuenta de la estrategia y siguen trabajando en grupos.

Análisis de dos grupos reincidentes				
Estudiante	Grupo de Trabajo	cantidad de aciertos	cantidad de errores	Notas
Estudiante reincidente 1	A	12	2	4.5
Estudiante reincidente 2	A	12	2	4.5
Estudiante reincidente 3	B	11	3	4.0
Estudiante reincidente 4	B	11	3	4.0

Cabe resaltar que estos cuatros estudiantes fueron tomados al azar del grupo de estudiantes reincidentes en formar grupos de trabajo. Luego repetimos el experimento tomamos dos grupos más al azar y se obtuvo un comportamiento similar. Esta vez se hizo en el examen final y los resultados se muestran en la siguiente tabla:

Análisis de dos grupos reincidentes				
Estudiante	Grupo de Trabajo	cantidad de aciertos	cantidad de errores	Notas
Estudiante reincidente 5	A	9	5	3.5
Estudiante reincidente 6	A	9	5	3.5
Estudiante reincidente 7	B	8	6	3.0
Estudiante reincidente 8	B	8	6	3.0

Estos estudiantes tomados al azar en exámenes diferentes, muestran un comportamiento que puede sustentar la hipótesis mencionada anteriormente, aun así, si dicha hipótesis es cierta, *DetectApp* permite la identificación de estos estudiantes con sospecha de fraude.

Como se menciona en [33] y [31] se piensa que los estudiantes que reinciden en sospecha de fraude tienen un comportamiento marcado a la hora de desarrollar un curso.

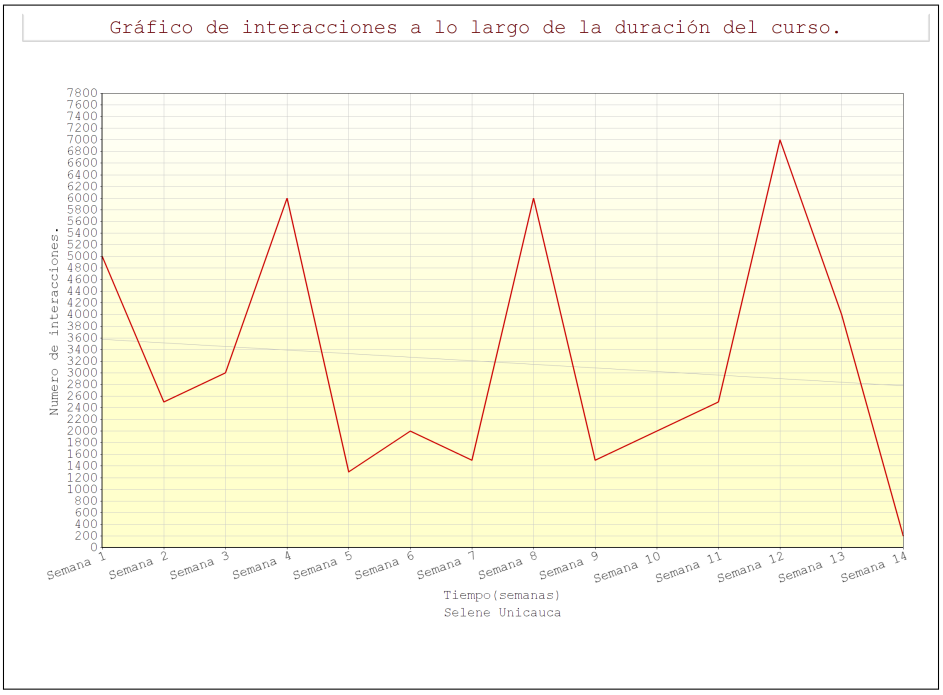


Figura 5.15: Gráfica de interacciones a través de la duración del curso.

Este comportamiento indica que los estos estudiantes no participan activamente en el curso y la mayoría de sus participaciones se realizan en las 24 horas previas al examen. En la Gráfica 5.15 se muestra picos de interacción más altos en las fechas cercanas a la programación de exámenes.

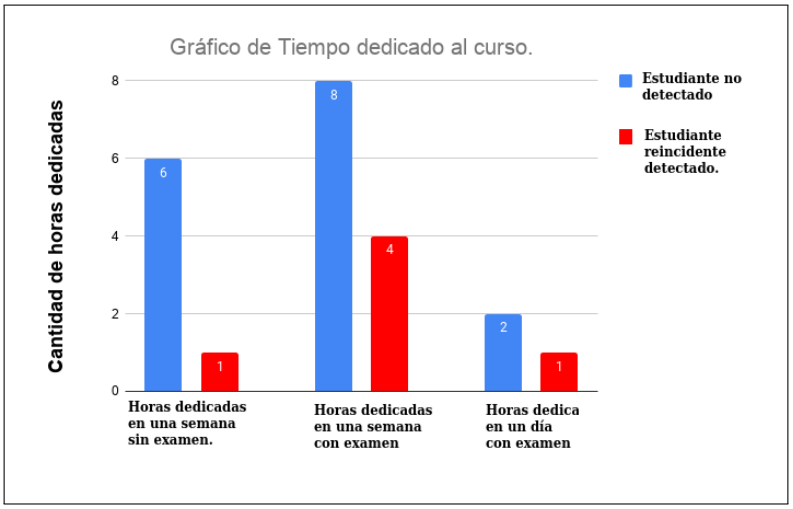


Figura 5.16: Cantidad de estudiantes que reinciden en formar grupos de trabajo y son detectados por *DetectApp*

En este caso de estudio se toma al azar 19 estudiantes que no son detectados durante todos los exámenes y 19 que si lo son, y se analiza las interacciones que estos han realizado. Dichas interacciones son horas de reproducción de contenido multimedia, participación en foros, lecturas del material compartido por el docente y horas de dedicación diarias al curso. Los resultados que se obtiene se muestran en la figura 5.15. Estos son el promedio de horas dedicadas por los estudiantes que no son detectados comparados con el promedio de los que si lo son. Los datos concuerdan con lo que se expresa en [33] y [31]. Según el artículo un estudiante que presenta comportamiento con sospecha de fraude no participa activamente en los foros como se muestra en la gráfica 5.17.

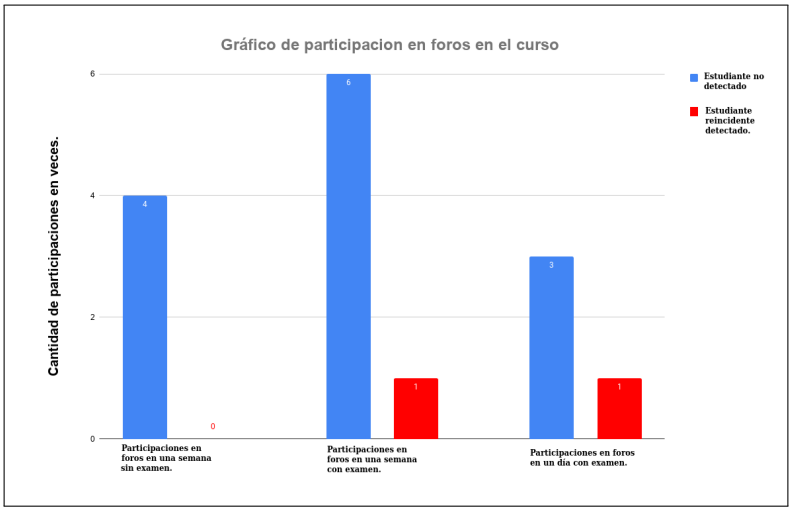


Figura 5.17: Gráfico de participación en foros en el curso

Además de no consumir activamente el contenido multimedia del curso, como vídeo tutoriales, vídeo clases y lecturas recomendadas, como se muestra en la figura 5.18.

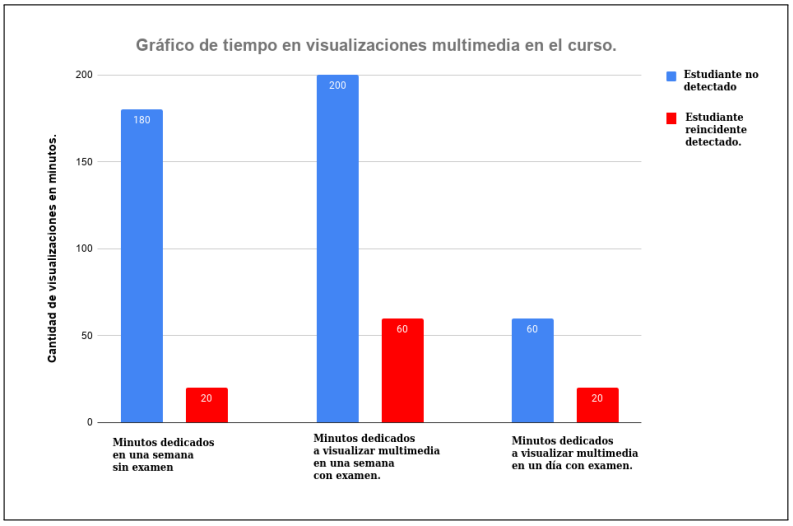


Figura 5.18: Cantidad de tiempo dedicado a visualizaciones de un estudiante no detectado (azul) VS uno reincidente (rojo) en formar grupos de trabajo y son detectados por *DetectApp*

Según este comportamiento un aporte que se puede hacer a la estrategia evaluativa es considerar la posibilidad de dar reconocimientos académicos a los estudiantes

que participen activamente en el desarrollo total del curso, creando la necesidad del habito de participación activa y con esta probablemente disminuyendo los posibles comportamientos fraudulentos.

Capítulo 6

Conclusiones

Los programas de educación superior han visto como la capacidad operacional en el sistema educativo tradicional se ha quedado corta frente a la necesidad de la misma que se presenta en la actualidad y una alternativa para mitigarla han sido los *MOOC*. Este tipo de cursos permiten llegar a mas estudiantes en comparación al método tradicional, lo que ha llevado a que estos cursos sean cada vez más incorporados por las instituciones de educación superior.

Dicha incorporación al ámbito educativo tradicional requiere que los *MOOC* sean validos por créditos académicos, lo cual incita a que en los mismos se presenten conductas con sospecha de fraude. Esto hace necesario el estudio y diseño de herramientas o métodos para detectar y/o mitigar estos comportamientos.

Actualmente los métodos para la detección, identificación y clasificación de conductas con sospechas de fraude presentan restricciones a la naturaleza de los cursos masivos en linea, si bien funcionan correctamente en la mayoría de casos, estas restricciones hacen que el problema inicial vuelva a presentarse y la ventaja de la masividad se vea limitada.

En el presente trabajo se realizó una revisión sistemática de la literatura existente para identificar las practicas de evaluación que se relacionan con las conductas con sospecha de fraude. Basados en éstas, se diseño una estrategia evaluativa que acompañada de un mecanismo tecnológico, permita la identificación de dichas conductas.

Para verificar su efectividad se aplicó a un caso de estudio en la Universidad del Cauca.

Al llevar a cabo lo mencionado anteriormente se obtienen las siguientes conclusiones:

- El método de evaluación mas común usado en los cursos en linea es el cuestionario de opción múltiple con única respuesta.
- Dado que el caso de estudio se realizó con estudiantes de una misma universidad, la cercanía y la posible relación entre estudiantes facilita las conductas con sospecha de fraude.
- El método de evaluación de cuestionario con preguntas de opción múltiple facilita conductas con sospecha de fraude como el compartir respuestas.
- Los estudiantes forman grupos de trabajo como conducta con sospecha de fraude predominante.
- El implementar una estrategia evaluativa en los métodos de evaluación aumenta la capacidad de detectar y confirmar algunos comportamientos con sospecha de fraude.
- Alternar diferentes versiones de un cuestionario, manteniendo el orden de las preguntas pero aleatorizando el orden de las opciones de respuesta, permite una mejor identificación de estudiantes con sospecha de fraude sin modificar la naturaleza del curso.
- El desarrollar un mecanismo para la implementación de la estrategia evaluativa brinda mas herramientas para la identificación de conductas con sospecha de fraude a diferencia de solo implementar una estrategia evaluativa.
- El mecanismo de detección *DetectApp* identificó dos conductas con sospecha de fraude que se mantuvieron a lo largo del desarrollo del caso de estudio. Tales son formar grupos de trabajo y copiar sus respuestas.
- *DetectApp* encontró que si los estudiantes forman grupos de trabajo y se ubican en la misma locación la efectividad de la estrategia evaluativa disminuye.

- *DetectApp* como mecanismo de detección, brinda las herramientas necesarias para detectar conductas con sospecha de fraude cuando la estrategia evaluativa disminuye su efectividad.
- Aunque la cantidad de estudiantes detectados que cumplen los tres criterios disminuyó a lo largo del caso de estudio, no es garantía de disminución de conductas con sospecha de fraude.
- A pesar de la implementación de la estrategia evaluativa, existe un grupo de estudiantes que se mantiene constante en cantidad, pero variando sus conductas con sospecha de fraude.
- Los estudiantes que han sido detectados con sospecha de fraude presentan hábitos de entrega similares entre ellos, como tiempos de entrega, cantidad de preguntas respondidas, aciertos y errores.
- La cantidad de aciertos en un examen sirve como criterio de identificación para la detección de conductas con sospecha de fraude.
- Aunque la herramienta no lo permite ver, al analizar el comportamiento a lo largo del caso de estudio de estudiantes detectados como sospechosos de conductas de fraude y comparado con estudiantes que no han sido detectados, permite concluir que la actividad de los estudiantes con sospecha de fraude en la plataforma es menor que la actividad de los estudiantes regulares.

En términos generales, al usar el método de evaluación en forma de cuestionario de opción múltiple con única respuesta y modificarlo para generar una estrategia evaluativa que no altere la naturaleza del curso *MOOC* “Introducción al emprendimiento con Lean StartUpƨ apoyarlo con el desarrollo e implementación de un mecanismo web, permite identificar conductas con sospecha de fraude que en condiciones normales se hace complejo dado la masividad del curso.

Si bien la estrategia evaluativa puede ser identificada y sorteada por los alumnos cuando estos se encuentran en el mismo lugar, la herramienta *DetectApp* permite seguir identificándolos por medio del uso de los parámetros definidos en la sección 4.2, Aunque el alumno note que el orden de las respuestas ha sido modificado, la

herramienta revisa si el alumno esta formando grupos de trabajo y de ser así lo clasifica como sospechoso de fraude cumpliendo así con el objetivo de identificación de conductas con sospecha de fraude.

Una manera de soportar esta clasificación cuando se presenta la situación antes mencionada, es revisando la cantidad de aciertos de las personas en el grupo de trabajo, si la coincidencia es significativa es una confirmación a la conducta con sospecha de fraude, debido a la masividad del curso hacer esta revisión de una manera manual se convierte en algo carente de usabilidad pero dicho proceso se puede automatizar lo cual se propone como mejora para trabajos futuros.

Bibliografía

[1] E.Francesc, "Bolonia y las TIC:de la docencia 1.0 al aprendizaje 2.0," Tech. Rep., Enero 2009.

[2] R. Moe, "The brief expansive history (and future) of the MOOC: Why two divergent models share the same name," 2015.

[3] J. Gomez Galan, "El fenómeno MOOC y la universalidad de la cultura: las nuevas fronteras de la educación superior," Tech. Rep., 2014.

[4] J. E. Borrero A., "Revista Educacion Virtual, ¿Pueden los cursos en linea evitar que los estudiantes hagan trampa?" Tech. Rep., Febrero 2013. [Online]. Available: https://revistaeducacionvirtual.com/archives/511.

[5] H. Corrigan Gibbs, E. Cutrell, N. Gupta, C. Northcutt, "Measuring and Maximizing the Effectiveness of Honor Codes in Online Courses," Tech. Rep., 2015.

[6] D. McCabe, L. K. Trevino, "Academic Dishonesty," Tech. Rep., 1993.

[7] A. D Ho, C. G Northcutt, I. L Chuang, "Detecting and Preventing "Multiple-AccountÇheating in Massive Open Online Courses," Tech. Rep., 2015.

[8] J. A. Ruiperez Valiente, A. Giora, C. ZhongZhou, D. Pritchard, "Using multiple accounts for harvesting solutions in mooc," *Proceedings of the Third, ACM Conference on Learning*, 2016.

[9] A. Bozkurt, I. De Waard, N. Ozdamar Keskin, "Research Trends in Massive Open Online Course (MOOC) Theses and Dissertations: Surfing the Tsunami Wave," Tech. Rep., Agosto 2016.

[10] R.Florido Bacallao, M.Florido Bacallao, "La Educación a Distancia, sus retos y posibilidades," Tech. Rep., Julio 2003.

[11] Universia Colombia, "10 ventajas de la educación virtual," Tech. Rep., Febrero 2020. [Online]. Available: https://noticias.universia.net.co/ciencia-tecnologia/noticia/2020/02/09/1167632/10-ventajas-educacion-virtual.html

[12] E. Dorrego, "Educación a distancia y evaluación del aprendizaje," Tech. Rep., 2016.

[13] H. Abelson, "The creation of OpenCourseWare at MIT," Tech. Rep., April 2008.

[14] P. Ruiz Martin, "Presente y futuro de los massive open online course (mooc)," *Universidad Complutense de Madrid*, 2013.

[15] J. Sanchez I., P. Castillo S., E. Cifuentes R., "WPD1.3 Informe sobre administración, gestión y planificación," Tech. Rep., 2016.

[16] R. Moe, "The brief expansive history (and future) of the MOOC: Why two divergent models share the same name," Tech. Rep., January 2015.

[17] A. Gonzales Aguilar, R. Poy Castro, "Factores de exito de los MOOC: algunas consideraciones criticas," Tech. Rep., Marzo 2014.

[18] A. Fidalgo Blanco, F. J. Garcia Peñalvo, M. L. Sein Echaluce, "Los MOOC: un análisis desde una perspectiva de la innovación institucional universitaria," Tech. Rep., 2017.

[19] Y. R. Jeffrey, "The Chronicle for Higher Education," Tech. Rep., February 2013.

[20] C. Delgado Kloos, P. J. Muñoz Merino, M. Muñoz Organero, C. Alario Hoyos, M. Perez San-Agustin, H. A. Parada G, J. A. Ruiperez, J. L. Sanz, "Experiences of running moocs ans spocs at uc3m," *EDUCON, vol. I*, pp. 884–891, 2014.

[21] A. Eisenberg, "The New York TIMES," Tech. Rep., March 2013. [Online]. Available: https://immagic.com/eLibrary/ARCHIVES/GENERAL/GENPRESS/N130302E.pdf

[22] H. Corrigan Gibbs, E. Cutrell, N. Gupta, C. Northcutt, "Deterring Cheating in Online Environments," Tech. Rep., 2015.

[23] S. Sequeria, E. Lopes, "Simple Method Proposal for cost estimation for work Breakdown structure," Tech. Rep., 2015.

[24] S. Lujan Mora, "MOOC (Massive Open Online Course)," Tech. Rep., 2013.

[25] A. Fidalgo, F. J. Garcia Peñalvo, M. L. Sein Echaluce, "MOOC cooperativo. Una integración entre cMOOC y xMOOC," Tech. Rep., 2013.

[26] G. Siemens, "Connectivism: creating a learning ecology in distributed environments," Tech. Rep., 2007.

[27] O. Popovic, R. Prokic, S. Trajkovic, "Massive Open Online Courses (MOOC) and Its Possibilities as Instrument of Formal, Nonformal, Informal and Lifelong Learning," Tech. Rep., 2016.

[28] P. Dawson, "The Conversation," Tech. Rep., November 2014. [Online]. Available: http://theconversation.com/explainer-what-is-a-small-private-online-course-34542.

[29] Anonimo, "Desarrollo Docente," Tech. Rep., July 2015. [Online]. Available: http://desarrollodocente.uc.cl/images/pdf/metodologias/Clase%20invertida.pdf.

[30] P. Dawson, R. Nelson, "The Conversation," Tech. Rep., October 2012. [Online]. Available: https://theconversation.com/moocs-and-exercise-bikes-more-in-common-than-youd-think-9726.

[31] D. Jaramillo Morillo, M. Solarte, G. Ramirez Gonzales, "Estrategia de seguimiento a las actividades de aprendizaje de los estudiantes en cursos en linea masivos y privados (mpoc) con reconocimiento académico en la universidad del cauca," *Séptima Conferencia de Directores de Tecnología de Información TICAL Gestión de las TICs para la Investigación y la Colaboración, San José*, 2017.

[32] W. Guo, "From SPOC to MPOC-The effective practice of Peking University online teacher training," Tech. Rep., 2014.

[33] D. Jaramillo Morillo, M. Solarte, G. Ramirez Gonzales, "Características del comportamiento de los estudiantes en un mpoc con reconocimiento académico sobre una instancia open edx," *Proceedings of the international conference mooc maker, Antigua Guatemala*, 2017.

[34] C. Alario Hoyos, G. Ramirez Gonzales, R. Ramirez Velarde, M. Solarte, "Kolb's learning styles, learning activities and academic performance in a massive private online course," *Advances in Soft Computing, I. Batyrshin, M. d. L. Martinez Villaseñor y H. E. Ponce Espinoza, Guadalajara*, pp. 327–341, 2018.

[35] Scopeo, "SCOPEO INFORME N2 MOOC: Estado de la situacion actual, posibilidades, retos y futuro," Tech. Rep., 2013.

[36] M. Padilla, M. A. Ramirez Fernandez, "Los MOOC en la Educación Superior. Un análisis comparativo de plataformas," Tech. Rep., 2016.

[37] I. Aguaded, R. Medina Salguero, "Los MOOC en la plataforma educativa MiriadaX," Tech. Rep., 2014.

[38] EduNext, "www.EduNext.com," Tech. Rep., Febrero 2018. [Online]. Available: https://www.edunext.co/es/open-edx/.

[39] J. E. S. P. V. J. B. F. M. E.Linares Espinós, V.Hernández, "Methodology of a systematic review," *Elsevier España*, 2018.

[40] K. m. Chang, X. Li, A. Hauptmann, "Massive Open Online Proctor: Protecting the Credibility of MOOC Certificates," Tech. Rep., 2015.

[41] T. Gill, "Proctorfree: Deterring online cheating," Tech. Rep., 2013.

[42] A. Chiappe, N. Hine, J. Martinez, "Literatura y practica: una revisión critica acerca de los MOOC," Tech. Rep., 2015.

[43] D. Wiley, "The MOOC Misstep and the Open Education Infrastructure," Tech. Rep., September 2014. [Online]. Available: https://opencontent.org/blog/archives/3557.

[44] P. Muñoz Merino, P. Diaz Pijeira, C. Delgado Kloos, J. A. Ruiperez Valiente, S. Ruiz, "Evaluation of a learning analytics application for open edx platform," *Computer Science and Information Systems, vol. 14*, p. 43, 2017.

[45] M. Solarte, G. A. Ramirez, D. Jaramillo, "Hábitos de ingreso y resultados en las evaluaciones en cursos en linea masivos con reconocimiento académico," *RIIN, vol. 5*, Mayo 2017.

[46] A. Melendez, M. Roman, M. Perez Sanagustin, J. Maldonado, "Calidad en Cursos Abiertos Masivos y en Linea. Revisión de literatura del 2012-2016," Tech. Rep., 2017.

[47] B. Perez, S. Reyna, "Estrategias Metodologicas y Evaluativas en el Ejercicio de la Docencia de los Profesores de Enseñanza Media en Pedagogia," Tech. Rep., 2013.

[48] J. Murillo Pacheco, "Principios y propiedades de la evaluacion," Tech. Rep., 2017.

[49] J. Matias Pereda, G. Lannelongue Nieto, "Tecnicas de ayuda en el proceso de aprendizaje: El caso de los Sistemas Anticopia," Tech. Rep., 2013.

[50] A. Berenice, R. Rodriguez, B. Perez, "Aplicacion del Modelo ELQ en la evaluacion de la calidad en la educacion a distancia impartida en la UPPUEBLA," Tech. Rep., 2013.

[51] K. V. G. D. M. P. D. K. C. Ruiperez Valiente Jose A., Joksimovic Srecko, "A Data-driven Method for the Detection of Close Submitters in Online Learning Environments," 2017.

[52] M. F. S. G. R. G. Daniel Jaramillo Morillo, José Ruipérez Valiente, "Identifying and characterizing students suspected of academic dishonesty in SPOCs for credit through learning analytics," 2020.

[53] Y. E.-A. Adi Friedman, Ina Blau, "Cheating and Feeling Honest: Committing and Punishing Analog versus Digital Academic Dishonesty Behaviors in Higher Education," 2016.

[54] L. M. L. A. Ferre Jáen, Elvira Del Río Alonso, "ExamRandomizeR: Una aplicación web para la generación de exámenes aleatorizados que faciliten situar el Examen como herramienta de aprendizaje y no solo de evaluación en el aula de matemáticas," 2017.

[55] J. J. E. B. Enrique Sánchez Acosta1, "ALEATORIEDAD EN LA EVALUACIÓN DE LOS MOOC," 2014.

Anexos

Anexo A

Manual de instalación de la aplicación web DetectApp

Trabajo de investigación de Pregrado
Yehison Javier Cuchumbe Pencua
Paula Andrea Vasquez Artunduaga
Director: Mag. Daniel Alberto Jaramillo Morillo

En este manual se encuentra la descripción del proceso de instalación de la herramienta DetectApp en un servidor.

Este proceso de divide en los siguientes pasos:

- Descarga:
  Para la descarga del codigo fuente se puede hacer uso del enlace de descarga https://github.com/javierc1993/DetectApp.git o bien hacer uso del CLI de github mediante el comando gh repo clone javierc1993/DetectApp.
- Instalación de dependencias:

  Para la instalación de las dependencias se debe previamente tener instalado Node JS y su gestor de paquetes npm.

Con el comando npm install “nombredeladependencia” se puede instalar cada una de las siguientes dependencias:

- Morgan.
- Express.
- Mongoose.
- EJS.
- Bootstrap.
- nodemon.

- Variables de entorno.

  Se debe crear un archivo .env donde se contengan las variables de acceso a la base de datos, como la dirección del dominio en donde se va a alojar la aplicación.

- Ejecución:

  Entrar al directorio raiz del proyecto. Correr npm install para descargar las dependencias especificadas en el package.json.
  Ejecutar node src/apps.js si es para entorno de desarrollo.
  Ejecutar node src/prod/apps.js si es para entorno de despliegue o bien usar el comando npm start que ejecuta el script de producción o npm dev que ejecuta el script de desarrollo.

Anexo B

Manual de Usuario de la aplicación DetectApp.

Trabajo de investigación de Pregrado
Yehison Javier Cuchumbe Pencua
Paula Andrea Vasquez Artunduaga
Director: Mag. Daniel Alberto Jaramillo Morillo

B.1. Inicio de sesión.

Para iniciar sesión debes ingresar a la pagina de bienvenida de la aplicación DetectApp, en esta encontrarás la siguiente interfaz:

Figura B.1: Interfaz Inicio de Sesión de la aplicación *DetectApp*

En esta se encuentra un formulario con los campos **Usuario** y **Contraseña**, estos son unicos e intransferibles y son asignados por el administrador de la aplicación unicamente a el docente o docentes encargados del curso.

Luego de diligenciar de manera correcta el formulario, se debe pulsar el botón identificado con la etiqueta “Ingresa”, al hacer esto y siempre y cuando la información ingresada sea correcta, DetectApp mostrara el listado de cursos disponibles.

B.2. Seleccionar curso.

Una vez se haya hecho de manera correcta el inicio de sesión, DetectApp mostrara la siguiente interfaz con el listado de cursos disponibles:

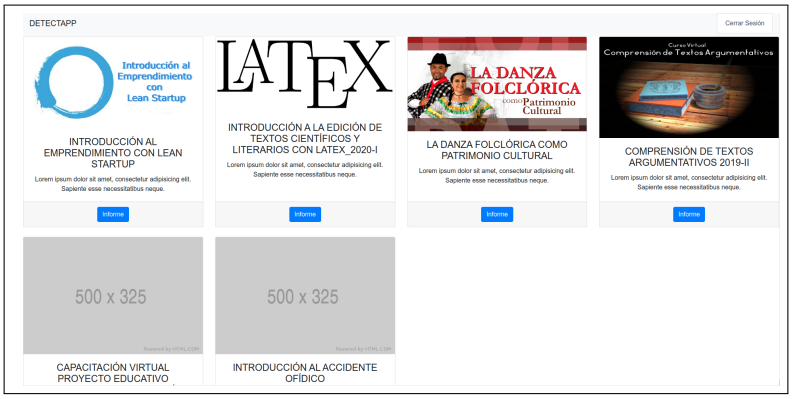


Figura B.2: Interfaz Listado de cursos disponibles en *DetectApp*

Se debe posicionar sobre la tarjeta del curso que se quiere visualizar los informes y dar clic sobre el botón con la etiqueta “informe” al hacerlo se nos desplegara la interfaz con el listado de exámenes.

B.3. Ver listado de exámenes.

Cuando se haya seleccionado el curso a analizar, DetectApp nos muestra el listado de exámenes que se encuentran disponibles al momento de hacer la consulta, este listado es como el que se muestra en la figura siguiente:

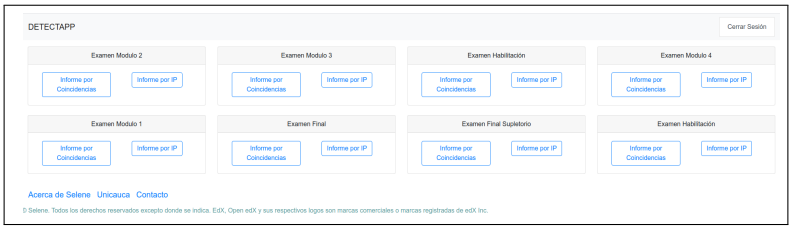


Figura B.3: Interfaz Listado de exámenes en *DetectApp*

cada examen nos muestra dos botones que nos conducirán a las dos alternativas de informe, para ver el informe por grupos de trabajo pulsamos el que tiene la etiqueta “Informe por IP” y para ver el informe por porcentajes de coincidencias pulsamos el que tiene la etiqueta “Informe por coincidencias”.

B.4. Consultar informe por porcentaje de coincidencias.

En este apartado se encuentra el listado de estudiantes que presentan un 90 % o mas de coincidencias en sus respuestas, para una mejor abstracción y análisis de información este a sido clasificado en cuatro posibles informes:

B.4.1. Listar estudiantes por porcentaje de coincidencias.

Esta opción es la que se muestra por defecto al seleccionar el informe por coincidencias, la vista se muestra como la figura siguiente:

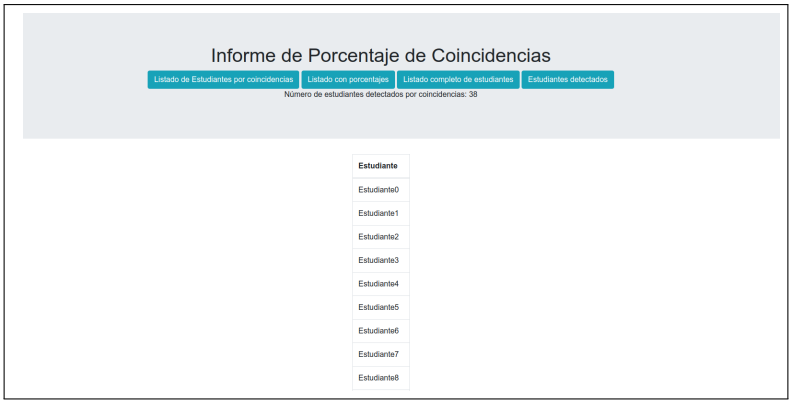


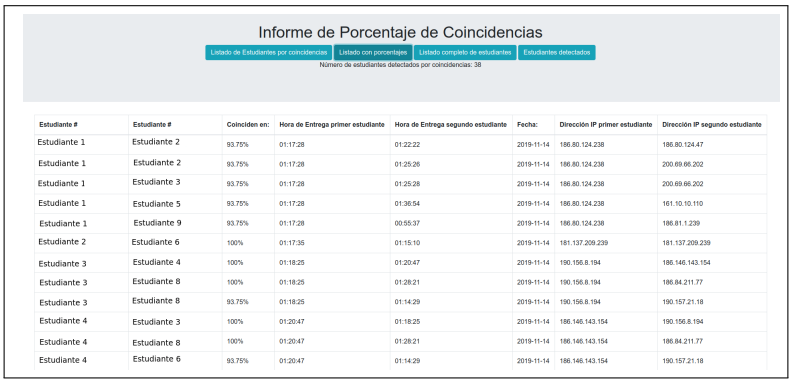
Figura B.4: Interfaz Listado de estudiantes con coincidencias superior al 90 % en *DetectApp*

en ella solamente se muestra los nombres de los estudiantes que tienen un porcentajes de coincidencias en sus respuestas igual o superior a un 90 % e igual o menor a 100 %.

B.4.2. Listar estudiantes detectados con sus porcentajes de coincidencia.

Al pulsar el botón con la etiqueta “Listado con porcentajes” se muestra una lista organizada en pares con los nombres de los estudiantes que tengan mas del 90 % de

coincidencias e información importante como el porcentaje total de coincidencias, el alumno con el cual coinciden en las respuestas, la hora de entrega de cada alumno, su fecha y la dirección IP de cada estudiante.

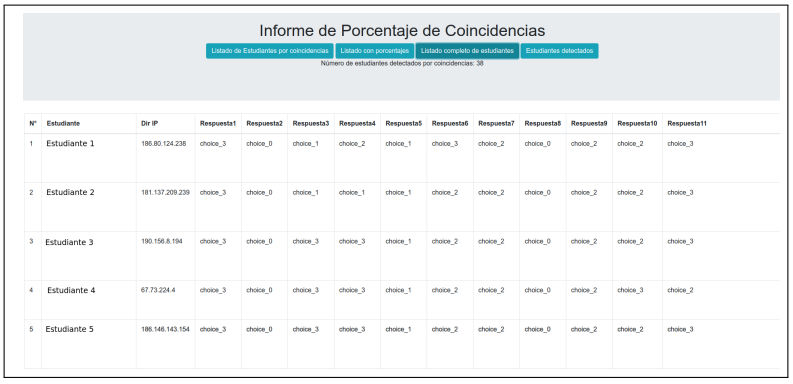


Estudiante #	Estudiante #	Coinciden en:	Hora de Entrega primer estudiante	Hora de Entrega segundo estudiante	Fecha:	Dirección IP primer estudiante	Dirección IP segundo estudiante
Estudiante 1	Estudiante 2	93.75%	01:17:28	01:22:22	2019-11-14	186.80.124.238	186.80.124.47
Estudiante 1	Estudiante 2	93.75%	01:17:28	01:25:26	2019-11-14	186.80.124.238	200.69.66.202
Estudiante 1	Estudiante 3	93.75%	01:17:28	01:25:28	2019-11-14	186.80.124.238	200.69.66.202
Estudiante 1	Estudiante 5	93.75%	01:17:28	01:36:54	2019-11-14	186.80.124.238	161.10.10.110
Estudiante 1	Estudiante 9	93.75%	01:17:28	00:55:37	2019-11-14	186.80.124.238	186.81.1.239
Estudiante 2	Estudiante 6	100%	01:17:35	01:15:10	2019-11-14	181.137.209.239	181.137.209.239
Estudiante 3	Estudiante 4	100%	01:18:25	01:20:47	2019-11-14	190.156.8.194	186.146.143.154
Estudiante 3	Estudiante 8	100%	01:18:25	01:28:21	2019-11-14	190.156.8.194	186.84.211.77
Estudiante 3	Estudiante 8	93.75%	01:18:25	01:14:29	2019-11-14	190.156.8.194	190.157.21.18
Estudiante 4	Estudiante 3	100%	01:20:47	01:18:25	2019-11-14	186.146.143.154	190.156.8.194
Estudiante 4	Estudiante 8	100%	01:20:47	01:28:21	2019-11-14	186.146.143.154	186.84.211.77
Estudiante 4	Estudiante 6	93.75%	01:20:47	01:14:29	2019-11-14	186.146.143.154	190.157.21.18

Figura B.5: Interfaz Listado con porcentajes *DetectApp*

B.4.3. Ver el listado completo de estudiantes.

Esta opción permite visualizar el listado completo de los estudiantes que han presentado el examen a consultar.



N°	Estudiante	Dir IP	Respuesta1	Respuesta2	Respuesta3	Respuesta4	Respuesta5	Respuesta6	Respuesta7	Respuesta8	Respuesta9	Respuesta10	Respuesta11
1	Estudiante 1	186.80.124.238	choice_3	choice_0	choice_1	choice_2	choice_1	choice_3	choice_2	choice_0	choice_2	choice_2	choice_3
2	Estudiante 2	181.137.209.239	choice_3	choice_0	choice_1	choice_1	choice_1	choice_2	choice_2	choice_0	choice_2	choice_2	choice_3
3	Estudiante 3	190.156.8.194	choice_3	choice_0	choice_3	choice_3	choice_1	choice_2	choice_2	choice_0	choice_2	choice_2	choice_3
4	Estudiante 4	67.73.224.4	choice_3	choice_0	choice_3	choice_3	choice_1	choice_2	choice_2	choice_0	choice_2	choice_3	choice_2
5	Estudiante 5	186.146.143.154	choice_3	choice_0	choice_3	choice_3	choice_1	choice_2	choice_2	choice_0	choice_2	choice_2	choice_3

Figura B.6: Interfaz Listado completo en *DetectApp*

En ella se muestra el nombre del estudiantes, su dirección IP , las respuestas escogidas, el curso, el código único de sesión, la fecha de presentación, la sección y subseción del curso, la hora y por ultimo la página del cuestionario.

Para visualizarlo se debe pulsar el botón identificado con la etiqueta “Listado completo de estudiantes”.

B.4.4. Ver el listado de estudiantes detectados que coinciden con los tres parámetros de detección.

Esta opción permite visualizar los estudiantes que además de coincidir en su 90 % de respuestas, también han formado grupos de trabajo y tienen horas de entrega cercanas.

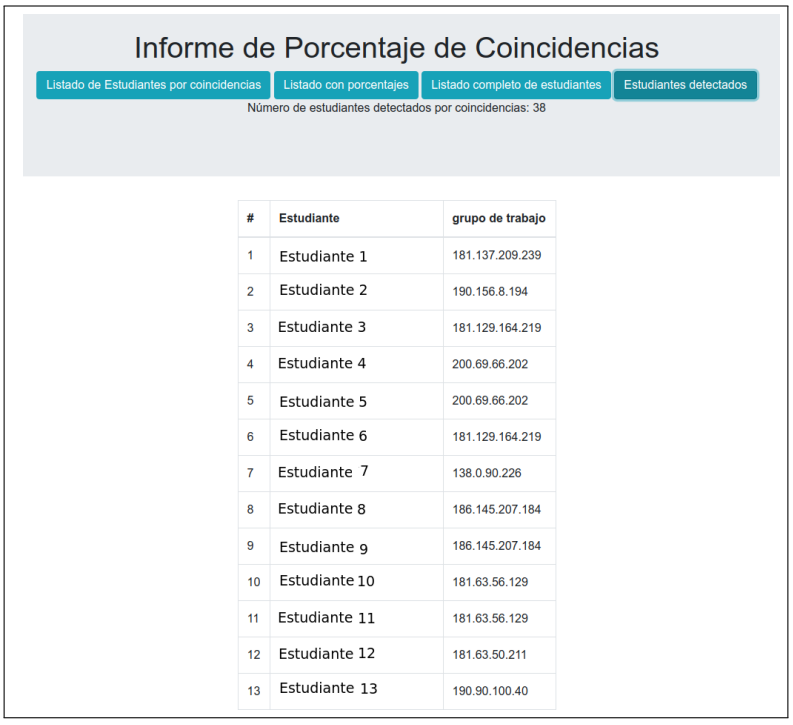


Figura B.7: Interfaz Estudiantes detectados por *DetectApp*

Para visualizarlo se debe pulsar el botón identificado con la etiqueta “Estudiantes detectados”.

B.5. Consultar informe por grupos de trabajo.

Esta opción permite que el docente pueda visualizar la cantidad de grupos de trabajo detectados y un listado de las direcciones IP de cada grupo.

Figura B.8: Interfaz Informe por IP *DetectApp*

Para visualizarlo se debe pulsar el botón identificado con la etiqueta “Informe por IP”.

B.5.1. Listar estudiantes detectados en el grupo de trabajo.

Es la opción que se muestra por defecto al pulsar el botón identificado con la etiqueta “Ver grupo de trabajo”.

Figura B.9: Interfaz Grupos de trabajo en *DetectApp*

En ella se encuentra un listado con los estudiantes que tienen la misma dirección IP y tiempos de envío cercanos, para hacer mas concluyente la información se muestra datos como el porcentaje de coincidencia, la hora de entrega de cada estudiante, la fecha de entrega y la dirección IP de cada uno.

B.5.2. Ver el listado de estudiantes detectados que coinciden con los tres parámetros de detección.

Esta opción permite visualizar los estudiantes que además de formar grupos de trabajo, coinciden de un 90 % a 100 % de respuestas y tienen horas de entrega cercanas.

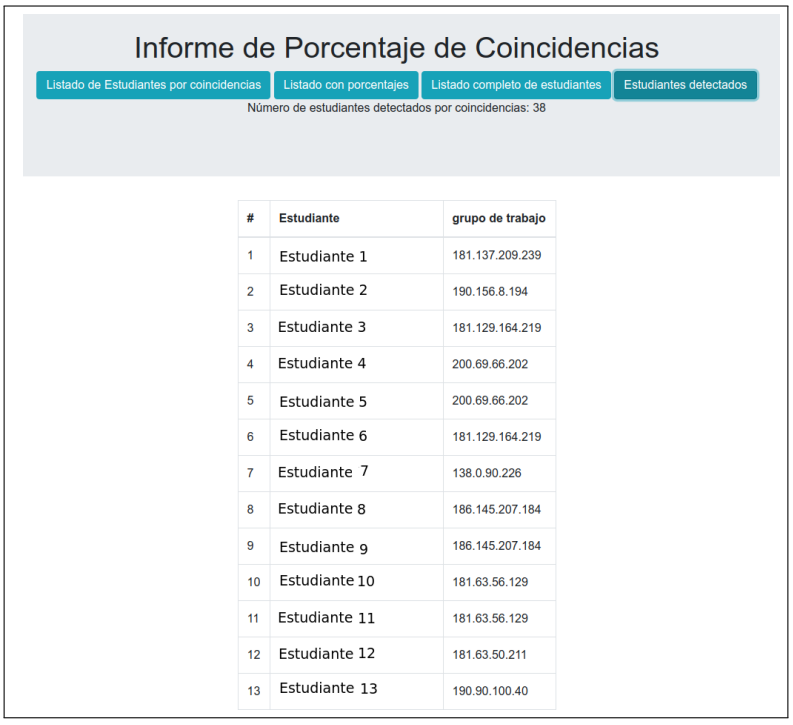


Figura B.10: Interfaz Estudiantes detectados *DetectApp*

Para visualizarlo se debe pulsar el botón identificado con la etiqueta “Estudiantes detectados”.

Anexo C

Artículo

Trabajo de investigación de Pregrado
Yehison Javier Cuchumbe Pencua
Paula Andrea Vasquez Artunduaga
Director: Mag. Daniel Alberto Jaramillo Morillo

Students sharing answers: Tool for detecting dishonest practices in Massive Private Online Courses

Abstract— **Considering the great popularity of Massive Open Online Courses (MOOCs), several higher education institutions have chosen to incorporate them into their professional programs and give them academic recognition. However, this has caused some to question the honesty of the students in these types of courses. There are still many challenges to overcome. This article presents DetectApp, a tool for the identification of students that are suspected of fraud through a report of several features, such as the similarity in responses or IP addresses. In addition, we present a use case applying the tool on a MPOC (Massive Private Online Course) offered by the University of Cauca during the second semester of 2019. Professors teaching MPOC courses on-campus and for credit can use DetectApp as a support tool to facilitate the detection of students that are likely committing fraud.**

***Keywords— MOOC, MPOC, Cheating, Academic Dishonesty, Assessment Design, Learning Analytics*.**

I. INTRODUCTION

Nowadays, there is a growing evolution of communication technologies that are facilitating the rise of online education [1]. For this reason, new educational modalities have been conceived, one of them has been the Massive Open Online Courses (MOOCs); which are online courses that have as main characteristics that they are massive and open [2], and inherit the advantages of traditional e-learning. Thus, they provide the opportunity to present diverse and flexible types of content and allow access to education at any time and place [3]. These advantages have led many universities and traditional education institutions to be interested in these new educational modalities as part of their teaching portfolio [4].

The MOOC model is gradually being incorporated into universities and their professional programs through strategies such as SPOCs and MPOCs (Small and Massive Private Online Courses), which are a MOOCs' variant that has been successfully adapted to not open educational settings. These modalities have the advantage of having a more controlled educational environment, which has led educational institutions to implement these courses as part of programs that provide academic credit or professional certifications [4].

With the arrival of these courses as an alternative way to obtain academic recognition, new requirements emerge. Therefore, there is a need to control for aspects such as the emergence of academically dishonest practices, to which these courses are vulnerable because they are taught completely online with online evaluations that have no proctoring control [4]. Behaviors such as impersonation, creation of multiple accounts, use of materials or sites that are not permitted, copying or cheating during exams, can be combated by implementing measures and proposing a solutions that would mitigate such behaviors and take full advantage of the important benefits offered by this form of education [5], [6], [7], [8].

In this article, we present the results after the implementation of DetectApp, a tool that allows detecting dishonest behaviors in Open edX course evaluations. This application allows the identification of behaviors with suspicion of fraud through Learning Analytics techniques and was tested in a MPOC in the context of the University of Cauca. The mechanism is based on an algorithm that compares the answers sent by the students and the IP address of their submissions, organizes the results in a database (mongodb), and presents the information to the teachers through a web application (nodejs).

There are high expectations about the impact of this proposal and it is expected that the results obtained will be very useful in this area.

II. RELATED WORK

A systematic review approach was adopted for the generation of the knowledge base. This review is defined as "a methodology that involves a critical and reproducible summary of the results of available publications on the same topic or specific issue". In this methodology, a search, detection, analysis and communication of information oriented to a certain topic in the field of science and technology is carried out" (E.Linares Espinós, 2018) [9]. In this systematic review, we have considered works related to fraud or academic dishonesty in online courses, as well as evaluative strategies that manage to identify, prevent and combat such behaviors. The most representative papers are presented below.

A. MOOC Fraud and Strategies for Reducing Academic Dishonesty in Online Learning Environments.

The MOOCs were conceived with the idea of being able to take the opportunity of education beyond the classrooms of a high school and to be able to turn the web into a great educational classroom that would allow access to education to any person regardless of their geographical location [10]. These advantages made universities attracted to this new educational model and decided to adopt it, trying to make these courses academically recognized in some of their programs [3], [11]. This recognition led to an increase in suspected fraud and the homologation of the MOOC contents for academic credits was more popular. Likewise, the

methods of committing fraud were more varied [4], in which the CAMEO technique [7] "Copying Answers using Multiple Existences Online" stands out.

Although Learning Analytics is used in [7] to detect CAMEO, this type of fraud technique is not implementable in MPOC environments. Due to its private nature, MPOC performs a login and logout control so that no student can have more than two accounts.

On the other hand, MOOCs, being an online environment, facilitate plagiarism and copying [4]. Here, techniques such as the misuse of web resources to search for answers, or having someone from the course send the correct answers to others before the exam is taken, come into play. As these fraudulent tendencies have increased, methods of prevention and correction have emerged. The honor code is a method brought from traditional teaching [6] and applied in online environments, but its effectiveness depends on the individual's ethical level. Therefore, an improvement has emerged, known as the "warning method" [5]. Which consists of displaying warnings prior to the presentation of the exams, so that the student is persuaded not to commit fraud, in order not to retaliate in the grade or stay in the course.

In addition, to deal with impersonation and plagiarism, surveillance systems have been put in place that use computer webcams to identify and monitor students taking an exam [12], [13]. Or the use of software elements such as "honeypots" [14] that consist of a website that contains the answers to the exam in order to identify the people who consult it at the time of the presentation of the test. However, these methods lead to a transformation of the learning philosophy of the MOOCs and a possible increase in the maintenance and sustainability of the platform, making the courses not very scalable, being very little applicable in MOOCs [15].

As stated in [4], [15], [16] personal supervision, validation by face-to-face examination, the use of third parties as guarantors [13], make its characteristics and the philosophy with which the concept of MOOC was born [17] be altered leaving it online or scalable.

In conclusion, none of the methods presented are totally effective and some present a limitation in the concept of massive and open, which is characteristic of the MOOC [15], so the gap is in finding effective strategies in the presentation of evidence that do not directly affect the concept of MOOC and that allow the characterization of fraudulent students and integration with the academy.

With regard to online courses, as defined in the ELQ (E-Learning Quality) model [18], assessment methods must implement strategies to deal with plagiarism, security and student authentication. In order to achieve this, they must institutionalize strategies through the creation of regulations, in such a way that the student can be guided on the correct way to consult and cite his or her sources.

III. METHODOLOGY

This section describes the methodology in two subsections. The first sub-section presents the specificities of the context and the case study, and the second the method that is applied and implemented for the identification of students suspected of academic dishonesty.

A. Description of the context and case study

The MPOC-type course "Introduction to Lean Startup Entrepreneurship", offered to students at the University of Cauca in Colombia, was used as a case study. This course was offered through the Selene learning platform, an instance of Open edX. The course emerged in the first period of 2018, so to date it has been offered on four opportunities, each with the participation of approximately 100 students. For this study, we have its last version, which was offered during the second semester of 2019.

The course was proposed to be developed as a FISH (Integral Social and Human Formation Component) elective and its fundamental purpose is to introduce students to the knowledge and use of one of the most successful metodologıas for the development of enterprises in recent times: the Lean Startup.

The course content was divided into 5 modules and an evaluation was made for each module. The evaluation was made taking into account each of the main topics, through online tests with multiple-choice questions that have only one correct answer. There were 12 to 14 questions with the possibility of taking the exam on two different dates.

For the presentation of each exam, the student had a time limit of 60 minutes to minimize the chances of fraud or shared answers. In addition, the course has an advantage, since access to the course is controlled by the platform administrator, this makes it impossible to create additional accounts and perform CAMEO.

However, other types of fraud methods can occur, such as sharing answers by other means or meeting in groups of students and solving tests. Therefore, we propose the creation of an evaluative strategy that will generate conditions for the identification of students with suspected fraud.

B. Identification of Students with Suspected Fraud

This section describes three parts: The method used for the identification of students with suspected fraud, the tool developed and the method used for the validation of the case study tool.

1) Criteria for student identification

Based on [19] where they work by comparing the time it takes for students to submit their answers, we have used two criteria to identify students who are suspected of fraud: students who have very similar test answers and match the IP address from which the test is sent.

For the identification of students by their answers on the tests, we use a similarity matrix. The matrix is made up of the calculation of matches by comparing each of a student's answers with the rest. Thus obtaining the DS matrix.

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

Where each entry $ds_{i,j}$ is the percentage of matching responses. Each element of the matrix is calculated by an exact match. For example, if one student matches another in all responses it will have a value of 100, otherwise a value of 0.

On the other hand, for the criterion of the IP address, all the students who send their exams from the same IP address have been grouped together. This case has been validated only for those addresses that are outside the university, thus being students who join to take the exams from their homes.

2) Tool for the identification of students with suspected fraud

The following is a general description of the architecture that was built using the 4+1 view model proposed by Philippe Kruchten [20]. This model allows the architecture to be represented in a standard way using UML diagrams. Figure 1 shows the deployment diagram of the built tool.

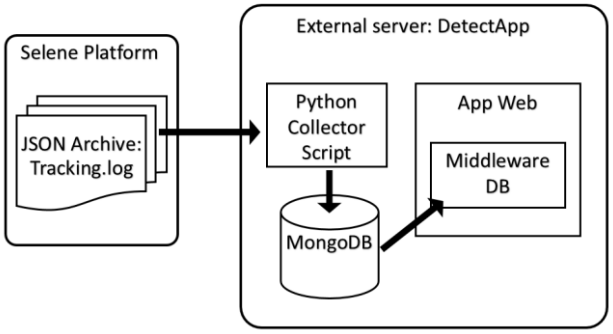


Fig. 1. System Deployment Diagram

- **Selene Platform:** This is an instance of Open edX that allows the offering of MPOCs for the Universidad del Cauca and is installed in a physical server of the university, the platform began its operation in the first period of 2016.
- **JSON Archive:** Each activity that students perform through the platform is recorded in a JSON format file called tracking.log. It is from this file that both the students' answers and the IP addresses for sending the exams are obtained. In addition to obtaining other information that helps to understand student behavior, such as: content viewed, video interactions, forum postings, etc., everything related to student interaction with the learning platform.
- **External server:** On the external server is the tool for identifying students with suspected fraud.
- **Python Collector Script:** This script is responsible for pre-processing the data in the JSON Tracking.log file. It takes all the records that are generated from the students' interactions with the learning platform, extracts the information and stores it in a MongoDB database in a more structured way. The information it retrieves is the student's idea, the activity performed, the IP address, the date and time the activity was performed, the location within the course, and in the case of assessments, it retrieves the answers sent by the student.
- **Middleware DB:** It structures the record data (student interactions) left in the database and runs the algorithm for both response comparison and IP address comparison. For identification by response, it selects from the match matrix students with more than 90% matches and for identification by IP address it classifies all students with the same address.
- **App Web:** It allows the visualization of the data delivered by the Middleware DB. The application allows teachers to visualize the courses they are in charge of and in each course to obtain a report on the students that have been detected as suspects of fraud.

C. Application in the case study

As a test, and in order to detect as many students as possible who share their answers, an evaluation strategy was developed so that students would fall into the trap. The strategy was to generate 5 different versions of questionnaires from one (5 per module). In this way, students were presented with questionnaires with the same questions and the same order, but with a subtle change in wording that made the answers to the questions different.

Then through the Selene platform, these versions of questionnaires were randomly delivered to the case study group of students to increase the probability of detection of possible fraud. In this way, if the students shared their answers but were not aware of the change in wording, they would have different scores. As expected, we received some complaints from some students that they had answered in the same way as their classmates but that they did not have the same grade. Validating that our strategy worked.

Thus, if there are more than 90% matches among students, it is considered a suspicion of fraud. With the help of the tool the teacher can check that the students have different questionnaires and it gives him useful information such as time of delivery, date of delivery and IP addresses, which will allow him to confirm whether there is fraud or not.

On the other hand, although the assessment strategy has the weakness of making students aware of the change in the statement, the selection of students who had the same IP address at the time of taking the exam was made. This indicates that they did collaborative work when taking the exam, which is not allowed in the case study and is classified as fraud. Thus, the tool shows the teacher, for each exam, a list of IP addresses in which by selecting one of them he displays a list of students with their names, date of delivery, time of delivery and their answers. This is to make it easier for the teacher to decide whether the suspected fraud is true or not.

IV. RESULTS

This section presents the tool that was built for the detection of students with suspected fraud and also some results of the application on the case study.

A. Web Application : DetectApp

The web application allows the visualization of some data of the students detected as suspects of fraud. Figure 2 shows the courses that the teacher is in charge of. This is the first interface shown to teachers.

Fig. 2. *DetectApp* Application Courses Interface.

The figure shows the administrator's view of the tool and displays the six active courses for the second semester of 2019, where each course has its corresponding report. Figure 3 shows all the exams contained in the course selected by the teacher. With the possibility of observing the report of each one of them.

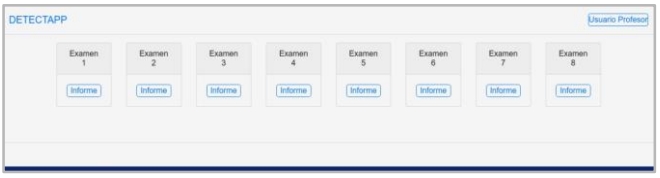


Fig. 3. *DetectApp* Application Modules Interface.

For each of the tests, two reports are shown, report by IP and report by answers match. Figure 4 shows the two types of report that the teacher can use as criteria to define the existence of fraud.

Fig. 4. *DetectApp* Application Reports Interface.

Figure 5 shows the report by IP. This interface shows the work groups identified by IP address matching in the test submission and in each of the groups the students with the same IP address can be queried.

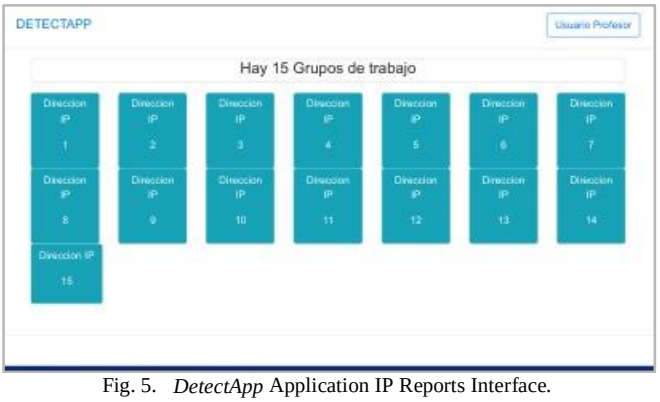


Fig. 5. *DetectApp* Application IP Reports Interface.

Figure 6 shows the report by response matches. In this interface the teacher can see the number of matches in the students' answers, as well as data such as time of delivery, date and IP address.

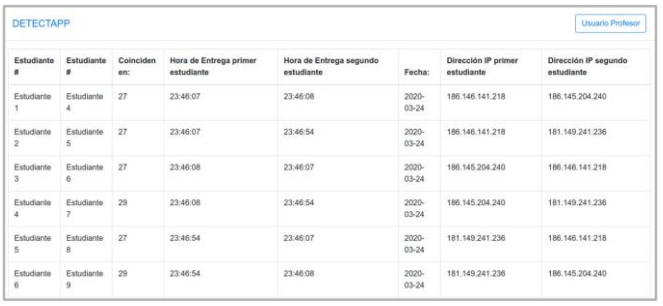


Fig. 6. *DetectApp* Application Matching Reports Interface.

B. Using the case study tool

After applying the method described in Section 3, item C, the following results were obtained. Table 1 shows the number of students who were identified according to the answer-matching report. This is one of the first indicators that was added to the tool, but it is expected to be integrated with the note achieved. It is possible that two students may answer the same questionnaire correctly, so the tool will identify them as suspects of fraud. Although it is recalled that 5 different questionnaires were created for each test and that these were given at random, so the probability of two students taking the same test is reduced. In addition, the results showed that 7 were the recurrent students in each of the tests. That is, the same 7 students were identified throughout the course and although they were not presented with the same questionnaires, they had the same answers and even the same errors.

	M1	**M2**	**M3**	**M4**	**Ex Final**
Matching Report	32	40	37	23	12
Repeat Students	NA	11	8	8	7
IP Report	15	12	12	11	11
Repeat Groups	NA	10	8	8	8

Table 1. Count of students identified in each report (answer-matching and IP address report) and count of recurring students.

We also identified that students worked in groups from outside the university and that 8 of the groups identified were maintained throughout the course. These students would meet to develop the exams from outside the university.

Figure 7 shows the consultation of one of the most suspicious groups. We found a group of 3 students who, according to the tool's criteria, presented suspicions of fraud. The Selene platform assigned a different questionnaire to each of them, and the DetectApp application found that all three coincided in the totality of the questionnaire responses, in the IP address and in the delivery times as they had a variation of a few seconds. The names of the students in the figure are anonymized.

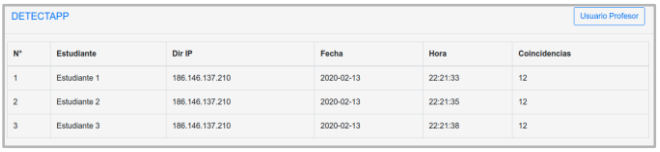


Fig. 7. Capturing and identifying suspicious students in the *DetectApp* application.

In addition, we noticed that the weakness of the evaluation strategy was evident when a group of students noticed the change in the wording of the questionnaires and shared their answers. However, we found that these students formed working groups with the help of the IP address report and manual review of the questionnaires. It was identified that although the coincidences in their answers were less than 90%, because they realized that the questionnaires were different and that the distracters in the exams are displayed randomly, they did commit fraud. Since they coincided in the number of incorrect answers in the same questions and made the same mistakes, and in addition to the fact that their answers had the same IP address, the dates and times of delivery were very close.

This confirms the recidivism of the groups despite the implementation of the evaluation strategy, which reaffirms the effectiveness of *DetectApp* to identify fraudulent behavior when evidencing this type of case.

V. CONCLUSIONS AND FUTURE WORK

In recent years, there has been great interest in incorporating the MOOC strategy into universities and allowing these types of courses to have a credit value and to be part of the professional programs offered. However, there are still many challenges. In this paper we presented a tool that aims to contribute to the field of academic dishonesty in MOOCs. DetectApp allows the identification of students with suspected fraud through a report of answer matches in exams and through a report of grouping by IP addresses.

The results showed that it is possible to identify students with suspected fraud through analysis of the responses sent. It was possible to identify 8 students who shared their answers throughout the course. It was also possible to identify that the students were working in groups outside the university through the IP address of their submissions and that they were meeting not only to study but also to answer the exams together.

One of the limitations of student identification through the answer-matching report is that students are identified when they have the same questionnaire and have answered everything correctly. However, the tool allowed teachers to compare the identified students with the grades achieved and so when students get a bad grade and have the same answers it is because they have made a mistake in the same way. We mean, they shared the answers. We are working to implement a new algorithm that performs the activity mentioned but automatically and that the teacher does not have to analyze it manually.

We hope that the results shown here will be a contribution to mitigate the academic dishonesty present in the MOOC courses and that in this way they will be more easily incorporated into the university environment.

REFERENCES

[1]. E.Francesc, “Bolonia y las TIC: de la docencia 1.0 al aprendizaje 2.0”, La Cuestion Universitaria, vol. 5, nº ISSN 1988-236X, Enero 2009.

[2]. R. Moe, “The brief & expansive history (and future) of the MOOC: Why two divergent models share the same name”, Current Issues in Emerging eLearning, vol. 2, nº Iss 1 Article 2, January 2015.

[3]. J. G. Galan, “El fenomeno MOOC y la universalidad de la cultura: las nuevas fronteras de la educacion superior”, Profesorado. Revista de curriculum y formacion de profesorado, vol. 18, nº 1, pp. 73-91, 2014.

[4]. J. E. Borrero, “Revista Educacion Virtual, Pueden los cursos en linea evitar que los estudiantes hagan trampa?”, Febrero 2013. Available: https://revistaeducacionvirtual.com/archives/511.

[5]. H. C. G. E. C. N. G. G. C. Northcutt, “Measuring and Maximizing the Effectiveness of Honor Codes in Onlines Courses”, pp. 223-228, 2015.

[6]. L. K. T. D. McCabe, “Academic Dishonesty”, The Journal of Higher Education , nº 44, pp. 522-538, 1993.

[7]. I. L. C. A. D. H. G. Northcutt, “Detecting and Preventing "Multiple-Account" Cheating in Massive Open Online Course”, CoRR, 2015.

[8]. A. G. c. Z. D. P. J. A. Ruiperez Valiente, “Using Multiple Accounts for Harvesting Solutions in MOOC”, Proceedings of the Tird, ACM Conference on Learning, 2016.

[9]. J. E. S. P. V. J. B. F. M. E. Linares Espinos, V. Hernandez, “Methodology of a systematic review”, Elsevier España, 2018.

[10]. S. Lujan Mora, “MOOC (Massive Open Online Course)”, Tech. Rep., 2013.

[11]. A. Fidalgo Blanco, F. J. Garcia Peñalvo, M. L. Sein Echaluce, “Los MOOC: un analisis desde una perspectiva de la innovacion institucional universitaria”, Tech. Rep., 2017.

[12]. K. m. Chang, X. Li, A. Hauptmann, “Massive Open Online Proctor: Protecting the Credibility of MOOC Certificates”, Tech. Rep., 2015.

[13]. T. Gill, “Proctorfree: Deterring online cheating”, Tech. Rep., 2013.

[14]. H. Corrigan Gibbs, E. Cutrell, N. Gupta, C. Northcutt, “Deterring Cheating in Online Environments”, Tech. Rep., 2015.

[15]. A. Chiappe, N. Hine, J. Martinez, “Literatura y practica: una revision critica acerca de los MOOC”, Tech. Rep., 2015.

[16]. D. Wiley, “The MOOC Misstep and the Open Education Infras tructure”,
Tech. Rep., September 2014. [Online]. Available: https://opencontent.org/blog/archives/3557.

[17]. G. Siemens, “Connectivism:creating a learning ecology in d istributed environments”, Tech. Rep., 2007.

[18]. A. Berenice, R. Rodriguez, B. Pérez, “Aplicación del Modelo ELQ en la evaluación de la calidad en la educación a distancia impartida en la UPPUEBLA”, RIED: revista iberoamericana de educación a distancia, vol. 16, no 1, pp. 155-172, 2013.

[19]. Ruipérez-Valiente José A., Joksimovic Srecko, Kovanovic Vitomir, Gasevic Dragan, Merino Pedro, Delgado-Kloos Carlos, “A Data-driven Method for the Detection of Close Submitters in Online Learning Environments”, 2017.

[20]. K. Hamilton y R. Miles, “Learning UML 2.0”, O´Reilly Meida, 2008.