

Contribuciones a la identificación automática del estilo de aprendizaje del estudiante en procesos formativos en línea masivos



MARIO FERNANDO SOLARTE SARASTY

Informe
Doctorado en Ingeniería Telemática

Director:
Ph.D. Gustavo Adolfo Ramírez González

Universidad del Cauca
Facultad de Ingeniería Electrónica y Telecomunicaciones
Departamento de Telemática
Línea de investigación: Aplicaciones y Servicios sobre Internet
Popayán, abril de 2019

MARIO FERNANDO SOLARTE SARASTY

Contribuciones a la identificación automática del estilo de aprendizaje del
estudiante en procesos formativos en línea masivos

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

Doctor en:
Ingeniería Telemática

Director:
Ph.D. Gustavo Adolfo Ramírez González

Popayán
2019

A Dios a quien le debo todo.

A mi padre por inculcarme el amor hacia la ciencia y el conocimiento.

A mi madre, hermanos, sobrinos y demás familiares por su amor y compañía.

*A mi esposa y a mi hija por constituirse en el motivo de vivir y mejorar como
persona todos los días.*

A Daniel Jaramillo por su invaluable contribución con la presente investigación.

A Gustavo Ramírez por su orientación y guía. A Carlos Serrano por ser mi mentor.

*A Jesús Cerón, Juan Pablo Arcos, Amanda Noguera, Fabián Anacona, Luís
Alejandro Cruz y Sonia Guerrero por su colaboración durante el tiempo de estudio.*

*A Mar Pérez Sanagustín y Carlos Alario Hoyos por sus orientaciones y apoyo
durante las pasantías en el extranjero.*

A los participantes del proyecto MOOC Maker por permitirme aprender a su lado.

*A Alejandro Maya, Isabel Ramos, Weimar Díaz y Christian Delgado por haberme
permitido el honor de orientar sus trabajos de grado durante el Doctorado.*

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

*A Elizabeth Tabares por confiar en mi desde el primer momento y a todos aquellos
a quienes no alcanzo a mencionar y que igual contribuyeron a mi formación como
investigador.*

Contenido

Lista de Figuras.....	7
Lista de Tablas.....	10
Capítulo 1	11
1.1 Introducción de la tesis doctoral.....	11
1.2 Escenario de motivación	11
1.3 Formulación del problema	12
1.4 Objetivos	14
1.4.1 Objetivo General	14
1.4.2 Objetivos Específicos.....	15
1.5 Metodología de desarrollo.....	15
1.6 Resultados conseguidos	16
1.7 Estructura de la monografía.....	16
1.7.1 Capítulo 1 Introducción a la tesis doctoral.....	16
1.7.2 Capítulo 2 Estado del arte	16
1.7.3 Capítulo 3 Extracción de datos de una instancia de Open EdX y construcción del dataset	16
1.7.4 Capítulo 4 Analítica del aprendizaje para correlación de variables	16
1.7.5 Capítulo 5 Analíticas del aprendizaje para predicción	17
1.8 Referencias bibliográficas	17
1.9 Anexos	17
1.9.1 Anexo A Descripción curricular del curso “Astronomía cotidiana” modalidad virtual del Componente de Formación Integral Social y Humana (FISH)	17
1.9.2 Anexo B Entrenamiento de algoritmos de aprendizaje de máquina supervisados	17
Capítulo 2	18
2.1 Estado del arte	18
2.2 Metodología de la revisión.....	18
2.2.1 Estrategia de búsqueda.....	18
2.2.2 Criterios de inclusión y exclusión	18
2.2.3 Identificación de estudios relevantes	19
2.2.4 Extracción y Síntesis de Datos	19
2.2.5 Resultados	19

2.3 Estado actual del conocimiento	19
2.3.1 MOOC y Analítica del Aprendizaje	19
2.3.2 Adaptación en los MOOC	21
2.3.3 Adaptación y personalización en e-learning a través de estilos de aprendizaje	23
2.4 Brechas encontradas en la revisión sistemática	24
Capítulo 3	26
3.1 Caso de estudio: Astronomía cotidiana	26
3.2 Extracción de datos de una instancia de Open EdX	28
3.2.1 Vista de escenarios.....	29
3.2.2 Vista lógica	29
3.2.3 Vista de despliegue	30
3.2.4 Resultados	33
3.3 Prueba con datos reales.....	34
3.3.1 Ingresos en el curso y sus contenidos	34
3.3.2 Actividad en videos	35
3.3.3 Actividad en Foros.....	36
3.3.4 Interacciones con Evaluaciones	37
3.3.5 Informes automáticos	38
3.4 Construcción de <i>dataset</i> (conjunto de datos).....	43
3.5 Dataset	46
3.6 Discusión y Conclusiones.....	49
Capítulo 4	50
4.1 Analítica del aprendizaje para correlación de variables	50
4.1.1 Comportamientos característicos de la calificación definitiva y la diferencia en el aprendizaje.....	52
4.1.2 Reducción de la dimensionalidad.....	59
4.2 Clasificación y clustering	66
4.2.1 Agrupación por Rendimiento Académico	66
4.2.2 Agrupación por actividades de aprendizaje.....	70
4.3 Rediseño del MPOC teniendo en cuenta los principios del Modelo de Estilos de Aprendizaje de Kolb.....	73
4.4 Analíticas del aprendizaje como aproximación a un análisis de comportamientos de los estudiantes.....	76
4.5 Discusión y Conclusiones.....	81

Capítulo 5	83
5.1 Analíticas del aprendizaje para predicción	83
5.2 Relaciones causales y modelos predictivos.....	83
5.1.2 Gestión de la incertidumbre con modelos probabilísticos	86
5.1.3 Gestión de la incertidumbre con los componentes principales.....	89
5.2 Analítica del aprendizaje para la identificación automática del estilo de aprendizaje según Kolb.....	91
5.3 Discusión y Conclusiones.....	96
Referencias bibliográficas	98

Lista de Figuras

Figura 1. Diagrama de casos de uso	29
Figura 2. Diagrama de clases para el mecanismo de seguimiento	30
Figura 3. Diagrama de Despliegue.....	30
Figura 4. Arquitectura general mecanismo propuesto	31
Figura 5. Script de extracción	31
Figura 6. Script de procesamiento de registros	32
Figura 7. Evento del archivo tracking.log	32
Figura 8. Captura Herramienta Selene. Historial de ingresos en el curso piloto	33
Figura 9. Eventos en el csv obtenido desde Selene	33
Figura 10. Historial de las interacciones del curso virtual de Astronomía Cotidiana (imagen ajustada al primer periodo académico de 2018).....	34
Figura 11. Ingresos al curso piloto mes de febrero.....	35
Figura 12. Historial de ingresos al curso piloto mes de febrero.....	35
Figura 13. Consulta interacciones de los estudiantes con los videos del curso piloto	36
Figura 14. Consulta de las interacciones de un estudiante específico con videos	36
Figura 15. Consulta de las interacciones en el foro del curso piloto	37
Figura 16. Consulta de las interacciones en el examen final del curso piloto	38
Figura 17. Ingresos por semana al Curso Virtual de Astronomía Cotidiana 2016-I	39
Figura 18. Ingresos por semana a los contenidos del curso piloto	39
Figura 19. Cantidad de ingresos a cada una de las unidades del curso	40
Figura 20. Cantidad de reproducciones de los videos por cada unidad del curso.....	41
Figura 21. Participación en los foros por semana	41
Figura 22. Participación en las evaluaciones por semana.....	42
Figura 23. Gráfica comparativa de los resultados finales del curso piloto	42
Figura 24. Metodologías de minería de datos. Fuente (Possos & Cruz, 2017)	43
Figura 25. Ciclo de vida CRISP-DM. Fuente (Wirth, & Hipp, 2000)	44
Figura 26. Total trabajo en línea vs estratos socioeconómicos	53
Figura 27. Total de actividades prácticas vs estilo de aprendizaje Kolb (1= Acomodador, 2= Asimilador, 3= Convergente, 4= Divergente).....	53
Figura 28. Total de actividades prácticas vs. Género (1 = femenino, 2 = masculino).....	54
Figura 29. Total de actividades prácticas vs. edad.....	55
Figura 30. Total trabajo práctico vs puntaje de ingreso.....	55
Figura 31. Distribución del desempeño académico (Definitiva) por género	56
Figura 32. La diferencia de aprendizaje por género.....	56
Figura 33. Histograma del rendimiento académico	57
Figura 34. Diferencia en el histograma de la diferencia de aprendizaje.....	57
Figura 35. Rendimiento académico vs Edad	58
Figura 36. Rendimiento académico vs Estrato socioeconómico.....	58
Figura 37. Rendimiento académico vs Estilo de Aprendizaje de Kolb	58
Figura 38. Rendimiento académico vs Puntaje de ingreso	58

Figura 39. Diferencia de aprendizaje vs Edad	59
Figura 40. Diferencia de aprendizaje vs Estrato socioeconómico.....	59
Figura 41. Diferencia de aprendizaje vs Estilo de aprendizaje Kolb.....	59
Figura 42. Diferencia de aprendizaje vs Puntaje de ingreso	59
Figura 43. Biplot de PCA 1 que muestra grupos de variables. El clúster azul trata sobre los resultados de la evaluación.....	61
Figura 44. Biplot de PCA 2 que muestra grupos de variables que indican diferentes evaluaciones	62
Figura 45. Biplot de PCA 3 que muestra grupos de variables que indican actividades que dan resultados similares.....	63
Figura 46. Biplot de PCA 4 que muestra grupos de variables que indican actividades que dan resultados similares.....	65
Figura 47. Diagrama de dispersión PCA 2 (Grades) de las puntuaciones F1 y F2. El color se basa en los estilos de Kolb.....	66
Figura 48. Diagrama de dispersión PCA 2 (Rendimiento académico) de las puntuaciones F1 y F2. El color se basa en la diferencia de aprendizaje	67
Figura 49. Diagrama de caja para el promedio de la diferencia de aprendizaje por edad	70
Figura 50. Diagrama de dispersión PCA 4 (Actividades) de las puntuaciones F1 y F2. El color se basa en los estilos de Kolb	71
Figura 51. Diagrama de dispersión PCA 4 (Actividades) de las puntuaciones F1 y F2. El color se basa en la diferencia de aprendizaje	72
Figura 52. Rendimiento académico por estilo de aprendizaje según Kolb	74
Figura 53. Rendimiento académico semanal por estilo en el MPOC rediseñado	75
Figura 54. Rendimiento académico semanal por estilo en el MPOC inicial.....	75
Figura 55. Visualizaciones videos del curso segundo semestre de 2016 (Fuente: Google Analytics)	76
Figura 56. Ingresos al curso por estudiantes en el segundo semestre de 2016	77
Figura 57. Ingresos al curso por estudiantes en el primer semestre académico de 2017	78
Figura 58. Resultados de evaluación U1T1 e ingresos fuera del examen	79
Figura 59. Resultados de evaluación U2T4 e ingresos fuera del examen	79
Figura 60. Resultados de evaluación U3T3 e ingresos fuera del examen	80
Figura 61. Desempeño académico vs Diferencia de Aprendizaje (en línea - presencial)	83
Figura 62. Desempeño académico vs Accesos a la plataforma.....	84
Figura 63. Desempeño académico vs Exámenes presentados	84
Figura 64. Desempeño académico vs Acceso a recursos en línea	84
Figura 65. Desempeño académico vs Acceso a recursos en video	84
Figura 66. Desempeño académico vs Acceso a prácticas web.....	85
Figura 67. Desempeño académico vs Total actividades prácticas	85
Figura 68. Diferencia de aprendizaje vs Exámenes presentados	85
Figura 69. Diferencia de aprendizaje vs Acceso a prácticas web	85
Figura 70. Diferencia de aprendizaje vs Total actividades prácticas.....	85
Figura 71. Diferencia de aprendizaje vs Videos vistos	85
Figura 72. Probabilidad definitiva>0.6 vs Accesos a la plataforma.....	87
Figura 73. Probabilidad definitiva>0.6 vs Exámenes presentados.....	87
Figura 74. Probabilidad definitiva>0.6 vs access to web resources.....	87

Figura 75. Probabilidad definitiva>0.6 vs Acceso a recursos en video	87
Figura 76. Probabilidad definitiva>0.6 vs Acceso a prácticas web.....	87
Figura 77. Probabilidad definitiva>0.6 vs Total actividades prácticas	87
Figura 78. Diferencia de aprendizaje vs Exámenes presentados	88
Figura 79. Diferencia de aprendizaje vs Acceso a prácticas web	88
Figura 80. Diferencia de aprendizaje vs Total actividades prácticas.....	88
Figura 81. Diferencia de aprendizaje vs Videos vistos	88
Figura 82. Desempeño académico vs F1 puntajes de PCA 2, datos de calificaciones.....	89
Figura 83. Diferencia de aprendizaje vs F1 puntajes PCA 2, datos de calificaciones	90
Figura 84. Probabilidad de Diferencia de aprendizaje > 0 vs F1 puntaje de PCA 2, datos de calificaciones	90
Figura 85. Desempeño académico vs F1 puntaje de PCA 4, actividades en plataforma	90
Figura 86. Probabilidad de Desempeño académico> 0.6 vs F1 puntaje de PCA 4, actividades en plataforma.....	90
Figura 87. Diferencia de aprendizaje vs F1 puntaje de PCA 4, actividades en plataforma	91
Figura 88. Probabilidad de Diferencia de aprendizaje > 0 vs F1 puntaje de PCA 4, actividades en plataforma.....	91

Lista de Tablas

Tabla 1. Cursos virtuales con reconocimiento académico en la Universidad del Cauca	28
Tabla 2. Instancia de Open edX (Selene).....	48
Tabla 3. Sistema de extracción de datos de Selene	48
Tabla 4. SIMCA	48
Tabla 5. Encuestas para determinación del estilo de aprendizaje (Google Forms y scripts en Matlab)	48
Tabla 6. PCA 2 Biplot Clusters	63
Tabla 7. PCA 3 Biplot Clusters	64
Tabla 8. PCA 4 Biplot Clusters	64
Tabla 9. Distribución de los estilos de aprendizaje de Kolb	67
Tabla 10. Recuento de muestras en el cuadrante de mayor rendimiento y porcentajes.....	68
Tabla 11. Promedios F1 por estilo de aprendizaje y género	68
Tabla 12. Diferencia de aprendizaje por estilo de aprendizaje y género.....	69
Tabla 13. Diferencia de aprendizaje vs estrato socioeconómico	69
Tabla 14. Conteo de muestras en el cuadrante de nivel de actividad más alto y porcentajes.....	72
Tabla 15. Promedios F1 (medida de actividad general) por estilo de aprendizaje y género	73
Tabla 16. Organización por cuartiles según ingresos previos a un examen (Fuente propia)	78
Tabla 17. Organización de estudiantes según los resultados de la evaluación	80
Tabla 18. Estadística R^2 para los modelos determinísticos de la Calificación final en comparación con los modelos probabilísticos.....	88
Tabla 19. Estadística R^2 para los modelos determinísticos de la Diferencia de aprendizaje en comparación con los modelos probabilísticos	89
Tabla 20. Resultado del entrenamiento de distintas categorías de algoritmos	92
Tabla 21. Mejor entrenado mecanismo árbol.....	93
Tabla 22. Acierto en la predicción Dataset de Entrenamiento	95
Tabla 23. Acierto en la predicción Dataset de Prueba	95

Capítulo 1

1.1 Introducción de la tesis doctoral

La Analítica del Aprendizaje, el Aprendizaje de Máquina –supervisado y no supervisado- y las estrategias de los cursos en línea en ambientes de masividad son técnicas que aparecieron en la primera década de este milenio y se consolidaron –en diversas manifestaciones- en la década siguientes en diferentes contextos.

Aunque los MOOC (Massive Open Online Courses - cursos en línea abiertos y masivos) surgieron en un ambiente de educación no formal, su reciente articulación en los currículos universitarios hace que sea necesario resolver distintos retos para minimizar las consecuencias de su modelo, como por ejemplo aquellos relacionados con su rigidez en materia pedagógica.

Lo anterior es, a su vez, una oportunidad magnífica para impulsar la introducción de tecnologías específicas para conocer y comprender el comportamiento de los distintos actores del proceso formativo, especialmente profesores y estudiantes teniendo en cuenta las significativas cantidades de personas involucradas en estos cursos.

Uno de los beneficios anunciados por la aplicación de las técnicas de la Analítica del Aprendizaje, unido al potencial de las aplicaciones web que dan soporte a los cursos en línea en ambientes de masividad es el aprovechamiento de las grandes cantidades de datos para poder comprender el comportamiento de los estudiantes.

Un aporte importante de la presente investigación es la integración de las tres técnicas indicadas anteriormente (aprendizaje de máquina, analíticas del aprendizaje, cursos masivos en línea) y su aplicación para el reconocimiento académico de conocimientos y habilidades adquiridos a través de esta modalidad educativa.

1.2 Escenario de motivación

La Universidad del Cauca, que hace presencia en el Departamento del Cauca, cercana a cumplir 200 de fundación y prestación del servicio de la educación superior, se ha caracterizado por ofrecer programas académicos en modalidad presencial tradicional, desperdiando las posibilidades que ofrecen las tecnologías telemáticas y los modelos híbridos que ellas pueden ofrecer en los procesos formativos.

La irrupción de los MOOC en el ámbito educativo internacional puede ser el catalizador que impulse la adopción de modelos educativos no presenciales en Instituciones de Educación Superior. Pero se puede correr el riesgo de un uso netamente instrumental de la tecnología, olvidando los propósitos máximos de los modelos formativos.

Dado que los MOOC fueron creados desde la perspectiva de la educación no formal (no conducente a título), su integración en procesos curriculares demanda nuevos retos (técnicos, administrativos, pedagógicos) que permitan la acreditación de conocimientos y habilidades adquiridas a través de los cursos masivos en línea sin afectar la calidad de la formación.

Las grandes cantidades de información que se pueden recoger a través de las plataformas MOOC deben servir de insumo para comprender los procesos de aprendizaje de los estudiantes en su interacción con los servicios de las aplicaciones web que dan soporte a una cantidad significativa de estudiantes.

Una vez organizados los complejos conjuntos de datos producidos tanto por las plataformas de soporte al aprendizaje como de sistemas de gestión académicos, integrando con información de otras fuentes (encuestas de caracterización, apreciaciones, evaluación, etc.) y otras variables de interés (estilos o preferencias de aprendizaje), se pueden entrenar algoritmos para predecir el comportamiento de distintas variables y tomar acciones para mejorar los indicadores del proceso, como la retención o el rendimiento académico.

1.3 Formulación del problema

Un MOOC es un curso ofrecido a través de Internet, sin pre-requisitos y abierto a una gran cantidad de estudiantes que integra el potencial de las redes sociales, la orientación de un experto disciplinar, millares de recursos abiertos y de libre acceso que se encuentran en Internet (McAuley et al. 2010) donde lo más importante es la participación activa de cientos o miles de estudiantes al mismo tiempo que se autorregulan de acuerdo a propósitos comunes, generalmente de aprendizaje. Aunque puede existir el apoyo docente asistentes y una duración de algunas semanas, generalmente no contiene obligaciones para los inscritos a excepción del acceso a internet y la motivación personal.

Los MOOC aparecen en el 2008 como evolución de los Recursos Educativos Abiertos (REA), como propuesta para universalizar la educación y ofrecer educación gratuita y de calidad a personas que residen en zonas lejanas o desfavorecidas, inspirados en el Conectivismo (Siemens, 2005), una innovadora Teoría Educativa propuesta por George Siemens en 2005. Se caracterizan por el ofrecimiento de cursos gratuitos accesibles a través de Internet (Wiley, 2012), no conducentes a titulaciones pero con la posibilidad de emisión de un certificado de aprobación después del respectivo pago (McAuley, 2010), generalmente de corta duración (Liyanagunawardena, Adams & Williams, 2013), centrados en los contenidos -que deben ser abiertos- básicamente tipo video (Gou et al. 2014), con actividades evaluativas relativamente sencillas (Roig, Mengual & Suarez, 2014) sin límite en la cantidad de inscritos (Jansen y Schuwer 2015).

Si bien los MOOC, aunque no con ese nombre, aparecieron en el 2008 con el desarrollo del curso "Conectivismo y conocimiento colectivo" ofrecido por Stephen Downes en la Universidad de Manitoba (Mackness & Williams, 2010), alcanzan un auge inusitado con la experiencia de Sebastian Thrun (profesor de la Universidad de Stanford) y Peter Norvig (Investigador de Google) al orientar un MOOC sobre Introducción a la Inteligencia Artificial donde se matricularon aproximadamente 160.000 estudiantes de los cuales alrededor 34.000 lo aprueban (Martin, 2012).

Los MOOC alcanzan fama mundial en el 2012 con la aparición de diversas iniciativas -la mayoría privadas- que utilizan plataformas telemáticas mediante las cuales se ofrecen

cursos de universidades de prestigio a nivel mundial, como son Udacity, Coursera, edX, MiriadaX, entre otras (Hernández, 2013), hasta tal punto que el New York Times los nombró el personaje del año (Pappano, 2012).

El alto éxito conseguido en los MOOC en poco tiempo, así como su rápida evolución, ha provocado la aparición de múltiples variantes, como por ejemplo los SPOC (Small Private Online Courses - cursos en línea pequeños y privados) que fueron introducidos en 2013 por Armando Fox, profesor de la Universidad de Berkeley, para caracterizar el uso de recursos producidos en MOOC en cursos presenciales (Fox, 2013) de ahí que no sean procesos abiertos ni masivos.

Un año después de los SPOC, Wenge Guo, profesor de la Universidad de Beijing, propone el término MPOC (Massive Private On line Courses - cursos en línea privados y masivos) bautizando su propuesta de usar la estrategia MOOC en ambientes educativos con reconocimiento académico manteniendo el componente no presencial del proceso educativo, pero sin profundizar en aquellos aspectos que necesariamente deben ser adecuados a su nuevo contexto de empleo.

De igual forma, muchas instituciones alrededor del mundo han definido un conjunto de criterios para convalidar los conocimientos y habilidades adquiridos a través de la estrategia MOOC (Solarte, 2016), sobresaliendo las siguientes estrategias:

- Reconocimiento académico de MOOC ya realizados.
- Integración curricular de MOOC propios o de otras instituciones.
- Integración curricular de un conjunto de MOOC como parte de una especialización o master presencial.

No existe consenso respecto al límite inferior en el número de estudiantes que pueda tener un MOOC; mientras (Neubaum et al, 2014) publicó una investigación sobre motivación para participación en un MOOC “semiabierto” donde se inscribieron 91 estudiantes, en (Delgado-Kloos et al, 2014) se describe experiencia del desarrollo de SPOC con más de 500 estudiantes regulares de la Universidad Carlos III de Madrid, es decir, la frontera entre un MPOC, un SPOC o un MOOC puede ser difusa en función del autor analizado.

En este punto aparece una primera pregunta de investigación ¿Cómo puede una Institución de Educación superior integrar a sus procesos educativos con reconocimiento académico la estrategia de los cursos en línea en ambientes de masividad y sacar provecho de la analítica del aprendizaje?

Con la consolidación de las plataformas que ofrecen MOOC, también empiezan a evidenciarse problemas y dificultades en el modelo inicial que los sustentó: bajas tasas de retención (Forbes, 2012), (Adamopoulos, 2013); calidad de la formación (Conole, 2013) eficacia pedagógica (Sonwalkar, 2013) y extrema rigidez (Cabero et al, 2014); carencia de atención personalizada (Zapata, 2013a); y sostenibilidad (Yuan & Powell, 2013), entre otros, que deben ser abordados desde distintas instancias y esfuerzos interdisciplinarios para asegurar la supervivencia de esta modalidad de la educación en línea.

En septiembre de 2013 Sonwalkar propone el concepto de aMOOC (adaptive MOOC) ofreciendo alternativas para empezar a abordar la dificultad de la falta de flexibilidad en los MOOC y, a su vez, abriendo nuevos campos para la investigación de las siguientes

cuestiones: ¿Cómo se diseña un MOOC para que sea adaptable?, ¿Cómo adaptar un MOOC?, ¿Qué se puede adaptar de un MOOC?, ¿Cómo identificar perfiles o patrones de comportamiento de los estudiantes al desarrollar un MOOC?, ¿Cuál es el efecto, en todos los niveles, de la adaptación de un MOOC?, ¿es posible llegar a personalizar el desarrollo de un MOOC?

Estudios posteriores han empezado a abordar temas relacionados con los estilos de aprendizaje aplicados a los MOOC. Según (Chang, Hung & Lin, 2015) existe un 90% de probabilidad que el estilo de aprendizaje del estudiante influye en sus intenciones respecto a un MOOC, lo cual da una idea de las motivaciones que tienen las personas para inscribirse a un MOOC, pero que no es suficiente para resolver la esencia de los problemas asociados a su rigidez.

Los estilos de aprendizaje son rasgos cognitivos, efectivos y fisiológicos, que sirven como indicadores relativamente estables de cómo los estudiantes perciben, interaccionan y responden en un ambiente de aprendizaje (Keefe, 1988). Honey y Munford en 2000 los definen como una descripción de las actitudes y comportamientos que determinan la forma preferida de aprendizaje del individuo. Respecto al uso de los estilos de aprendizaje en ambientes web, (Manocher, 2006) afirma que tener en cuenta a los primeros es significativamente importante para el rendimiento y apropiación de conocimiento.

Respecto a los procesos formativos mediados por tecnologías telemáticas, (Xu et al, 2014) afirman que las posibilidades de adaptación y personalizado mejoran la eficacia de aprendizaje en línea en términos de examinación, satisfacción y la autoeficacia.

Una de las principales críticas a los MOOC es su falta de pedagogía (Zapata, 2012), circunstancia heredada por los aMOOC, así como también la rigidez pedagógica (Sonwalkar, 2013), aspecto de vital importancia si de cuestiones de calidad educativa se trata. Aunque se han formulado metodologías para el diseño de MOOC adaptables (Fidalgo, 2013), falta un gran trecho para conciliarlas con técnicas de diseño instruccional que propenden por un proceso educativo personalizado. De otra parte, se ha detectado una tendencia a utilizar distintos tipos de estilos de aprendizaje como elemento de adaptación del MOOC (Daradoumis et al, 2013) sin elaborar estudios comparativos sobre sus bondades o deficiencias. De igual forma, los MOOC siguen evolucionando de manera vertiginosa y si bien han llegado para quedarse, seguramente su forma definitiva será muy diferente a la actual (Zapata, 2013a), siendo los SPOC una de las variantes que más expectativas están generando a futuro (Delgado-Kloos et al, 2014).

De lo anterior surge la siguiente pregunta de investigación: ¿Cómo identificar de manera automática el estilo o las preferencias de aprendizaje de los usuarios para una posterior adaptación de procesos formativos en línea masivos bien sean abiertos o privados?

1.4 Objetivos

1.4.1 Objetivo General

Proponer un mecanismo para la identificación automática del estilo de aprendizaje en procesos de formación en línea masivos abiertos o privados.

1.4.2 Objetivos Específicos

- Construir un conjunto de datos que relacione la interacción de los estudiantes con una plataforma MOOC con sus estilos de aprendizaje según el modelo de Kolb.
- Determinar el tipo de algoritmo de aprendizaje de máquina supervisado que mejor clasifique un estilo de aprendizaje en función del conjunto de datos construido en el proyecto.
- Desarrollar un estudio de caso que permita verificar la identificación automática de estilos de aprendizaje en un curso en línea masivo con reconocimiento académico en la Universidad del Cauca.
- Aplicar técnicas de Analítica del Aprendizaje para encontrar asociación de variables de interés involucradas en el caso de estudio.

1.5 Metodología de desarrollo

Para la consecución de los resultados de la investigación se siguieron las siguientes metodologías:

- Mapeo sistemático para la elaboración del estado del arte.
- 4+1 vistas para la descripción de la Arquitectura para la captura de requisitos, diseño e implementación de un prototipo que permitió tener acceso a la información de interacción de los estudiantes con una instancia de Open edX instalada en la Universidad del Cauca.
- CRISP-DM para la construcción de los conjuntos de datos empleados en la investigación.
- Análisis estadísticos para la verificación de hipótesis de trabajo con aplicaciones como R.
- Análisis de Componentes Principales para reducción de la dimensionalidad en los conjuntos de datos.
- Diseño del caso de estudio mediante la metodología de Robert Yin.

1.6 Resultados conseguidos

Durante la investigación se consiguieron los siguientes resultados:

- Construcción de dos conjuntos de datos (*dataset*) iniciales con más de 70.000 registros relacionados con la actividad de profesores, tutores, administrador de plataforma y estudiantes del curso Astronomía cotidiana.
- Construcción de dos *dataset* procesados con información de interacción de estudiantes asociadas al estilo de aprendizaje según Kolb e información demográfica y académica.
- Definición de *dataset*s de prueba y entrenamiento a partir de los *dataset* procesados.
- Aplicación de un *dataset* de entrenamiento con distintos algoritmos de aprendizaje supervisado para identificar el más adecuado para la predicción.
- Predicción del estilo de aprendizaje según Kolb del *dataset* de prueba.
- Reducción de complejidad y análisis estadístico de las variables recopiladas durante la investigación.

Participación en el proyecto MOOC Maker entre octubre de 2015 a octubre de 2018, financiado por la Comisión Europea a través del programa Erasmus+.

1.7 Estructura de la monografía

1.7.1 Capítulo 1 Introducción a la tesis doctoral

Donde se presenta de manera resumida la motivación, formulación del problema, objetivos de la investigación abordados, la metodología de desarrollo y resultados conseguidos.

1.7.2 Capítulo 2 Estado del arte

Apartado que contiene una breve descripción del estado actual de conocimiento referido a diversos tópicos relativos a los cursos abiertos en ambientes de masividad, estilos de aprendizaje, identificación de perfiles de aprendizaje en MOOC en particular y en educación en línea en general, y las brechas encontradas a partir de las cuales se determinó el alcance de la presente investigación.

1.7.3 Capítulo 3 Extracción de datos de una instancia de Open EdX y construcción del *dataset*

Que contiene la descripción y diseño del caso de estudio, además del proceso de diseño e implementación de un prototipo para la extracción y organización de datos de interacción de los estudiantes con la plataforma web que alojó el caso de estudio, así como la descripción del *dataset* construido.

1.7.4 Capítulo 4 Analítica del aprendizaje para correlación de variables

En donde se detalla el estudio que pretende contribuir a determinar la correlación o asociación entre las variables recopiladas y dar respuesta a aspectos tales como:

- Determinar si hay diferencias en el rendimiento académico entre los diferentes estilos de aprendizaje según Kolb.

- Determinar si existen correlaciones entre las actividades realizadas en la plataforma, tales como número de accesos, vídeos reproducidos, recursos web accedidos, participación en foros, laboratorios de prácticas en línea, etc.
- Clasificar a los estudiantes por sus actividades de plataforma y determinar que dicha clasificación está relacionada con los estilos de aprendizaje de Kolb.
- Determinar si los factores demográficos influyen en el desempeño de los estudiantes, entre otros

1.7.5 Capítulo 5 Analíticas del aprendizaje para predicción

Apartado donde se aplican diferentes técnicas para intentar encontrar, por un lado, modelos predictivos para el rendimiento académico y la diferencia de aprendizaje, variable que relaciona el desempeño académico de un estudiante en el curso en línea con su promedio académico -presencial-. Finalmente, se ilustran los resultados del entrenamiento del algoritmo J48 para predecir el estilo de aprendizaje en función de las variables asociadas a la interacción de los estudiantes con la aplicación web donde se alojó el caso de estudio.

Al final de cada capítulo se presenta la discusión de resultados, las conclusiones y los trabajos a future propuestos.

1.8 Referencias bibliográficas

1.9 Anexos

Información adicional no consignada en la monografía se ha organizado a través de los siguientes anexos:

1.9.1 Anexo A Descripción curricular del curso “Astronomía cotidiana” modalidad virtual del Componente de Formación Integral Social y Humana (FISH)

1.9.2 Anexo B Entrenamiento de algoritmos de aprendizaje de máquina supervisados

Capítulo 2

2.1 Estado actual del conocimiento

En este apartado contiene una breve descripción del estado actual de conocimiento referido a la adaptación en cursos masivos abiertos y en línea y a la identificación de perfiles de aprendizaje en MOOC en particular y en educación en línea en general, y las brechas encontradas a partir de las cuales se determina el alcance de la presente investigación.

2.2 Metodología de la revisión

Siguiendo lo propuesto en (Petersen et al, 2008) para la realización del mapeo sistemático a continuación, se describen las fases más relevantes:

2.2.1 Estrategia de búsqueda

La búsqueda de referencias bibliográficas útiles se realizaba cada seis meses, la última de ellas en octubre de 2018 en las siguientes bases de datos electrónicas: Web of Science, Scopus y Google Scholar. Las cadenas de búsqueda estuvieron construidas alrededor de grupos de palabras clave y sus distintas combinaciones:

- Cursos en línea en ambientes de masividad (MOOCs, MPOC, SPOC y sus variantes)
- Estilos de aprendizaje en educación en línea
- Identificación automática del estilo de aprendizaje

2.2.2 Criterios de inclusión y exclusión

Para ser incluido, un artículo debe cumplir los siguientes requisitos:

- estar en idioma inglés o español;
- contener al menos dos de las cadenas de búsqueda
- encontrarse no más allá de los primeros 200 resultados en cada búsqueda
- si contiene la palabra MOOC, no haber sido publicado antes del 2008.

2.2.3 Identificación de estudios relevantes

Para la inclusión de los artículos se tuvo en cuenta –en primera instancia- su título, luego resumen y conclusiones. Se excluyeron aquellos artículos que no cumplieron los criterios de inclusión. Los artículos potencialmente elegibles se obtuvieron para una lectura completa y luego se determinó su inclusión si se consideraron relevantes en su contenido. Se eliminan aquí los artículos repetidos.

2.2.4 Extracción y Síntesis de Datos

Para la gestión de la información contenida en los artículos relevantes se desarrolló un banco de datos con los contenidos extraídos de los artículos incluidos en la revisión. Las características de interés para esta revisión fueron:

- Diseño de MOOC
- Analítica del aprendizaje en MOOC
- Reconocimiento académico de MOOC
- Estilos de aprendizaje
- Técnicas para identificación automática del perfil de estudiante

2.2.5 Resultados

Con la estrategia de búsqueda planteada, a lo largo de casi cinco años que duró la investigación, se obtuvieron 10.620 artículos, de los cuales fueron seleccionados 1379 después de considerar su título, el número se redujo a 201 luego de revisar el resumen y las conclusiones. Estos fueron descargados y sometidos a los criterios de inclusión y exclusión. Finalmente se tuvieron en cuenta 98 artículos que aprobaron los criterios de inclusión para esta revisión.

2.3 Marco conceptual y trabajos desarrollados

A continuación, se presentan las bases teóricas y los referentes conceptuales empleados en el desarrollo de la investigación.

2.3.1 MOOC y Analítica del Aprendizaje

Los MOOC son una de las formas más recientes del Movimiento Educativo Abierto (Baraniuk, 2007), cuyos principios se han construido sobre la ideología acerca que el conocimiento es un bien común (Ehlers, 2011) el cual pertenece a la humanidad en su conjunto. Este movimiento considera a la educación como un motor de desarrollo que debe incentivar la creación y difusión universal del conocimiento a través de diversos canales, incluyendo las Tecnologías de la Información y la Comunicación - TIC (Dans, 2009). Exponentes de este movimiento se arrojan bajo el término genérico de Educación Abierta,

y pueden ir desde los REA, materiales educativos digitales generalmente dispuestos a través de repositorios (Campbell, 2004) hasta los MOOC de nuestros días.

El auge de los REA, cuya máxima expresión seguramente fue la apertura a cientos de contenidos de prestigiosas universidades alrededor del mundo entero a través de iniciativas conocidas como Open Courseware (Tovar & Divoska, 2013), pero miradas críticas a este proceso, como la realizada por Parrish en 2004, señalan que los cambios en las prácticas educativas, (tanto de profesores como de estudiantes), han sido mínimas, lo cual hace pensar que producir y utilizar REA no es suficiente para generar innovación educativa (Ehlers, 2011). Una alternativa a esta situación, ha resultado aparentemente muy difícil de implementar en la práctica: en lugar de enfocarse en los contenidos abiertos, el énfasis habría que hacerlo en hacer más abiertas las prácticas educativas, perspectiva desde la cual surgió una interesante práctica educativa abierta: la enseñanza abierta, la cual ha encontrado una manera contemporánea de implementación en forma de lo que se conoce como MOOC (Chiappe, Hine & Martínez, 2015).

Los MOOC son cursos en línea, cuyo objetivo principal es brindar alternativas educativas soportadas en tecnologías de Internet de acceso abierto. Su clasificación más general los divide en dos, los cMOOC y los xMOOC (Guárdia, Maina & Sagrá, 2013) (Williams, Adams, 2013) (SCOPEO, 2013). Los cMOOC tienen como objetivo la generación del conocimiento de manera colaborativa y se basan en la cantidad de aportes que pueden ser generados por sus participantes. Sin embargo, estos generalmente no cuentan con una estructura, ni con una evaluación que permita medir de alguna forma el conocimiento o el nivel de aprendizaje que adquiere el estudiante. Por otra parte, los xMOOC son cursos que cuentan con un contenido estructurado y evaluaciones, se basan en la presentación de contenidos a través de videos cortos y documentos de apoyo; son los más parecidos a un curso en línea tradicional y se han vuelto muy populares, debido a que muchas de las plataformas están orientadas a ofrecer este tipo de cursos (SCOPEO, 2013).

Los MOOC ofrecen a los estudiantes la oportunidad de tomar cursos de instituciones y/o profesores de prestigio, sin ningún tipo de requisitos previo; se desarrollan en un tiempo predefinido, y siguen un plan de estudios, los estudiantes consultan sesiones grabadas en video acompañados de conjuntos de tareas semanales, recursos en línea, lecturas en línea, preguntas, y exámenes -parciales y finales- (Creed-Dikeogu & Clark, 2013). Sin embargo, no hay expectativas de participación predefinidas, los estudiantes pueden completar a su propio ritmo tan poco o tanto del curso como deseen (Martin, 2012).

Pero los MOOC no son una panacea educativa (Creed-Dikeogu & Clark, 2013). Entre los problemas principales se encuentra la evaluación de las grandes cantidades de estudiantes que se inscriben, motivo por el cual la mayoría de las evaluaciones son simplemente preguntas de selección múltiple. Los estudiantes tampoco pueden construir relaciones con los profesores o conseguir la realimentación en profundidad sobre su progreso educativo y que carecen de un diseño instruccional efectivo (Mahraj, 2012), lo cual sugiere la pérdida de interacción cara a cara necesaria para el éxito del proceso formativo presencial.

En cuanto lo pedagógico, los MOOC tiene su fundamentación epistemológica en el Conectivismo (Siemens, 2004). Downes (2011) afirma que “... *los MOOCs combinan el contenido abierto y la enseñanza abierta, pero también son compatibles con la participación masiva; eso se logra mediante la adopción de una pedagogía y una estructura conectivista*”. Siemens considera que el conectivismo es una teoría que se caracteriza por considerar el aprendizaje como una extensión del conocimiento y de la comprensión a través de una red

personal, base sobre la cual se organizan actividades y recursos en la programación educativa.

Zapata (2013) critica este modelo proponiendo un enfoque para él revolucionario: la individualización del aprendizaje. Asegura que la relación entre profesores y estudiantes en un MOOC a través de materiales, en realidad aísla a la mayoría de estudiantes y sólo es aprovechado por una minoría de alumnos aventajados, con un alto costo social asociado. Aunque existe una rama del *e-learning* conocido como *e-learning* adaptativo, entendido como programas computacionales capaces de mostrar información para usuarios (de forma individual) tomando en cuenta un modelo detallado de sus objetivos, intereses y preferencias (Berlanda & García, 2004), Zapata propone individualizar no solamente la presentación del contenido sino el diseño instruccional en su conjunto para adaptarlo a las características de aprendizaje detectadas en el entorno personal de los estudiantes.

De igual forma, Zapata (2013b) avizora que la educación superior de los próximos años, será un producto híbrido con pluralidad de opciones metodológicas donde el parámetro ayuda pedagógica tendrá valores que oscilen entre lo que son ahora los xMOOC puros y una individualización basada en el perfil de aprendizaje y en la analítica que se haga para cada caso; con ello introduce el elemento de la Analítica del Aprendizaje (Duval, 2011) como elemento mediador entre la captura del perfil de un estudiante y el diseño instruccional capaz de construir y ofrecer MOOC personalizados. A pesar de lo anterior, en este trabajo no se menciona ni la técnica para diseñar MOOC personalizables, ni tampoco la manera en la cual la analítica del aprendizaje identifica las variables que corresponden a la caracterización del perfil de los estudiantes.

La Analítica del Aprendizaje, es una de las macro tendencias de la actualidad y una de las grandes innovaciones del mundo educativo global del presente milenio. Se trata del análisis de los *big data* -gran cantidad de datos generados por los estudiantes en su proceso de formación y aprendizaje- que permite predecir tendencias y patrones en el comportamiento de los estudiantes y personalizar los entornos de aprendizaje en función de las necesidades de cada individuo (Fournier, Kop, & Sitlia, 2011). George Siemens y otros autores, lo definen como *“la medición, recolección, análisis y presentación de datos sobre los aprendices y sus contextos, para propósitos de entendimiento y optimización del aprendizaje y el entorno en el que se produce”* [10].

El uso de la analítica del aprendizaje, junto al análisis de redes sociales, en los equipos responsables del diseño y ofrecimiento de un MOOC, podría dar respuesta a las preguntas típicas y proporcionar elementos para corregir temas de deserción, bajos resultados de aprendizaje y ofrecer un entorno de aprendizaje adaptativo (Clark, 2013).

2.3.2 Adaptación en los MOOC

Como se comentó en la formulación del problema del Capítulo 1, el concepto aMOOC fue postulado en septiembre de 2013 en (Sonwalkar, 2013) donde se presentan los resultados del primer MOOC adaptativo soportado en una plataforma específica para su ofrecimiento llamada AMOL (*Adaptive Mobile Learning*) implementada en los servicios en la nube de Amazon, apoyado en un marco pedagógico para la adaptación dinámicas de contenidos a cuatro dimensiones de aprendizaje y cinco estrategias educativas propuestas por el mismo autor. Dicha investigación se centra en los resultados de la arquitectura distribuida empleada para la atención de miles de estudiantes, pero no profundiza en la aproximación pedagógica empleada, llamada *“learning cube”*, ni tampoco en el comportamiento de los

estudiantes durante la experiencia; de igual forma, el estudio no indica la forma en la cual AMOL detecta o se informa del perfil del estudiante.

Luego de la experiencia anterior se han comenzado a divulgar diversos trabajos que investigan alrededor de la adaptabilidad en los MOOC, bien sea desde el punto de vista de las plataformas que los deben ofrecer, las metodologías para su diseño, y los elementos que se pueden variar en un curso masivo en función de las características de los usuarios; poco se ha escrito sobre la identificación automática de estilos de aprendizaje, donde la mayoría de autores sigue empleando la aplicación de una encuesta en línea. A continuación, se resumen los principales de ellos:

(Fidalgo, García & Seins, 2013) presenta una aproximación a una propuesta para el desarrollo de cMOOC adaptables que se espera útil para mejorar las tasas de deserción y poca cooperación entre estudiantes. Propone que la adquisición de datos y evaluación se haga a través de Analítica del Aprendizaje, aunque no dice cómo, tampoco menciona aspectos técnicos de la plataforma que soportará el ofrecimiento de los MOOC. Esta propuesta no ha sido corroborada mediante casos de estudio o desarrollos experimentales.

(Rimbaud et al., 2013), presenta una propuesta soportada en taxonomías y técnicas de aprendizaje adaptativo para mejorar los resultados del aprendizaje del idioma inglés para estudiantes que no lo hablan. No ha sido verificada mediante un estudio de caso o desarrollo experimental.

(Bansal, 2013) integra un sistema de recomendación para la adaptación de MOOC que puede adaptarse a las necesidades de los estudiantes y proporcionar recomendaciones personalizadas según un modelo de participantes según su nivel de conocimiento. No emplea estilos de aprendizaje. (Torres, Infante, Díaz, 2014) plantea algunas reflexiones acerca del potencial de las técnicas de inteligencia artificial para el desarrollo de MOOC personalizables. El trabajo sólo profundiza en una descripción modular de un sistema de recomendaciones para la tutorización, pero no indica cómo se resuelven los problemas técnicos asociados a la integración de los sistemas de recomendación con las plataformas para el ofrecimiento de MOOC.

(Baer et al., 2013) propone una estrategia para la transformación y sostenimiento de sistemas de aprendizaje personalizados para ofrecer experiencias formativas individualizadas desde una perspectiva institucional. La propuesta no tiene en cuenta las características de los MOOC, ni tampoco hace alusión a estilos de aprendizaje.

(Fasihuddin, Skinner & Athauda, 2014), introduce un enfoque basado en las ciencias cognitivas para mejorar ambientes abiertos de aprendizaje y proporcionar personalización, utilizando el modelo de estilos de aprendizaje de Felder y Silverman que utiliza un agente de identificación para la determinación del perfil según patrones de interacción con materiales y emplea un agente recomendador responsable de soportar navegación adaptable. No se ha hecho una implementación de la propuesta.

(Armendariz, McHardy & García, 2014) presenta una propuesta de herramienta (Online Course Tool for Adaptive Learning -OCTAL-) para la personalización de ejercicios acorde a un modelo del estudiante de MOOC con el propósito de mejorar la experiencia metacognitiva. Se encuentra en marcha un proceso de verificación sin resultados publicados.

(Grünewald, Totschnig & Willems, 2013) expone reflexiones y lineamientos para el diseño de MOOC que contemplen el estilo de aprendizaje según Kolb y el uso técnicas de gamificación y laboratorios virtuales. No hace identificación automática del perfil de aprendizaje y sus recomendaciones metodológicas son para un curso de “Internetworking con TCP/IP”, no necesariamente son extensibles a otras temáticas.

(Sunar, et al, 2015) realiza un estudio del estado del arte sobre personalización en MOOC, incluyendo aspectos como el pedagógico, la sostenibilidad del aprendizaje y los modelos de negocio, siendo especialmente crítico en los componentes metodológicos para el diseño de MOOC adaptables, indicando que las implementaciones más populares abarcan la evaluación y realimentación personalizada y los sistemas de recomendación de contenidos.

(Yan-hong, Bo, & Jian-hou, 2015) propone un Modelo de Aprendizaje Personalizado para e-learning, que igualmente puede ser aplicado a MOOC, que hace uso de los principios del *e-learning* adaptativo. Este modelo no se verificó mediante alguna implementación.

(Onah, & Sinclair, 2015) propone un marco de trabajo para recomendar materiales instrucciones en función del perfil de los estudiantes y los objetivos de aprendizaje fijados por ellos mismos. En el perfil no incluyen los estilos de aprendizaje ni tampoco ha sido verificado mediante un desarrollo experimental o estudio de caso.

En resumen, se ha trabajado más en la adaptación de algunos componentes de los MOOC según alguna característica de los estudiantes, que en la identificación automática de sus estilos de aprendizaje.

2.3.3 Adaptación y personalización en *e-learning* a través de estilos de aprendizaje

Respecto a las técnicas de adaptación y personalización en *e-learning* a través de estilos de aprendizaje, se destaca el trabajo de Alshammari, Anane y Hendley con el desarrollo de complejo marco de trabajo para la construcción de rutas de aprendizaje en función del estilo de aprendizaje de Felder – Silverman y la evaluación de diversos aspectos como satisfacción del estudiante, la evaluación, seguridad computacional, usabilidad, entre otros Alshammari, M., Anane, R. & Hendley, J. (2014, 2015a, 2015b). Estos trabajos se han aplicado en procesos de *e-learning* tradicional, no en MOOC, ni tampoco usan otros estilos de aprendizaje.

(Ghazali, Noor, & Saad, 2015) realiza una revisión sobre métodos y aproximaciones en ambientes de *e-learning* para su personalización, clasificando las técnicas para detección de los estilos de aprendizaje y proponiendo su propio modelo para la personalización del proceso de aprendizaje. Este trabajo, además de no contemplar las características específicas de los MOOC, tampoco ha sido verificado por una implementación.

(Fasihuddin, Skkiner, & Athauda, 2015) propone un modelo para identificar automáticamente el estilo de aprendizaje de Felder – Silverman mediante la búsqueda de patrones de comportamiento con objetos de aprendizaje, el modelo ha sido verificado a través de un estudio de caso que no fue un MOOC sino un curso convencional.

(Hung, Chang, & Lin, 2016) propone la utilización de un algoritmo K-means para la identificación del estilo de aprendizaje Felder – Silverman en función de los hábitos de los estudiantes cuando usan la resolución de problemas como técnica de estudio.

(Ahmad, Ishak, Alias & Mohamad, 2015) propone los mapas autoorganizados para clasificar diferentes grupos de estudiantes, con el fin de presentar contenidos adaptados a sus perfiles. Presenta como trabajo a futuro, el procesamiento en tiempo real de los datos, con miras a mejorar el rendimiento y la interacción de los estudiantes con los LMS.

(Hamtni, & Ateia, 2015) utilizando un sistema basado en reglas, detectaron automáticamente el estilo de aprendizaje según el modelo VAK, a partir de la interacción de los estudiantes con los contenidos, estudios de casos, ejemplos, ejercicios y evaluaciones en un curso de *e-learning*. La precisión de su algoritmo alcanzó el 52,78%, relativamente bajo según los propios autores.

(Abdullah et al, 2015) presenta un enfoque para la detección automática del estilo de aprendizaje según Felder – Silverman, a través de árboles de decisión, Los resultados encontraron un 76% de precisión.

(Dissanayake et al, 2018) expone un mecanismo de aprendizaje de máquina no supervisado que comparan con el rendimiento del estudiante en un MOOC para encontrar la correlación temporal con las interacciones de vídeo, el análisis se realiza con apenas un video en dos cursos diferentes y no especifica un modelo específico de estilos de aprendizaje.

En resumen, los trabajos sobre técnicas automáticas para identificar estilos de aprendizaje no se han aplicado en MOOC, muestran diversos niveles de resultados y se centran en encontrar el estilo de aprendizaje según el modelo de Felder – Silverman.

Un mapeo sistemático elaborado por el autor (Solarte, 2015) determinó una baja disponibilidad de conjuntos de datos sobre MOOC, como consecuencia del modelo de negocio de muchos proveedores de MOOC de utilizar los datos obtenidos en sus plataformas como elemento comercial (Yuan & Powell, 2013); ninguno de conjuntos de datos a los que se tuvo acceso contó con información asociada a Modelos de Estilos de Aprendizaje de los estudiantes (Solarte, 2015).

2.4 Brechas encontradas en la revisión sistemática

Luego de la elaboración del estado de conocimiento alrededor de la personalización y adaptación de MOOC, incluyendo la identificación del estilo de aprendizaje; a continuación, se resumen las brechas de conocimiento encontradas:

- No se encontró evidencia sobre mecanismos para la identificación automática del estilo de aprendizaje de un estudiante en función de su interacción con la plataforma en cursos en línea en ambientes de masividad con reconocimiento académico.
- No se halló referencias acerca de los resultados de la aplicación de distintos algoritmos de aprendizaje de máquina supervisados, según los datos que se puedan conseguir de la interacción de estudiantes en cursos en línea en ambientes de masividad con reconocimiento académico
- No existe un estudio, en el contexto de los MOOC, SPOC o MPOC, que permita establecer ventajas y desventajas del uso de distintos estilos de aprendizaje para la adaptación de ellos.
- No se consiguen conjuntos de datos sobre MOOC o MPOC libres que incluyan el estilo de aprendizaje de los estudiantes.

- No existe una aproximación metodológica clara para el diseño de MOOC que sean adaptables a un perfil del estudiante; la mayoría parten de un diseño único de un MOOC para adaptar bien sea la presentación de contenidos o la evaluación a algún indicador proporcionado por el estudiante.
- No hay estudios comparativos que indiquen las ventajas o desventajas de distintos mecanismos software empleados para implementar adaptabilidad en el desarrollo de MOOC o MPOC.
- No existe consenso sobre la caracterización de un estudiante en un MOOC que sirviera como referente universal para implementar la personalización de un curso masivo.
- Aunque muchos autores hablan de personalización en lugar de adaptación, ninguno resuelve el problema fundamental de atender necesidades individuales formativas a través de un MOOC.
- No hay claridad sobre la forma de articular arquitectónicamente las herramientas ya conocidas para la Analítica del Aprendizaje con las plataformas de software libre que soportan el ofrecimiento de MOOC.
- No se conocen estudios que indiquen el impacto en el estudiante del desarrollo de MOOC flexibles.

Ante la magnitud de las brechas detectadas, el presente trabajo de Doctorado se centrará en las cuatro primeras de la lista, aunque de igual forma puede abordar parcialmente las demás.

La hipótesis de trabajo de la presente investigación queda definida como: el entrenamiento de algoritmos de aprendizaje de máquina supervisados permitirá una adecuada identificación automática del estilo de aprendizaje de los estudiantes en cursos en línea masivo en ambientes de masividad y válidos por créditos académicos en educación superior.

Capítulo 3

3.1 Caso de estudio: Astronomía cotidiana

La alternativa seleccionada como caso de estudio, teniendo en cuenta la metodología (Yin, 2003) para la implementación del modelo MOOC dentro del Componente FISH de la Universidad del Cauca fue la de cursos en línea, privados y masivos, conocidos como MPOC. Un MPOC se define como un MOOC de menor alcance, dada que la aceptación para el ingreso se restringe al criterio de haber inscrito académicamente el curso, pero el número de admitidos es suficientemente alto en comparación con un curso tradicional lo cual también puede implicar la imposibilidad de realizar actividades de seguimiento y realimentación individualizadas.

Para la implementación del MPOC se reconvirtió a virtual el curso electivo FISH "Astronomía y sociedad" que se había ofrecido en modalidad presencial entre los años 2011 y 2013, que a su vez era una evolución del también curso electivo FISH "Fundamentos de Astronomía" ofrecido de manera presencial entre los años 2008 y 2009. La nueva asignatura, denominada "Astronomía cotidiana" se ofreció con un cupo de 400 estudiantes, 10 veces más estudiantes que un curso presencial tradicional.

Para responder a la pregunta ¿Es equitativo el modelo MOOC respecto a la evaluación según las preferencias de aprendizaje de los estudiantes? se definió un esquema de evaluación donde se harían exámenes en línea semanales de las temáticas vistas, así como también la aplicación en línea de un test de Kolb para la identificación de las preferencias de aprendizaje de los estudiantes.

En resumen, el Modelo de Estilos de Aprendizaje de David Kolb conocida como Teoría del Aprendizaje Experiencial (Kolb, 2014) afirma que el aprendizaje es el resultado de la forma como las personas perciben y luego procesan lo que han percibido. Identificó además que algunas personas prefieren percibir a través de la experiencia concreta (EC) mientras otras prefieren la conceptualización abstracta y las generalizaciones (CA). También que algunas personas prefieren procesar a través de la experimentación activa (EA) mientras otras lo hacen a través de la observación reflexiva (OR). La yuxtaposición de las formas de percibir y de procesar crea un modelo de cuatro cuadrantes, que a su vez caracteriza cada uno de los estilos de aprendizaje propuestos por Kolb: Acomodador (EC-EA), Asimilador (CA-OR), Convergente (EC-OR), Divergente (CA-EA).

Por el algoritmo definido para la definición de los estilos, es posible que una persona presente una combinación de estilos o presente características de todos ellos simultáneamente sin ninguna preferencia en particular, de tal forma que un estudiante puede pertenecer a uno de nueve grupos diferentes. El diseño del MPOC fue orientado por las directrices del proyecto MOOC-Maker (Alario-Hoyos, C., Delgado Kloos, C., & Rayyan, S. 2016), teniendo además en consideración las características de un buen aprendizaje definiendo actividades que fomenten reflexión; estimulen y faciliten el diálogo; posibilite la aplicación de conceptos teóricos mediante el desarrollo de prácticas; estimulen la creatividad generando preguntas y la búsqueda de respuestas del propio estudiantado (Sánchez & Prendes, 2015), específicamente en el primer diseño, no se tuvieron en cuenta los estilos de aprendizaje.

El curso se organizó a través de tres unidades temáticas, cada una de ellas compuesta a su vez por cinco temas, para un total 15 (de los dos últimos no son evaluables). Cada tema se desarrolló en una semana calendario que se evaluó de manera virtual con cuestionarios de preguntas de selección múltiple con única respuesta correcta y con varias respuestas correctas. Cada cuestionario constaba de diez preguntas, se habilitaba cada semana a la misma hora y durante 60 minutos para minimizar las posibilidades de fraude o filtración de respuestas. Para motivar tanto la recolección de información como el uso de determinados servicios de la aplicación web donde se alojó el curso, actividades como el diligenciamiento de encuestas, test y foros se califican usando el criterio de la participación y se les da un porcentaje (bajo) en cada nota que se sube al sistema de información respectivo. Para efectos de la presente investigación, sólo se tuvieron en cuenta los resultados de los exámenes en línea. Dadas las dificultades con la técnica de evaluación por pares (Conole, 2014), no se definieron actividades en este sentido. Los objetivos de cada tema semanal se desarrollaron a través de videos elaborados por el profesor del curso, las dispositivas de clase, recursos web videos complementarios (no elaborados en el curso, pero que complementan los contenidos propios) y al menos un foro de discusión.

Dependiendo de la naturaleza de la temática las unidades se tuvieron aplicaciones web interactivas para la realización de simulaciones y prácticas, y talleres que implicaron la construcción manual de modelos, experimentos e instrumentos a manera de trabajo independiente de los estudiantes. En los exámenes en línea semanales se preguntaron aspectos asociados a cada uno de dichos recursos didácticos.

Por la naturaleza de la asignatura y la filosofía del Componente FISH, el diseño de los exámenes en línea se centró en comprobar el dominio de habilidades de conocimiento, comprensión y aplicación de conocimiento de la Taxonomía de Bloom (Castleberry, & Brandt, 2016).

Como parte del caso de estudio, se instaló una instancia de Open edX en un servidor propio de la Universidad del Cauca bajo el nombre "Selene". Dado que aún no está integrado al Sistema de Información institucional, la creación de cuentas de usuario hizo de forma manual a partir del listado oficial de estudiantes inscritos. De igual forma, se desarrollaron algunas aplicaciones para facilitar el seguimiento de las actividades de aprendizaje que no implementa la versión de Open edX instalada (Jaramillo, & Solarte, 2016).

Dadas las características el curso: asignatura con reconocimiento académico (alto nivel de alineación con el currículo) y un apoyo medio a nivel institucional (el uso de una máquina

servidora no en las mejores condiciones técnicas y un talento humano escaso en número), el MPOC en cuestión queda a mitad de camino entre un “MOOC como reemplazo” y un “MOOC como un modelo orientador” según la clasificación propuesta en (Alario et al. 2014), siendo un esquema relativamente fácil de replicar en otras experiencias formativas.

A pesar de contar con recursos educativos de los cursos previos, la elaboración de los videos representó en esfuerzo significativo para el profesor del curso, dado que la Universidad del Cauca no dispone de una unidad para la producción de contenidos especializada y dedicada al cuerpo profesoral. Por ello se integró un ambiente basado en servicios web (Jaramillo, D. & Solarte, M., 2016) que facilitó el desarrollo de los contenidos en curso en un tiempo aceptable para el desarrollo de la experiencia. La evaluación de la experiencia, desde el punto de vista de los estudiantes fue altamente positiva (Solarte & Ramírez, 2016).

La Tabla 1 muestra los cursos en modalidad virtual y con reconocimiento académico que se han ofrecido en la Universidad del Cauca a partir del entorno creado por el caso de estudio de la presente investigación y el número de estudiantes que se inscribieron, no es el mismo número de quienes concluyeron los cursos. Con más de 4000 inscritos, aproximadamente la mitad ellos nunca habían tenido la oportunidad de participar y/o matricular en un curso de estas características.

	1-2016	2-2016	1-2017	2-2017	1-2018	2-2018
Astronomía cotidiana	433	428	517			
Comprensión y producción de textos argumentativos	105					
Comprensión y de textos argumentativos		110	109	103	97	102
Introducción a los drones			133	101	106	
Introducción a la edición de textos científicos y literarios en LaTeX		102	99	104	100	120
Introducción al emprendimiento con Lean Startup				265	317	270
La danza folclórica como patrimonio cultural						172

Tabla 1. Cursos virtuales con reconocimiento académico en la Universidad del Cauca

3.2 Extracción de datos de una instancia de Open EdX

El diseño de un mecanismo para el seguimiento de las actividades de aprendizaje de estudiantes en cursos en línea, abiertos y masivos se realizó teniendo en cuenta la arquitectura y funcionamiento de la plataforma Open edX. Para describir la arquitectura del prototipo a implementar se optó por utilizar el modelo de 4+1 vistas, ya que permite representar de forma estándar la arquitectura a través de diagramas UML.

A continuación, se describe la vista de escenarios, la vista lógica y la vista de despliegue, suficientes para describir la herramienta a construir.

3.2.1 Vista de escenarios

La descripción de la arquitectura en esta vista se hace mediante Diagramas de Casos de uso a partir de los cuales se relacionan las cuatro vistas restantes. La vista de Escenarios es obligatoria cuando se utiliza el modelo 4+1 vistas (Kruchten, Obbink, & Stafford, 2006) ya que todos los elementos de la arquitectura se derivan de los requerimientos que ahí se presentan. En la Figura 1 se presenta el Diagrama de Casos de Uso que orienta el diseño arquitectónico de la herramienta.

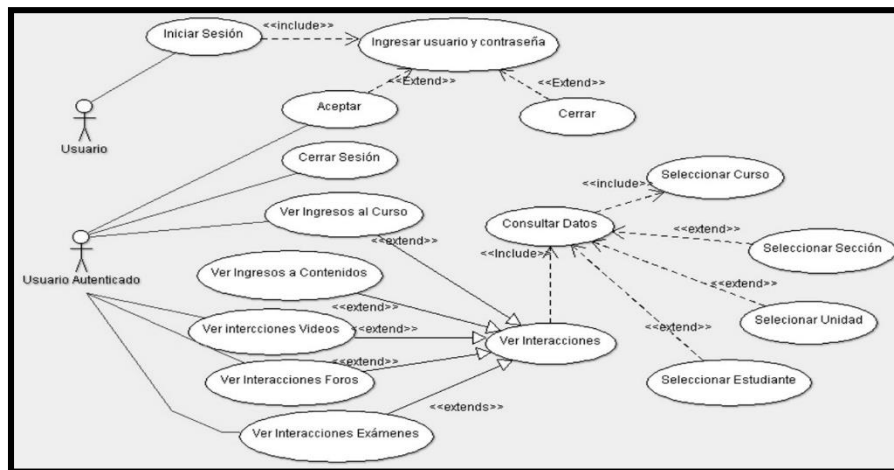


Figura 1. Diagrama de casos de uso

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

3.2.2 Vista lógica

En esta vista se representa la funcionalidad que el sistema proporcionará a los usuarios finales. Es decir, lo que el sistema debe hacer, y las funciones y servicios que ofrece. La representación se realiza mediante Diagramas de Clases. En la Figura 2 se presenta a continuación un diagrama general para el mecanismo de seguimiento, de manera que permita entender cómo funciona el mecanismo de una manera más simple.

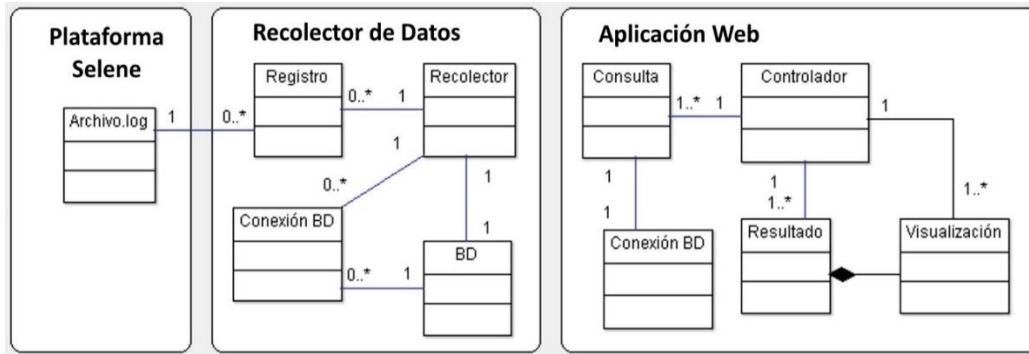


Figura 2. Diagrama de clases para el mecanismo de seguimiento

3.2.3 Vista de despliegue

En esta vista se muestra todos los componentes físicos del sistema, así como las conexiones que conforman la solución (incluyendo los servicios). En la Figura 3 se muestra el Diagrama de Despliegue del mecanismo para el seguimiento.

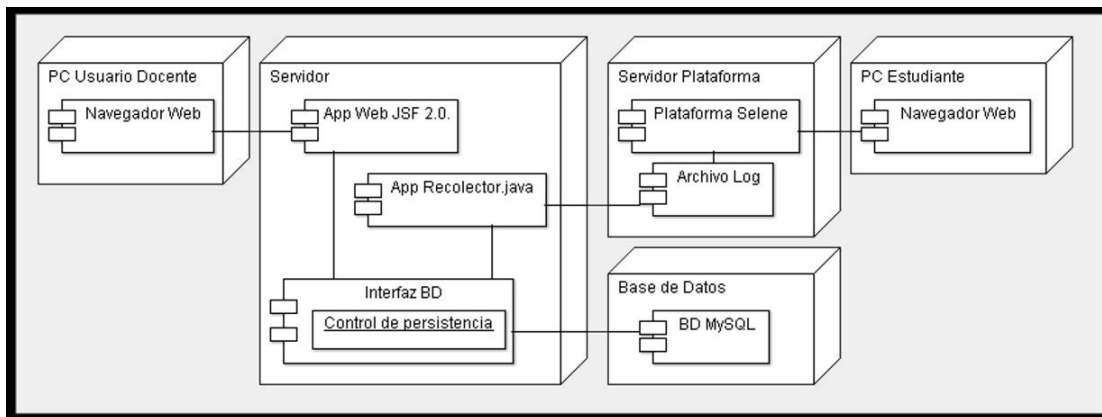


Figura 3. Diagrama de Despliegue

En la Figura 4 se presenta la arquitectura de implementación del mecanismo a construir. A continuación, se describe cada uno de los componentes arquitectónicos y su relación con los demás elementos dentro de la arquitectura.

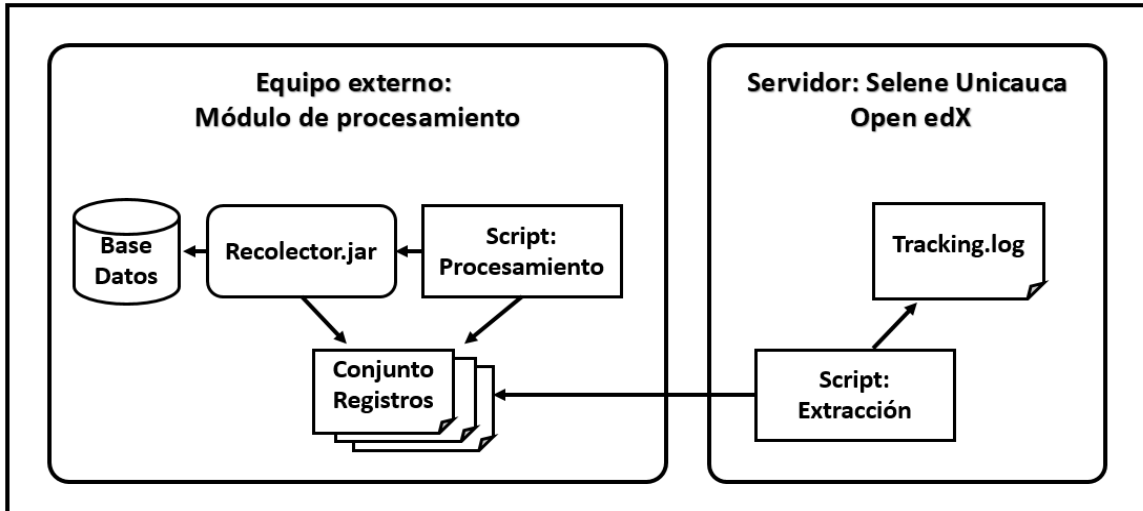


Figura 4. Arquitectura general mecanismo propuesto

Tracking.log

La plataforma Selene se despliega sobre un equipo servidor con una versión de Ubuntu server 14.04 y es una instancia de Open edX. En la plataforma se genera un archivo de registro llamado tracking.log. En este archivo se registran todas las interacciones de los estudiantes con el entorno de aprendizaje Selene, tales como: inicio de sesión, ingreso en los cursos, navegación en contenidos, presentación de foros e información de la interacción con las actividades evaluativas. La ubicación del archivo tracking.log dentro de la instancia es: /edx/var/log/tracking/. Aquí periódicamente dependiendo de la cantidad de datos de interacción generados se crea un nuevo archivo de registro y se comprime y almacena el anterior dejando registro de las interacciones en plataforma desde que ésta es dada de alta.

Script: Extracción

Los scripts son un conjunto de instrucciones generalmente almacenadas en un archivo de texto que deben ser interpretados línea a línea en tiempo real para su ejecución. El script de extracción lo que contiene son los comandos utilizados para realizar una copia de toda la carpeta que contiene los registros tracking.log de la plataforma y una línea de sincronización de dicha carpeta en un equipo externo. De esta manera todos los registros pueden ser procesados en tiempo real. En la Figura 5 se muestra el script ejecutado cada 5 minutos.

```

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

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

```

Figura 5. Script de extracción

Script: Procesamiento

Una vez con todos los registros disponibles en el equipo externo, el script de procesamiento captura el archivo tracking.log que es el archivo que tiene la información de las interacciones en tiempo real, realiza una copia del mismo, lo ubica en una dirección específica en donde el Recolector.jar puede procesarlo y repite esta operación periódicamente cada 5 minutos a través del programa Cron¹ presente en Ubuntu. De esta manera los datos procesados de la interacción pueden ser utilizados para realizar seguimiento a los estudiantes en todo momento. Para obtener los datos de eventos pasados, el script fue modificado a fin de que descomprimiera uno por uno los archivos de registro y ordenara la ejecución del Recolector.jar para cada registro presente en la carpeta sincronizada, así se logró crear un Data set con datos de interacción de los estudiantes desde el momento en que se creó la plataforma. En la Figura 6 se muestra el script de procesamiento el cual es ejecutado cada 5 minutos.

```

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

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

echo "tracking procesado"

```

Figura 6. Script de procesamiento de registros

Recolector.jar:

Es un ejecutable creado en código java, que al ejecutarse busca el archivo tracking.log en una ubicación específica, lo lee tomando evento por evento presente en el archivo de registro, guarda en un buffer los eventos relacionados a las actividades que se desean capturar para luego procesarlos, obtener la información de manera ordenada y entendible para un humano, y los guarda en una base de datos (MySQL). El archivo tracking está escrito en Json, por lo que el recolector debe interpretar dicha información de manera adecuada. En la Figura 7 se muestra un ejemplo de los eventos registrados en el archivo tracking.log.

```

{"username": "Luis [REDACTED]", "event source": "browser", "name": "play_video", "acces
pt_language": "es-ES,es;q=0.9", "time": "2018-09-16T18:56:06.506539+00:00", "agent": "Mozi
lla/5.0 (Windows NT 10.0; Win64; x64) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/68.0.3
440.106 Safari/537.36", "page": "http://selene.unicauca.edu.co/courses/course-v1:Unicauca+
LaTEX_Fish+2018-II/courseware/714c09237b02479285423321dd8f326f/779d9398f56f4f6a9d931df0c31
50bb7/", "host": "selene.unicauca.edu.co", "session": "ee9e0282fa511c756ebd8a2aab6cec6c",
"referer": "http://selene.unicauca.edu.co/courses/course-v1:Unicauca+LaTEX_Fish+2018-II/co
urseware/714c09237b02479285423321dd8f326f/779d9398f56f4f6a9d931df0c3150bb7/", "context": {
"user_id": 455, "org_id": "Unicauca", "course_id": "course-v1:Unicauca+LaTEX_Fish+2018-II",
"path": "/event"}, "ip": "10.0.2.2", "event": "{\"code\": \"1ZNPbYVCNMg\", \"id\": \"2bb
f8ee0e3f948e9a992ecf87e8247a1\", \"currentTime\": 7.006137015735626}", "event_type": "play
video"}

```

Figura 7. Evento del archivo tracking.log

Dentro de la línea de registro se guarda toda la información necesaria para el seguimiento a las actividades. Para el ejemplo, la pareja “name:play_video”, identifica que un estudiante reprodujo un video dentro de la plataforma de aprendizaje, incluyendo información del id del estudiante, curso, fecha, sección y subsección en el contenido del curso, etc. De esta manera es posible capturar los eventos de interacción del estudiante en cuanto a acceso, contenidos, recursos, foros y evaluaciones.

¹ Cron es un administrador regular de procesos en segundo plano que ejecuta procesos o scripts a intervalos regulares de tiempo (por ejemplo, cada minuto, día, semana o mes).

3.2.4 Resultados

A partir de la ejecución del mecanismo con los archivos de registro de Selene, se logró la construcción de un *dataset*. Aquí se encuentran los datos de interacción de cuatro cursos que han sido ofrecidos en la Universidad del Cauca y han sido reconocidos académicamente.

A continuación, se muestra un ejemplo de consulta realizada en la herramienta de seguimiento. En la Figura 8 se muestra una consulta de ingresos en un curso en la herramienta Selene. Se tiene 1523 ingresos en el curso de Astronomía cotidiana para el mes de marzo, provenientes de un total de 344 estudiantes que estuvieron activos en dicho periodo.

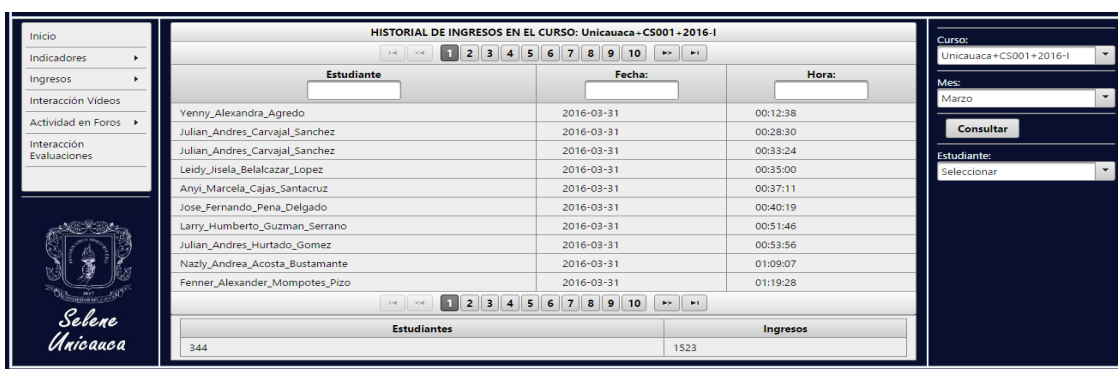


Figura 8. Captura Herramienta Selene. Historial de ingresos en el curso piloto

Es posible ordenar los resultados por el número de ingresos en el curso, número de ingresos en los contenidos del curso, número de videos reproducidos, participación en los foros y número de evaluaciones realizadas.

Uno de los cursos que mayor participación ha logrado es el curso de “Astronomía Cotidiana”, que para el primer semestre del 2017 logro obtener 517 estudiantes inscritos en un mismo curso, más de 10 veces la cantidad de inscritos que se logran con un curso presencial convencional en la Universidad del Cauca, catalogándolo como un MPOC por la característica de ser privado. Con el fin de contribuir al entendimiento de los comportamientos de los estudiantes en los MOOC, se obtuvo desde la base de datos las interacciones de los estudiantes del curso de “Astronomía Cotidiana” para el primer semestre del 2016. Para el análisis de los datos, se generó un archivo csv mediante una consulta a la base de datos MySQL. La figura 9 se muestra una captura de pantalla del csv analizado.

7242	Esteban	Unicauca+As	14:23:40	7/02/2017	a8a9d8f5663	pause_vid	30ad398096a	f2f154481e4	sDuy0Nf2Cwg
7243	Esteban	Unicauca+As	14:23:40	7/02/2017	a8a9d8f5663	stop_video	30ad398096a	f2f154481e4	sDuy0Nf2Cwg
7244	Jhonatan	Unicauca+As	14:24:03	7/02/2017	a64eb97b2fc	pause_vid	30ad398096a	f2f154481e4	ZgVJKrcD-kc
7245	Esteban	Unicauca+As	14:24:51	7/02/2017	a8a9d8f5663	pause_vid	30ad398096a	f2f154481e4	ZgVJKrcD-kc
7246	Esteban	Unicauca+As	14:24:59	7/02/2017	a8a9d8f5663	pause_vid	30ad398096a	f2f154481e4	ZgVJKrcD-kc
7247	Jhonatan	Unicauca+As	14:25:41	7/02/2017	a64eb97b2fc	pause_vid	30ad398096a	f2f154481e4	ZgVJKrcD-kc
7252	Jhonatan	Unicauca+As	14:32:31	7/02/2017	a64eb97b2fc	pause_vid	30ad398096a	f2f154481e4	ZgVJKrcD-kc

Figura 9. Eventos en el csv obtenido desde Selene

Con el csv se realizó un pequeño análisis estadístico que sirvió de insumo para el seguimiento que el docente realizaba sobre los estudiantes. En la Figura 10 se muestra los

datos del comportamiento del curso relacionado a la cantidad de interacciones realizadas por los estudiantes a lo largo del curso con los contenidos, videos, foros y exámenes. Es claro como los estudiantes en un MPOC con reconocimiento académico se comportan en torno a las actividades evaluativas. Como se muestra en la imagen, los estudiantes tienen grandes picos de interacción en las fechas en las cuales fueron programados los exámenes.

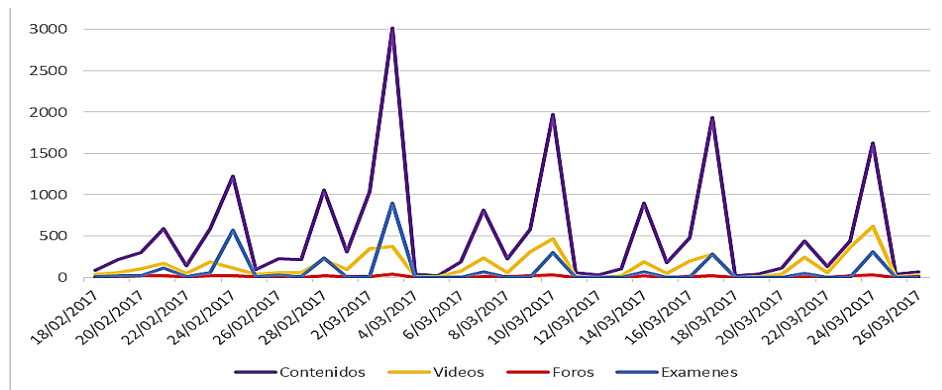


Figura 10. Historial de las interacciones del curso virtual de Astronomía Cotidiana (imagen ajustada al primer periodo académico de 2018)

3.3 Prueba con datos reales

El seguimiento se hizo al curso virtual “Astronomía Cotidiana”, de la Universidad del Cauca, ofrecido en modalidad MPOC como electiva del componente de Formación Integral Social y Humana en la plataforma Selene.

El curso Astronomía cotidiana tiene como propósito introducir a los estudiantes en los principios, métodos, y herramientas de la educación en línea para el estudio de conceptos básicos de la Astronomía y su relación con otras ciencias de la tierra y el espacio. El curso se organizó en tres temas principales y por cada uno de ellos los estudiantes debieron realizar una evaluación. Se tuvo como duración un periodo de 18 semanas comenzando el 15 de febrero y terminando el 10 de junio de 2016 y contó con una participación de 403 estudiantes, diez veces la cantidad de estudiantes tradicionalmente admitida en cursos regulares con reconocimiento académico.

El prototipo implementado permite obtener información sobre el comportamiento del curso casi en tiempo real. A continuación, se describe los resultados obtenidos y organizados acorde a los indicadores deseados.

3.3.1 Ingresos en el curso y sus contenidos

Una de las inquietudes del docente fue conocer los ingresos de los estudiantes los primeros días del curso, para saber si el procedimiento definido para inscripción de estudiantes a la plataforma era el correcto y si los estudiantes habían recibido las comunicaciones orientadoras al comienzo de la asignatura. En la Figura 11 se muestran los ingresos al curso para el mes de febrero, mes en que dio inicio el curso piloto. Como se observa, se registraron 1508 ingresos en el curso y el número de estudiantes fue 401.

HISTORIAL DE INGRESOS EN EL CURSO: Unicauca + CS001 - 2016-I		
<input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/> <input type="text"/>		
Estudiante	Fecha:	Hora:
Diana_Isabel_Paz_Pino	2016-02-29	00:03:54
Kevin_Steven_Castillo_Solarte	2016-02-29	00:12:14
Oscar_Ivan_Marcillo_Eche	2016-02-29	00:20:05
Nancy_Yamileth_Cifuentes	2016-02-29	00:26:06
Diego_Alejandro_Vezez_Mosquera	2016-02-29	00:28:19
Luis_David_Medina_Sandoval	2016-02-29	00:53:28
Alvaro_Javier_Manrique_Zuniga	2016-02-29	01:04:39
Wesnert_Alegria_Riascos	2016-02-29	01:32:05
Alvaro_Javier_Manrique_Zuniga	2016-02-29	01:35:08
Dajaramillo	2016-02-29	01:45:00
Estudiantes		Ingresos
401		1508

Figura 11. Ingresos al curso piloto mes de febrero

La Figura 12 muestra el historial de ingresos obtenidos para el mes de febrero. En la semana 1 (día 15 al 19) es donde se presentan mayores ingresos, debido a la expectativa del curso. Se ve que para fechas posteriores los ingresos disminuyen y se mantienen. De la misma forma la herramienta permite ver los contenidos que se han consultado para fechas específicas.



Figura 12. Historial de ingresos al curso piloto mes de febrero

3.3.2 Actividad en videos

En cuanto a las interacciones de los estudiantes con los videos la plataforma permite ver la información de cada video y también de cada estudiante. La pestaña “Interacción en Videos” en el menú de navegación es el acceso a las dichas consultas.

En la Figura 13 se observa la consulta hecha al finalizar el curso para el video “GQQg83MSuvY” (identificador del video en Youtube), pero además se puede ver la sección y unidad de ubicación del video dentro del curso.

CURSO: Unicauca + CS001 + 2016-I						
Estudiante	Video:	Acción:	Fecha:	Hora:		
Laura_Maritza_Hernandez_Lasso	GQQg83MSuvY	stop_video	2016-04-07	14:12:21	Unidad	
Hernan_Castillo_Kevin	GQQg83MSuvY	stop_video	2016-04-02	00:21:01	Unidad	
Angellin_Katherine_Nieto_Ley	GQQg83MSuvY	play_video	2016-04-02	04:49:30	Unidad	
Angellin_Katherine_Nieto_Ley	GQQg83MSuvY	play_video	2016-04-02	05:53:25	Unidad	
Ludwing_Nicole_Palomino	GQQg83MSuvY	play_video	2016-04-01	00:57:09	Unidad	
Jhon_Jairo_Cordoba_Ayala	GQQg83MSuvY	play_video	2016-04-01	01:08:41	Unidad	
Angie_Dayana_Palechor	GQQg83MSuvY	play_video	2016-04-01	01:54:46	Unidad	
Cristhian_Sebastian_Bolanos	GQQg83MSuvY	play_video	2016-04-01	02:49:25	Unidad	
Diego_Alejandro_Martinez	GQQg83MSuvY	play_video	2016-04-01	02:51:33	Unidad	
Diego_Alejandro_Martinez	GQQg83MSuvY	stop_video	2016-04-01	02:59:45	Unidad	

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

Figura 13. Consulta interacciones de los estudiantes con los videos del curso piloto

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

Además, es posible la consulta de las interacciones de los videos por cada estudiante. En la Figura 14 se muestra un ejemplo donde un estudiante, registró 226 interacciones que corresponden a 127 reproducciones de las cuales 99 veces llego hasta el final del video. Miró 106 videos diferentes, de los cuales 90 fueron observados en su totalidad. También la herramienta entrega la lista de los diferentes videos vistos por el estudiante. La consulta fue realizada al finalizar el curso.

CURSO: Unicauca + CS001 + 2016-I						
Estudiante	Video:	Acción:	Fecha:	Hora:		
Adrian_Felipe_Vargas_Arias	9vk6Kn2VXhQ	play_video	2016-03-13	01:23:30	Unidad	
Adrian_Felipe_Vargas_Arias	9vk6Kn2VXhQ	play_video	2016-03-13	01:23:38	Unidad	
Adrian_Felipe_Vargas_Arias	9vk6Kn2VXhQ	stop_video	2016-03-13	01:26:06	Unidad	
Adrian_Felipe_Vargas_Arias	iFe8JULpkA	play_video	2016-03-13	01:26:07	Unidad	
Adrian_Felipe_Vargas_Arias	LoSOZ9r_UvU	play_video	2016-03-13	01:33:58	Unidad	
Adrian_Felipe_Vargas_Arias	xluMl0YWm4E	play_video	2016-03-13	01:34:03	Unidad	
Adrian_Felipe_Vargas_Arias	LoSOZ9r_UvU	play_video	2016-03-13	01:53:34	Unidad	
Adrian_Felipe_Vargas_Arias	LoSOZ9r_UvU	stop_video	2016-03-13	02:04:34	Unidad	
Adrian_Felipe_Vargas_Arias	Z2yCHMBFNyM	play_video	2016-03-13	02:35:51	Unidad	
Adrian_Felipe_Vargas_Arias	xluMl0YWm4E	stop_video	2016-03-13	02:37:06	Unidad	

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

Figura 14. Consulta de las interacciones de un estudiante específico con videos

3.3.3 Actividad en Foros

En cuanto a los foros se puede obtener el total de las participaciones de los estudiantes, dichas participaciones corresponden a creaciones de temas nuevos (“hilos”), respuestas en dichos temas y comentarios de las mismas del curso en general, además, de la misma forma que para los otros indicadores, es posible hacer consultas por cada estudiante. En la

Figura 15 se muestra los datos de las participaciones del curso piloto en donde se contó con la participación de 330 estudiantes. Consulta realizada al final del curso.

CURSO: Unicaauaca+ CS001 - 2016-I				
1 2 3 4 5 6 7 8 9 10				
Estudiante	Foro:	Fecha:	Hora:	
John_Edinson_Benavides_Clavijo	edx.forum.thread.created	2016-06-16	00:22:59	Aplicaciones práctic
John_Edinson_Benavides_Clavijo	edx.forum.thread.created	2016-06-16	00:28:44	Recurso de astron
John_Edinson_Benavides_Clavijo	edx.forum.thread.created	2016-06-16	01:21:57	Relación de progr
Jimmy_Jesid_Castro_Burbano	edx.forum.thread.created	2016-06-16	03:06:39	CALENTAMIENTO
Nazly_Andrea_Acosta_Bustamante	edx.forum.thread.created	2016-06-16	05:00:03	, \u00bfrqu\u00e99
Diego_Alejandro_Martinez	edx.forum.thread.created	2016-06-16	05:11:15	Calentamiento glo
Nazly_Andrea_Acosta_Bustamante	edx.forum.thread.created	2016-06-16	05:17:43	ARTE EN LA ASTR
Nazly_Andrea_Acosta_Bustamante	edx.forum.thread.created	2016-06-16	05:51:48	astronomia en ing
Daniel_Andres_Munoz_Salazar	edx.forum.thread.created	2016-06-16	15:34:31	peque\u00f1a ser
Lineth_Vanessa_Yanten_Zuniga	edx.forum.thread.created	2016-06-16	15:57:02	TURISMO ASTRON
1 2 3 4 5 6 7 8 9 10				
Estudiantes	Hilos	Respuestas	Comentarios	Participaciones
330	1340	1005	65	2410

Figura 15. Consulta de las interacciones en el foro del curso piloto

3.3.4 Interacciones con Evaluaciones

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

La característica de la cantidad de repeticiones se agregó a la herramienta debido a que Open edX permite configurar en los exámenes el número de intentos, es decir, el número de veces que un estudiante puede repetir un examen o ejercicio. En la Figura 16 se muestran los resultados obtenidos para el examen final del curso piloto. Se puede apreciar como de 403 estudiantes inscritos, 358 presentaron el examen final del curso, también se muestran el número de participaciones y cuantas veces fue repetido el examen.

CURSO: Unicaauca+CS001+2016-I				
Estudiante		Fecha:	Hora:	
Miguel_Eduardo_Mosquera		2016-06-16	18:05:06	
Jose_Ignacio_Romero_Urbano		2016-06-16	18:14:29	
Pablo_Albeyro_Gaviria_Bastidas		2016-06-16	18:16:13	
Eduardo_Jose_Ordenez_Hurtado		2016-06-16	18:16:40	
James_Gabriel_Zambrano_Mendez		2016-06-16	18:16:47	
Bayron_Leonardo_Usama_Noguera		2016-06-16	18:18:18	
David_Eduardo_Urrea_Galvis		2016-06-16	18:19:03	
Wesnert_Alegria_Riascos		2016-06-16	18:19:17	
Marlio_Alejandro_Chicue		2016-06-16	18:20:28	
Jenny_Carolina_Jojoa_Cifuentes		2016-06-16	18:21:11	
Estudiante	Intentos	Repeticiones	Unidades realizadas	Estudiantes
	400	42	1	358

Figura 16. Consulta de las interacciones en el examen final del curso piloto

3.3.5 Informes automáticos

Otra funcionalidad que se le incorporó a la herramienta fue la creación y envío de reportes en formato csv de manera automática. La herramienta recupera la información del curso directamente desde la base de datos y la exporta en seis archivos:

- Ingresos_Curso.csv
- Ingresos_Contenidos.csv
- Interacciones_Videos.csv
- Interacciones_Foros.csv
- Interacciones_Exámenes.csv
- Indicadores_compuestos.csv

La creación y envío de los archivos se hace mediante un script que se ejecuta periódicamente como una tarea “Cron” cada semana. En principio esta funcionalidad es estática y no es modificable, pero se espera para trabajos posteriores agregar a la aplicación web una interfaz que permita controlar su configuración. Las Figuras 13 a 23 fueron generadas a partir del último reporte entregado por la herramienta.

En la Figura 17 se observa el comportamiento de los estudiantes en cuanto a ingresos durante todo el curso. Se observa como los ingresos tienen datos altos en la semana de inicio en donde se atribuye el hecho a la expectativa de los estudiantes por el curso y que va disminuyendo con el tiempo. Por otra parte, para las semanas 6, 11, 12, 16 y 17 el incremento de ingresos corresponde a las semanas en donde se programó actividades evaluativas, este hecho es validado por los resultados que se muestran más adelante en las Figuras 18 y 19.

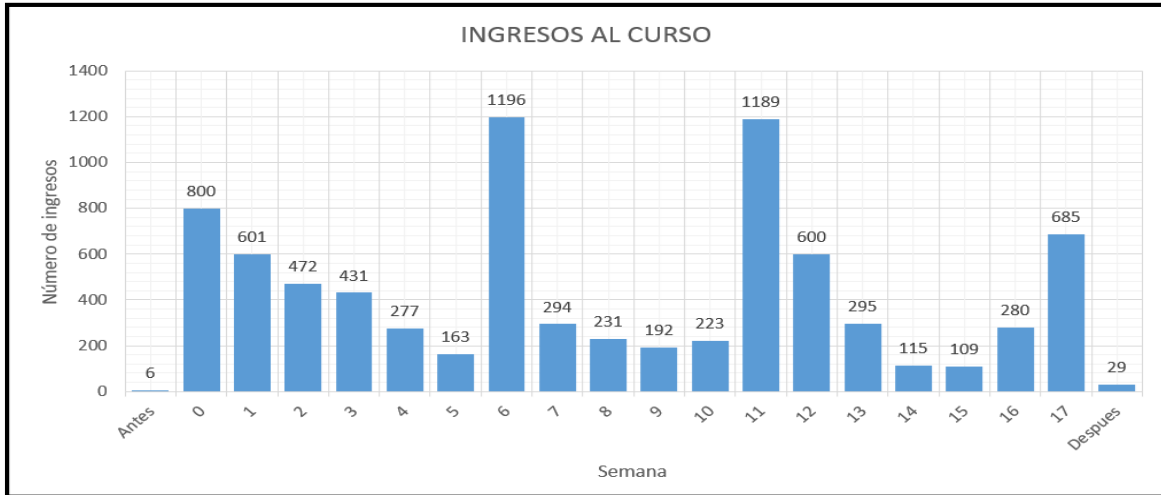


Figura 17. Ingresos por semana al Curso Virtual de Astronomía Cotidiana 2016-I

Por otra parte, en la Figura 18 se muestran los ingresos hechos en los contenidos del curso. Al igual que para la anterior gráfica, los resultados se incrementan justo en las semanas en donde se programó actividades evaluativas. Sin embargo, se ve como las interacciones de los estudiantes en las primeras semanas es considerable y constante, por lo menos hasta cubrir los temas que abarcaron el primer examen realizado.

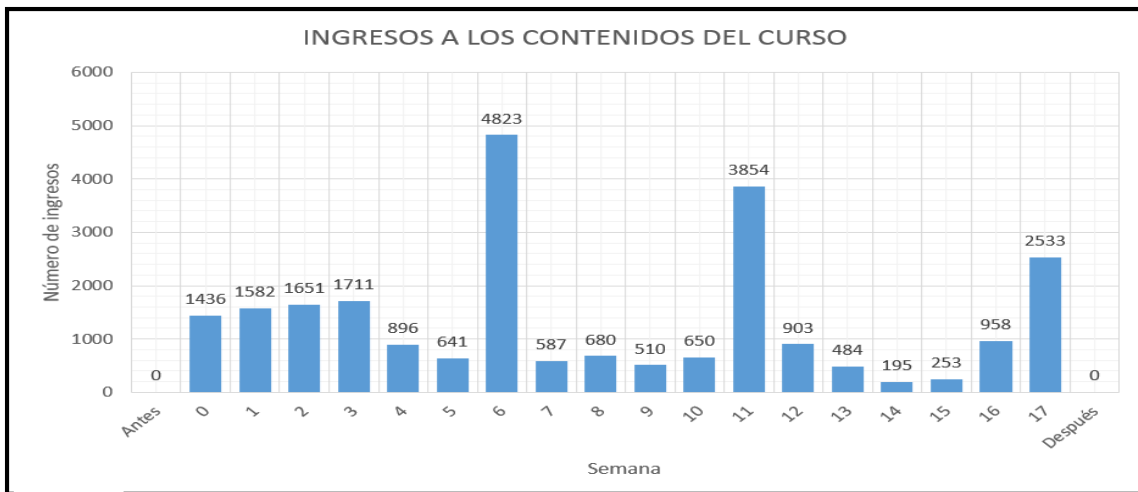


Figura 18. Ingresos por semana a los contenidos del curso piloto

Otra forma de mostrar los datos se presenta en la Figura 19, en donde se grafica la cantidad de ingresos por cada una de las unidades y también tiene una relación directa con las evaluaciones del curso, pues la unidad más visitada es la de Calificaciones.

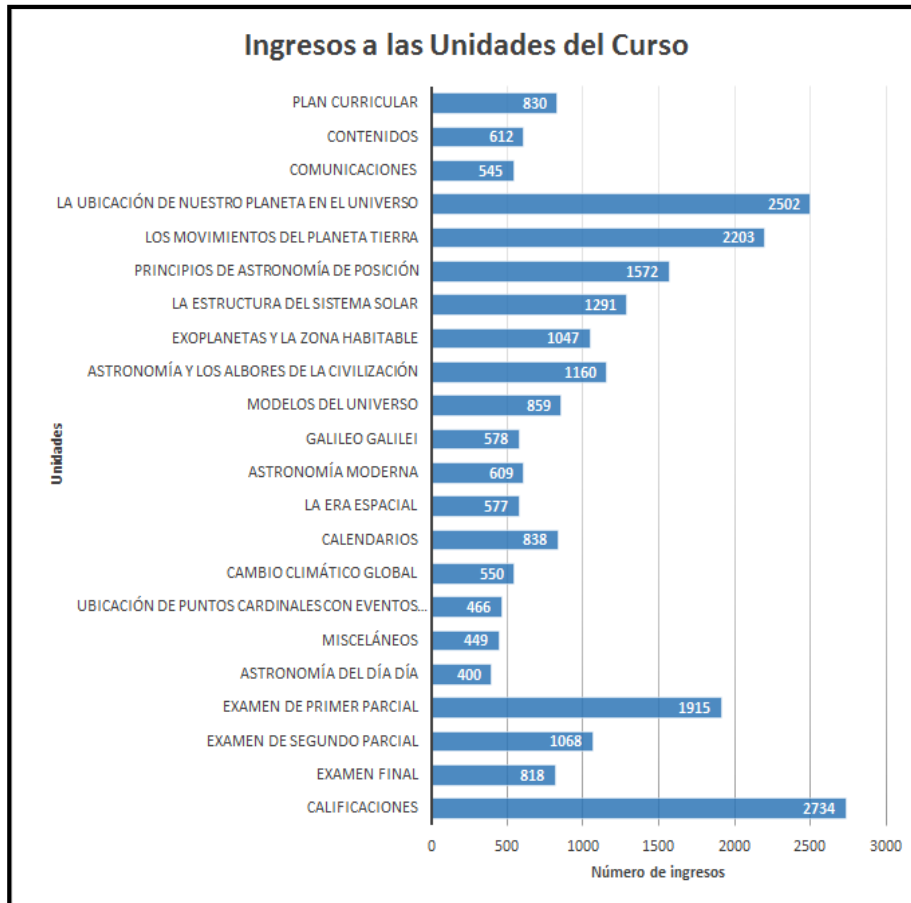


Figura 19. Cantidad de ingresos a cada una de las unidades del curso

En la Figura 20 se muestra la cantidad de reproducciones de los videos de cada unidad que forma parte del contenido. En la gráfica se muestra una gran cantidad de reproducciones y se aprecia como los resultados obtenidos de los videos vistos no contrastan demasiado con los resultados de los videos vistos completamente. Sin embargo, se muestra como los estudiantes van perdiendo el interés con el tiempo, pues las unidades se encuentran ordenadas según los tiempos de presentación en el curso.

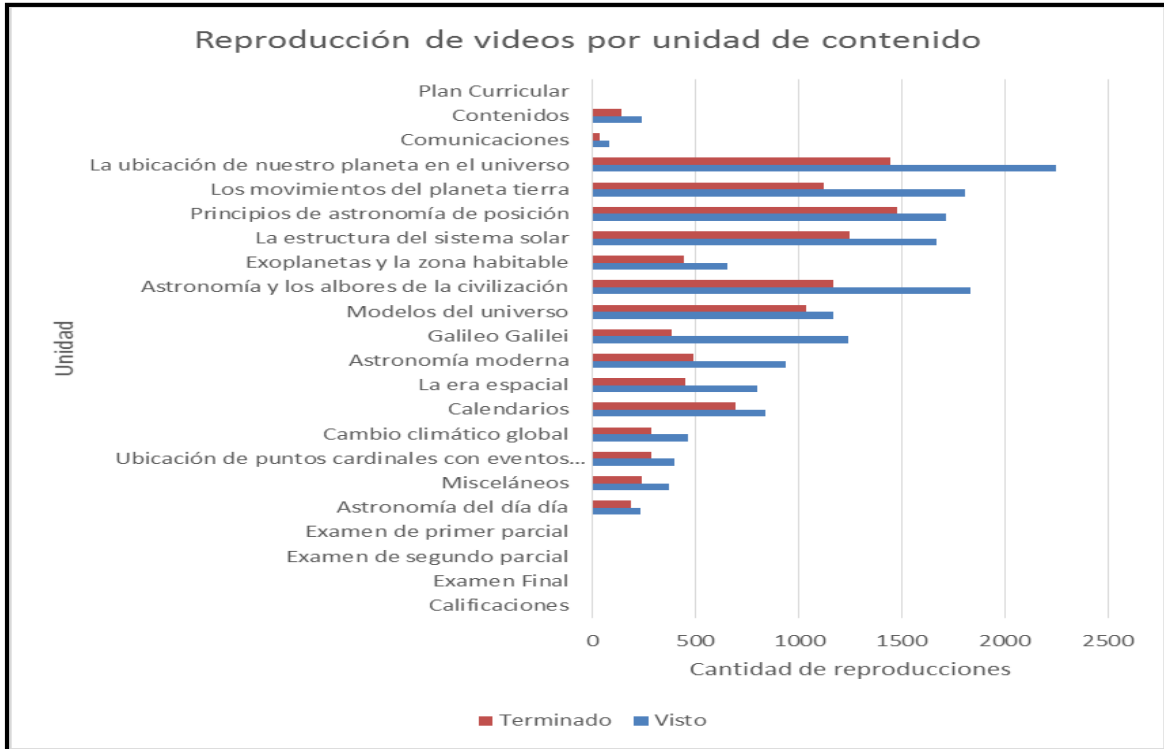


Figura 20. Cantidad de reproducciones de los videos por cada unidad del curso

En la Figura 21 se muestra las participaciones de los estudiantes en los foros graficada por cada semana de ejecución del curso. Al igual que para los resultados obtenidos con los videos, se ve una mayor participación en las semanas de inicio del curso, teniendo una baja participación en las semanas finales.

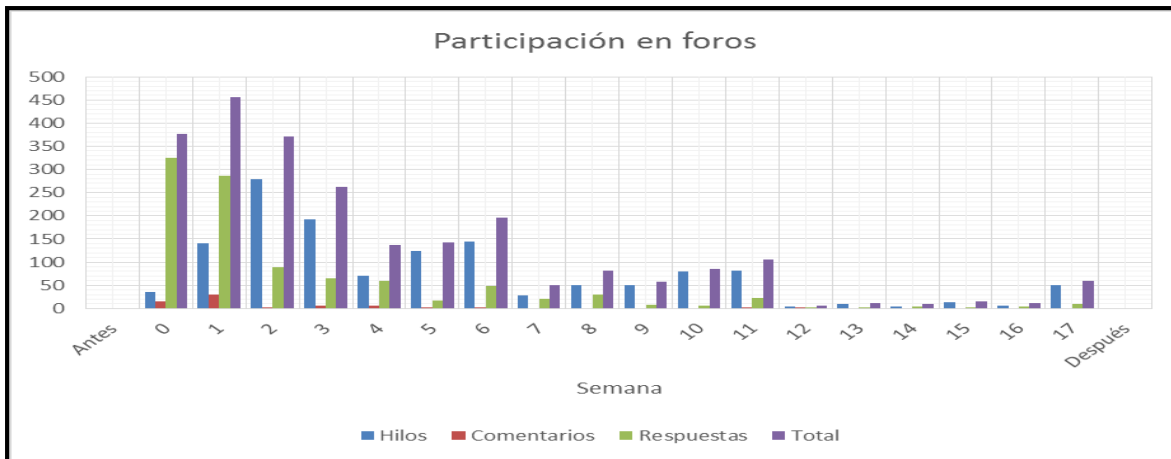


Figura 21. Participación en los foros por semana

Otro de los resultados obtenidos de los reportes entregados por la herramienta es la participación en las evaluaciones. La Figura 22 muestra la cantidad de estudiantes que presentaron las diferentes actividades evaluativas.

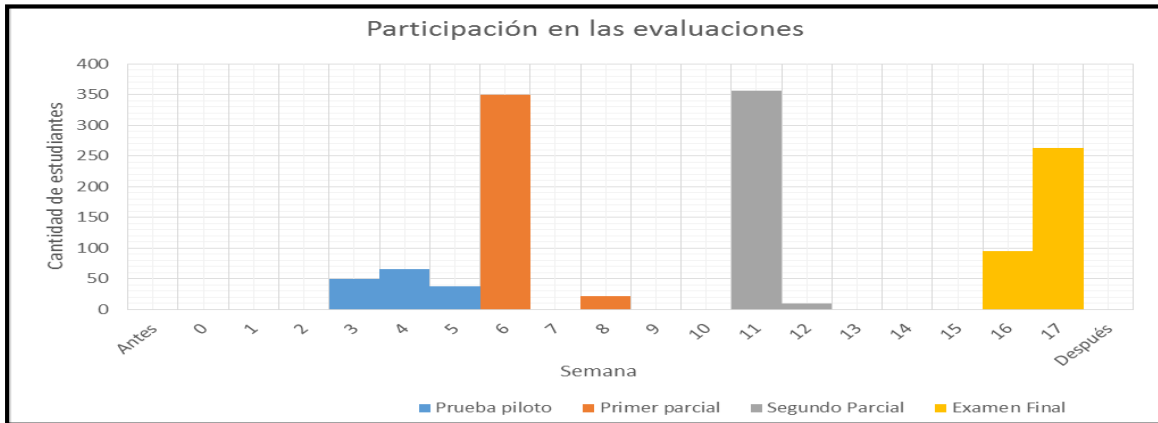


Figura 22. Participación en las evaluaciones por semana

Entre las semanas tres a cinco -marcadas en azul- se dejaron activos exámenes de autoevaluaciones de prueba, realizados con el fin de familiarizarse con el tipo de ejercicios y evaluación que ofrece Open edX. Por otra parte, en las semanas 8 y 12 se realizaron exámenes supletorios para los estudiantes que no pudieron realizarlos en las fechas estipulas. El examen final se podía presentar en la semana 16 o en la 17, la mayoría prefirió realizar el examen en el último plazo.

En la Figura 23 se muestra una gráfica comparativa de las interacciones de los estudiantes. Es claro que el comportamiento de los estudiantes se acomoda a las fechas de las evaluaciones.

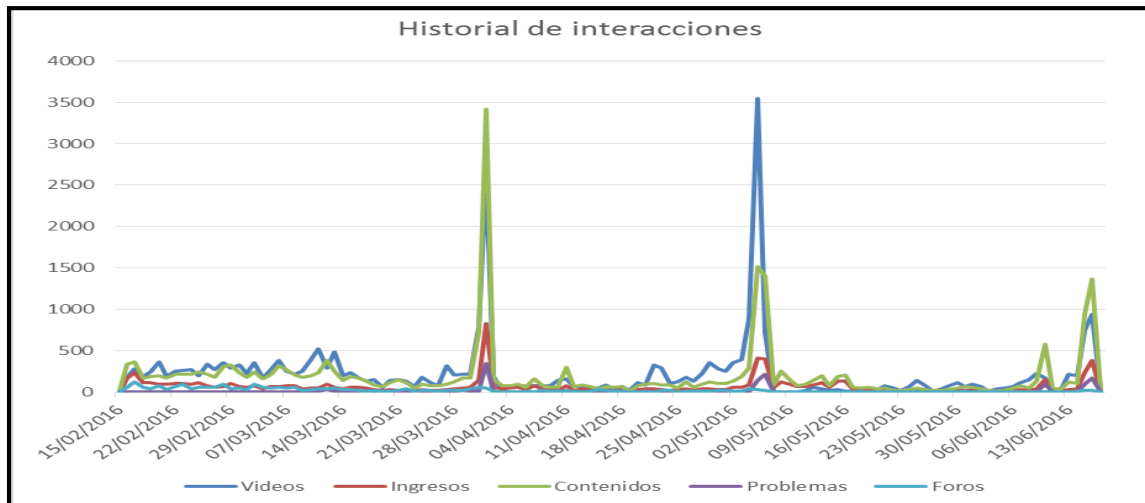


Figura 23. Gráfica comparativa de los resultados finales del curso piloto

Hubo casos en que se utilizó la herramienta para evitar problemas con algunas tácticas de los estudiantes para lograr presentar los exámenes por fuera de las fechas y horas estipulas, pues haber realizado alguna actividad (sobre todo evaluativas) y culpar a la plataforma tecnológica por no haber almacenado resultados, cuando en realidad los estudiantes ni siquiera habían ingresado en el curso.

Por otra parte, el cursó conto con una nota de participación en los foros, en esta tarea la herramienta ayudó ostensiblemente, pues permitió identificar cual fue la participación de los estudiantes en los foros.

3.4 Construcción de *dataset* (conjunto de datos)

La construcción de un dataset a partir de la información obtenida descrita en el apartado anterior, requiere una metodología adecuada a la naturaleza y propósitos de la investigación a realizar.

Desde la Minería de Datos, han surgido distintos enfoques para la elaboración datasets como KDD, SEMMA, CRISP-DM y Catalyst, de las cuales las dos últimas definen una serie de lineamientos para la elaboración de la construcción (Possos & Cruz, 2017).

CRISP-DM (además de ser de uso libre) es la metodología más utilizada actualmente en iniciativas de este tipo según se puede observar en la Figura 24, posición que ha mantenido desde el año 2007. Lo anterior posibilita encontrar una gran cantidad de información y referencia sobre su uso y se evidencia los resultados satisfactorios de su implementación.

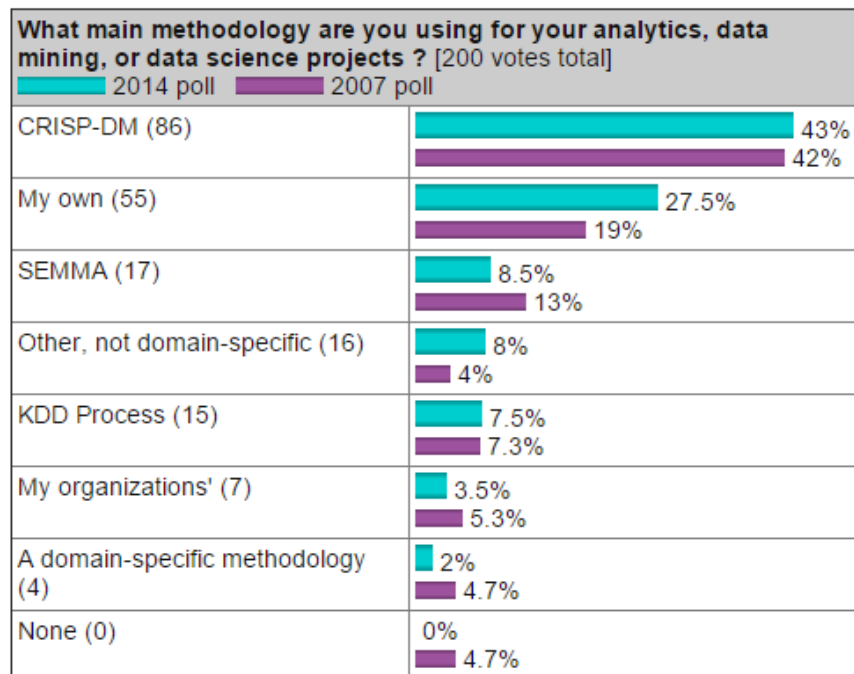


Figura 24. Metodologías de minería de datos. Fuente (Possos & Cruz, 2017)

La metodología CRISP-DM está descrita en términos de un modelo de proceso jerárquico, a través de en un conjunto de tareas descritas en cuatro niveles de abstracción (de lo general a lo específico):

- fase,
- tarea genérica,
- tarea especializada,

- instancia de proceso.

En el nivel superior, el proceso de minería de datos es organizado en seis fases. Cada fase consiste de varias tareas genéricas de segundo nivel. El segundo nivel es llamado genérico pues está destinado a ser bastante general para cubrir todas las situaciones posibles a explotar mediante técnica de minería de datos (o aprendizaje de máquina). Las tareas genéricas están destinadas a ser tan completas y estables como sea posible. *Completo* significa que cubre tanto al proceso entero de minería de datos y todas las aplicaciones de minería de datos posibles. *Estable* significa que el modelo debería ser válido para acontecimientos normales y aún para desarrollos imprevistos como técnicas de modelado nuevas.

El nivel de tarea especializado es el lugar para describir cómo deberían ser realizadas las acciones en las tareas genéricas dadas ciertas situaciones específicas. Por ejemplo, en el segundo nivel podría haber una tarea genérica llamada limpieza de datos. El tercer nivel describe cómo esta tarea se diferencia en situaciones diferentes, como la limpieza de valores numéricos contra la limpieza de valores categóricos, o si el tipo de problema es de agrupamiento o de modelado predictivo.

La descripción de fases y tareas como pasos discretos realizados en un orden específico representa una secuencia idealizada de eventos. En la práctica, muchas de las tareas pueden ser realizadas en un orden diferente, y a menudo será necesario volver a hacer tareas anteriores repetidamente, así como repetir ciertas acciones.

El nivel de la instancia de proceso, es un registro de las acciones, decisiones, y de los resultados de un trabajo real de minería de datos.

Según el modelo de referencia CRISP-DM, el ciclo de vida de un proyecto de minería de datos consiste en seis fases, como se muestra en la Figura 25. Estas no tienen una secuencia rígida y es comúnmente requerido el movimiento hacia adelante y hacia atrás entre fases diferentes.



Figura 25. Ciclo de vida CRISP-DM. Fuente (Wirth, & Hipp, 2000)

En proyectos de naturaleza académica, algunas actividades en cada fase no son estrictamente desarrolladas. Lo anterior es posible por la flexibilidad de la metodología.

Resumen de las fases y niveles de CRISP-DM:

Comprensión del negocio: Objetivos y requerimientos desde una perspectiva no técnica.
<ul style="list-style-type: none">▪ Establecimiento de los objetivos del negocio (Contexto inicial, objetivos, criterios de éxito)▪ Evaluación de la situación (Inventario de recursos, requerimientos, supuestos, terminologías propias del negocio)▪ Establecimiento de los objetivos de la minería de datos (objetivos y criterios de éxito)▪ Generación del plan del proyecto (plan, herramientas, equipo y técnicas)
Comprensión de los datos: Familiarizarse con los datos teniendo presente los objetivos del negocio.
<ul style="list-style-type: none">▪ Recopilación inicial de datos▪ Descripción de los datos▪ Exploración de los datos▪ Verificación de calidad de datos
Preparación de los datos: Obtener la vista minable o <i>dataset</i> .
<ul style="list-style-type: none">▪ Selección de los datos▪ Limpieza de datos▪ Construcción de datos▪ Integración de datos▪ Formateo de datos
Modelado: Aplicar las técnicas de minería de datos a los dataset.
<ul style="list-style-type: none">▪ Selección de la técnica de modelado▪ Diseño de la evaluación▪ Construcción del modelo▪ Evaluación del modelo
Evaluación: De los modelos de la fase anteriores para determinar si son útiles a las necesidades del negocio.
<ul style="list-style-type: none">▪ Evaluación de resultados▪ Revisar el proceso▪ Establecimiento de los siguientes pasos o acciones
Despliegue: Explotar utilidad de los modelos, integrándolos en las tareas de toma de decisiones de la organización.
<ul style="list-style-type: none">▪ Planificación de despliegue▪ Planificación de la monitorización y del mantenimiento

- | |
|---|
| <ul style="list-style-type: none">▪ Generación de informe final▪ Revisión del proyecto |
|---|

3.5 Dataset

De la revisión de literatura (Jaramillo, 2017), y el conocimiento de los servicios prestados y usados en la instancia Selene, se definieron los indicadores deseados para la construcción del dataset, a saber

Interacción Estudiante – Evaluación:

- Resultados de Evaluaciones
- Número de Evaluaciones
- Intentos por estudiante
- Intentos por curso
- Bitácora de Evaluaciones

Interacción Estudiante – Plataforma de aprendizaje:

- Ingresos a un curso por estudiante
- Total de ingresos a un curso
- Ingresos en secciones y subsecciones por estudiante
- Ingresos en secciones y subsecciones por curso
- Bitácora de Ingresos

Interacción Estudiante – Contenidos (Videos):

- Vistos de un video
- Repeticiones de un video
- Repeticiones de un video por estudiante
- Videos vistos por estudiante
- Videos vistos por curso
- Progreso

Interacciones Estudiante - Herramientas de comunicación:

- Foros creados por estudiante
- Foros creados por curso
- Respuestas creadas por estudiante
- Respuestas creadas por curso

- Comentarios creados por estudiante
- Comentarios creados por curso
- Votos recibidos

Teniendo como fuentes de información:

- Instancia de Open edX (Selene)
- Sistema de extracción de datos de Selene
- SIMCA (Sistema de Información, Matrícula y Control Académico) de la Universidad del Cauca.
- Encuestas para determinación del estilo de aprendizaje (Google Forms y scripts en Matlab)

Los campos del dataset definidos –específicamente para el curso de Astronomía cotidiana– son los siguientes, en función de la fuente de formación (Tablas 2 a 5):

Ítem	Explicación
ES01	Resultado examen semana 1
ES02	Resultado examen semana 2
ES03	Resultado examen semana 3
ES04	Resultado examen semana 4
ES05	Resultado examen semana 5
ES06	Resultado examen semana 6
ES07	Resultado examen semana 7
ES08	Resultado examen semana 8
ES09	Resultado examen semana 9
ES10	Resultado examen semana 10
ES11	Resultado examen semana 11
ES12	Resultado examen semana 12
ES13	Resultado examen semana 13
R1	Resultado examen remedial 1
R2	Resultado examen remedial 2
Parcial.70	Nota parcial 70% normalizada a 1
Parcial.30	Nota parcial 30% normalizada a 1
Definitiva	Nota Definitiva normalizada a 1
P70.Cual	Nota parcial 70% cualitativa
P30.Cual	Nota parcial 30% cualitativa
Def.Cual	Nota Definitiva cualitativa
Pro6p	Promedio de los seis primeros exámenes presentados
Pro7p	Promedio de los seis mejores exámenes de los primeros siete exámenes

Pro5f	Promedio de los cinco últimos exámenes presentados
-------	--

Tabla 2. Datos proporcionados por la Instancia de Open edX (Selene)

Ítem	Explicación
Ingresos	Cantidad de ingresos al curso
Exámenes.Presentados	Cantidad de intentos de envío de examen y autoevaluación
Accesos.Diapositivas	Cantidad de accesos a diapositivas del curso
Accesos.Paginas.Video	Cantidad de acceso a páginas de video del curso
Play.Video	Número de pausas en los videos del curso
Stop.Video	Número de videos completos vistos sin pausas
Accesos.Recurso.Web	Cantidad de acceso a páginas con recursos web del curso
Acesos.Simuladores	Cantidad de accesos a simuladores
Accesos.Practicas	Cantidad de accesos a prácticas
Accesos.Talleres	Cantidad de accesos a talleres
Foros.Leídos	Cantidad de Foros leídos
Foros.Creados	Cantidad de Foros creados
Total.Actividades.Practicas	Acesos.Simuladores + Accesos.Practicas + Accesos.Talleres
Email	Correo electrónico reportado

Tabla 3. Datos proporcionados por el Sistema de extracción de datos de Selene

Ítem	Explicación
Genero	Género
Programa	Programa Académico
Semestre	Semestre académico matriculado
Promedio.Semestre	Promedio del semestre académico
Diferencia.Aprendizaje	Definitiva - Promedio.Semestre
Puntaje.Ingreso	Puntaje de ingreso a la Universidad
Prioridad	Prioridad del programa académico matriculado
Estrato	Estrato socioeconómico
Nacimiento	Año de nacimiento
Edad	Edad
Municipio	Municipio de procedencia
Departamento	Departamento de procedencia
Tipo.Bachillerato	Bachillerato público o privado

Tabla 4. Datos proporcionados por SIMCA

Ítem	Explicación
Estilo.segun.Kolb	Estilo de aprendizaje según Modelo de Kolb

Tabla 5. Encuestas para determinación del estilo de aprendizaje (Google Forms y scripts en Matlab)

3.6 Discusión y Conclusiones

En los últimos años los cursos en línea en ambiente de masividad se han posicionado como una nueva tecnología educativa que poco a poco se está abriendo paso en la educación superior. Sin embargo, aún hay desafíos que superar, como el adecuado ofrecimiento de los cursos para logren fortalecer habilidades y competencias de los estudiantes. Una alternativa que se presenta es el análisis de las grandes cantidades de datos que se obtienen en las plataformas que permitirán entender cómo se desarrollan los procesos de aprendizaje en entornos virtuales escalables.

La investigación logró mostrar la posibilidad y la metodología utilizada para extraer los datos desde una instancia de Open edX alojada en la Universidad del Cauca denominada Selene e iniciar los primeros pasos para la realización de análisis que permitirán el entendimiento del proceso de aprendizaje de los estudiantes usando como caso de estudio el curso “Astronomía Cotidiana” en modalidad virtual en los términos de interacciones de los estudiantes.

CRIPS-DM resultó ser una metodología robusta y flexible para la construcción de un dataset para aplicación de técnicas de minería de datos o aprendizaje de máquina.

Con el prototipo integrado, se pudo construir un dataset con información de interacción de los estudiantes con una plataforma de MOOC, resultados de aprendizaje, datos demográficos proporcionados por SIMCA y el estilo de aprendizaje según Kolb. No se conoce un *dataset* libre o publicado que relacione información de comportamiento con estilos de aprendizaje según Kolb con lo cual se constituye un importante aporte para los investigadores en este campo del saber.

Capítulo 4

4.1 Analítica del aprendizaje para correlación de variables

Aplicación de procedimientos estadísticos y clasificadores a los conjuntos de datos del curso virtual “Astronomía cotidiana” que incluye datos de interacción de los usuarios con la plataforma MOOC Open edX y sus estilos de aprendizaje según el Modelo de David Kolb.

La principal herramienta utilizada para analizar los cursos es el Análisis de Componentes Principales (PCA) descrito en (Jolliffe, 2002). PCA, al encontrar los valores propios y el vector propio de la matriz de correlación, determina el eje ortogonal de varianza máxima, llamado componente principal o PC. El primer componente o factor principal (F1) captura la mayor cantidad de variabilidad, el segundo (F2) la siguiente mayor variabilidad y así sucesivamente. Las variables se proyectan en aquellos ejes que llaman a estos "factores de carga". Mientras que las observaciones, cuando se proyectan en el eje se llaman "puntajes de factor". Las cargas de los factores muestran cómo las variables se agrupan. La dimensión de un conjunto de datos se puede reducir cuando las cargas de factores de diferentes variables son bastante cercanas. Los puntajes de los factores muestran cómo se agrupan las observaciones. Con el tiempo, se pueden observar grupos de observaciones que sugieren similitudes y diferencias entre los grupos. Además, si los factores resultantes de PCA son adecuados e interpretables, los puntajes de los factores también se pueden utilizar para analizar correlaciones y construir modelos predictivos de los resultados deseados. Se llevó a cabo el siguiente análisis de componentes principales (Solarte, 2018):

1. PCA1: análisis de todas las variables involucradas
2. PCA2: análisis de variables relacionadas con el rendimiento académico
3. PCA3: análisis de actividades realizadas en una instancia de Open edX con algunos datos de perfil
4. PCA4: análisis exclusivamente de actividades realizadas en una instancia de Open edX

Los datos son una consolidación de las actividades de los estudiantes sobre la plataforma, test de estilos de aprendizaje, encuestas, sistemas de información universitario y actividades realizadas fuera de línea (talleres). Por ejemplo, se consideró el cálculo para la calificación del curso del 70%, los mejores 6 resultados de la evaluación de los mejores 8 exámenes, mientras que la calificación para el 30% final se integra con las cinco últimas pruebas: 9 a 13. Además, la clasificación en Kolb los estilos de aprendizaje se realizaron a través de una encuesta de Google Docs que se procesó automáticamente para determinar nueve combinaciones posibles que se redujeron a cuatro estilos principales de aprendizaje.

Los datos combinados contienen 54 variables descritas para 284 muestras de una cohorte original de cerca de 400 estudiantes inscritos en el MOOC. Las variables incluidas en el estudio se describen de la siguiente manera:

Variable	Description
ES01	Exam week 1 (Between 0 and 1)
ES02	Exam week 2 (Between 0 and 1)
ES03	Exam week 3 (Between 0 and 1)
ES04	Exam week 4 (Between 0 and 1)
ES05	Exam week 5 (Between 0 and 1)
ES06	Exam week 6 (Between 0 and 1)
ES07	Exam week 7 (Between 0 and 1)
ES08	Exam week 8 (Between 0 and 1)
ES09	Exam week 8 (Between 0 and 1)
ES10	Exam week 10 (Between 0 and 1)
ES11	Exam week 11 (Between 0 and 1)
ES12	Exam week 12 (Between 0 and 1)
ES13	Exam week 13 (Between 0 and 1)
R1	Remedial exam 1 (Voluntary)
R2	Remedial exam 2 (Voluntary)
Parcial.70	Grade for first 70% course (Between 0 and 1)
Parcial.30	Grade for last 30% course (Between 0 and 1)
Definitiva	Final Grade (Between 0 and 1)
P70.Cual	Qualitative 70% grade (1 <= A <= 0.9 < B <= 0.8 < C <= 0.7 < D <= 0.6 < E <= 0.5 < F)
P30.Cual	Qualitative 30% grade (1 <= A <= 0.9 < B <= 0.8 < C <= 0.7 < D <= 0.6 < E <= 0.5 < F)
Def.Cual	Qualitative final grade (1 <= A <= 0.9 < B <= 0.8 < C <= 0.7 < D <= 0.6 < E <= 0.5 < F)
Pro6p	Average of the first 6 exams
Pro7p	Average of the best 6 exams of the first 7 exams
Pro5f	Average of the last 5 exams
Estilo.segun.Kolb	Style according to Kolb
Kolb	Kolb style code (C=Accomodating S=Assimilative, O=Convergent I=Divergent)
Kolb.Num	Kolb style number (1=Accomodating, 2=Assimilative, 3=Convergent, 4=Divergent)
Genero	Gender 1 (Femenine), 2 (Masculine)
Programa	Academic program
Semestre	Academic semester
Promedio.2016.2	Academic average
Diferencia.Aprendizaje	Learning difference=Definitiva - Promedio.2016.2
Puntaje.Ingreso	Entrance examination average
Prioridad	Academic program priority

Estrato	Social-economic strata
Nacimiento	Year of birth
Edad	Age
Municipio	Current residence municipality
Departamento	Current residence department
Tipo.Bachillerato	Public or private highschool
Ingresos	Number of accesses to online course
Exámenes.Presentados	Number of exams solved
Accesos.Diapositivas	Number of hits on course slides
Accesos.Paginas.Video	Number of hits on course videos
Play.Video	Number of pauses while viewing videos
Stop.Video	Number of videos completed without pauses
Accesos.Recurso.Web	Number of hits on pages with course resources
Accesos.Simuladores	Number of hits on simulators
Accesos.Practicas	Number of hits on labs
Accesos.Talleres	Number of hits on workshops
Foros.Leidos	Number of read fora
Foros.Creados	Number of created fora
Total.Actividades.Practicas	Simulator hits + Labs hits + workshop hits
Code	Código para sincronización de las diferentes fuentes de información

4.1.1 Comportamientos característicos de la calificación definitiva y la diferencia en el aprendizaje

Usando la variable Total.Actividades.Practicas que indica la adición de todas las actividades en línea llevadas a cabo por los estudiantes. En las figuras 26, 27, 28, 29 y 30 se puede apreciar hay diferencias entre las diferentes variables demográficas.

En la Figura 26, se muestra la suma de todas las actividades prácticas, que representan el trabajo en línea de los estudiantes, frente a los estratos socioeconómicos. En la variable de estratos, 1 significa más bajo, 4 significa más alto.

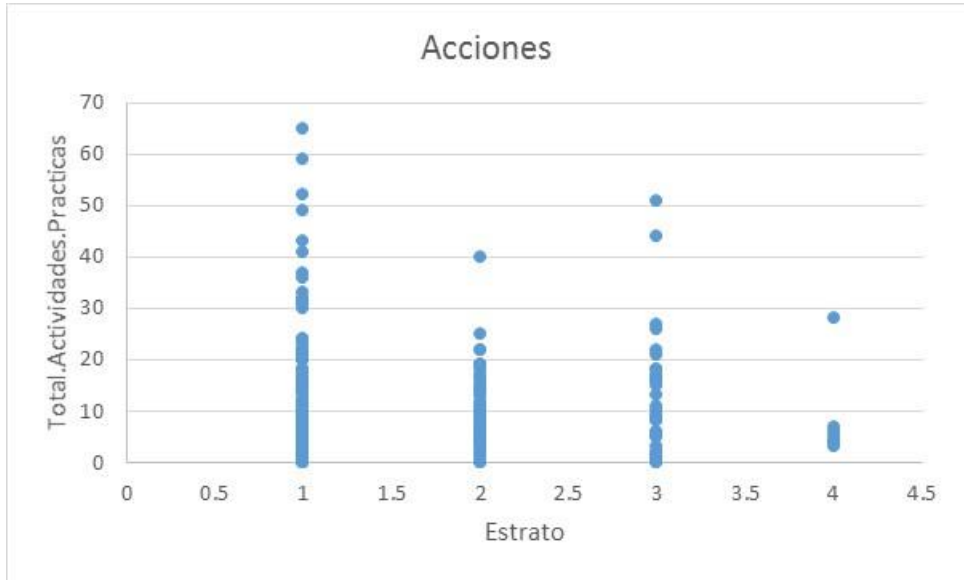


Figura 26. Total trabajo en línea vs estratos socioeconómicos

En la Figura 27 observamos que existe una tendencia de las personas de clase alta a hacer mucho menos trabajo global. Sin embargo, es importante señalar que las personas de la clase más alta representan una minoría; solo 7 de 253. Por lo tanto, la diferencia en la variabilidad podría no ser tan significativa como parece al principio.

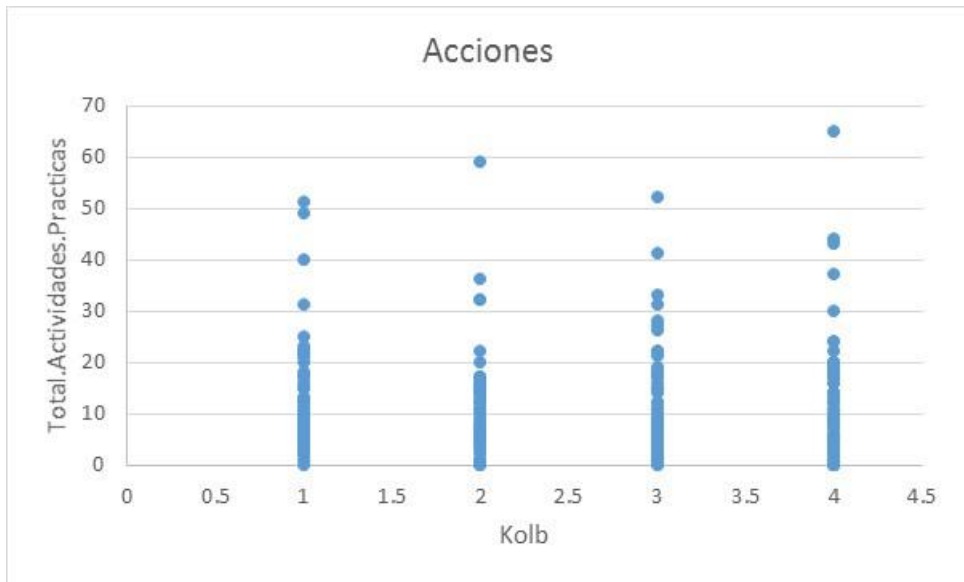


Figura 27. Total de actividades prácticas vs estilo de aprendizaje Kolb (1= Acomodador, 2= Asimilador, 3= Convergente, 4= Divergente)

En la Figura 28 se muestra el valor del trabajo práctico total frente al estilo de aprendizaje de Kolb. Parece que hay muy poca diferencia en el rendimiento del trabajo en línea por estilo.

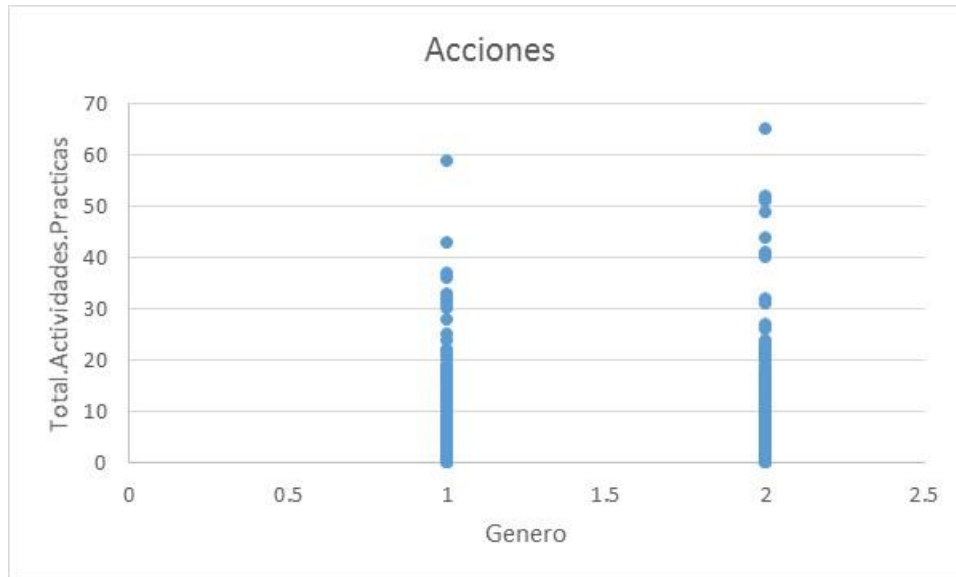


Figura 28. Total de actividades prácticas vs. Género (1 = femenino, 2 = masculino)

En la Figura 29 se muestra el trabajo práctico total contra el género. 1 es femenino, 2 es masculino. Parece que hay una clara ventaja para los hombres en el rendimiento. Pero, de nuevo, las mujeres son una minoría. La muestra analizada contenía 103 mujeres y 150 hombres. El promedio para las mujeres fue de 11.44 y para los hombres de 11.33. Entonces, parece que las mujeres eran un poco más activas que los hombres. Pero, una vez que se realizó una prueba de hipótesis de dos poblaciones frente a la media, resultó que con un valor de p de 0,939 frente a un alfa de 0,05, no podemos descartar la hipótesis, con una certeza del 93,86%, de que ambos promedios son estadísticamente los mismos. Por lo tanto, no parece haber diferencia entre el rendimiento en línea de los hombres frente a las mujeres.

En la Figura 29 mostramos de nuevo el total de actividades prácticas esta vez frente a la edad. Parece que cuanto más jóvenes son los estudiantes son más propensos a llevar a cabo las actividades prácticas que son. Pero una vez más, las personas mayores en el grupo son una minoría muy pequeña como se identificó en los datos. La gran mayoría de la población de estudiantes se encuentran en el rango 19-22, que en la Figura 29 es el área con mayor variabilidad.

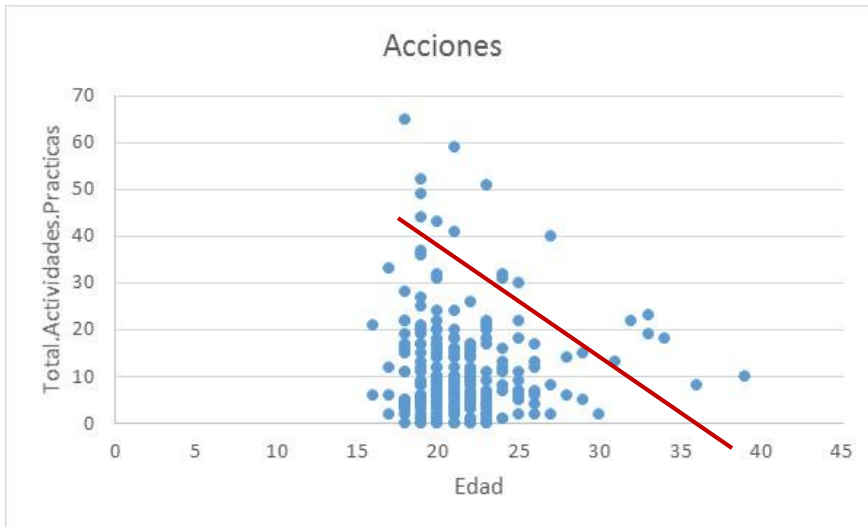


Figura 29. Total de actividades prácticas vs. edad

En la Figura 30 mostramos el trabajo práctico total frente al puntaje del examen de ingreso. No parece haber diferencia entre los diferentes puntajes y el desempeño práctico del trabajo que no puede explicarse por las diferencias en las muestras.

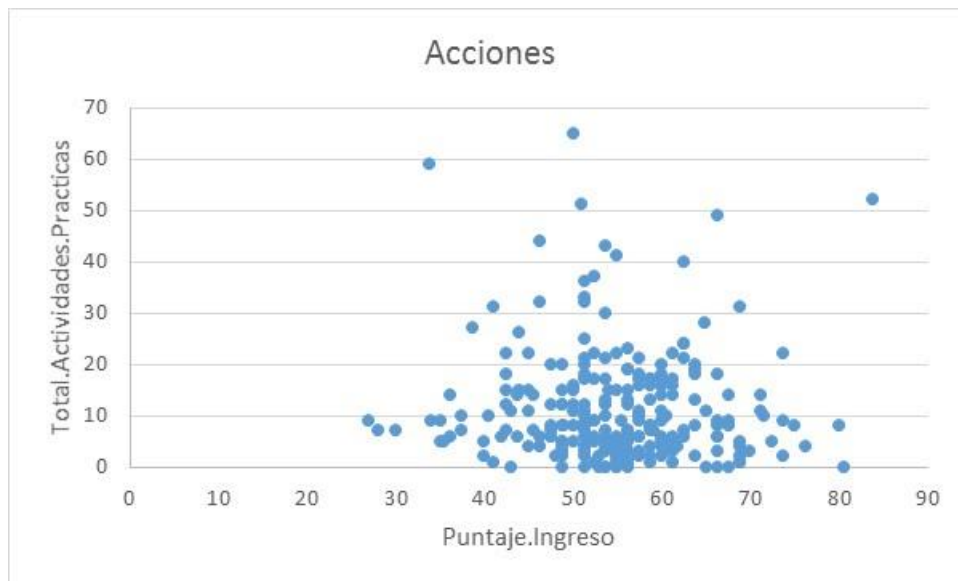


Figura 30. Total trabajo práctico vs puntaje de ingreso

Los resultados del aprendizaje tienen dos medidas: Definitiva (calificación final) y Diferencia.Aprendizaje, o diferencia de aprendizaje = Definitiva - Promedio.2016. La Figura 31 muestra la distribución de la calificación final por género.

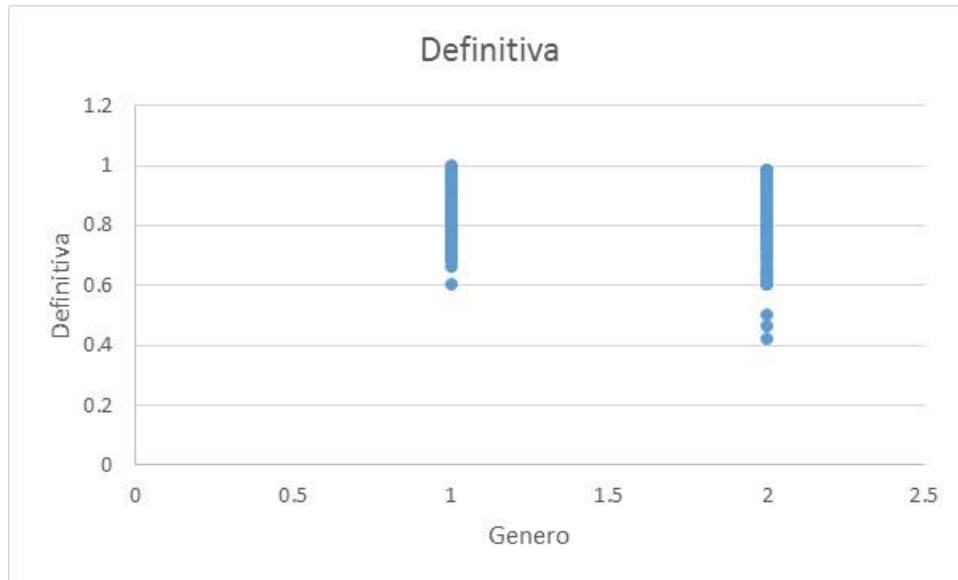


Figura 31. Distribución del desempeño académico (Definitiva) por género

En la figura 32 parece mostrar que las mujeres no tienen calificaciones tan bajas como los hombres. Sin embargo, en promedios, son lo mismo. Las mujeres tienen un promedio de 0.8222 mientras que los hombres tienen un promedio de 0.8292. La prueba de dos muestras indica que con un valor p de 0.611 no podemos rechazar la hipótesis de que ambas medias son estadísticamente iguales.

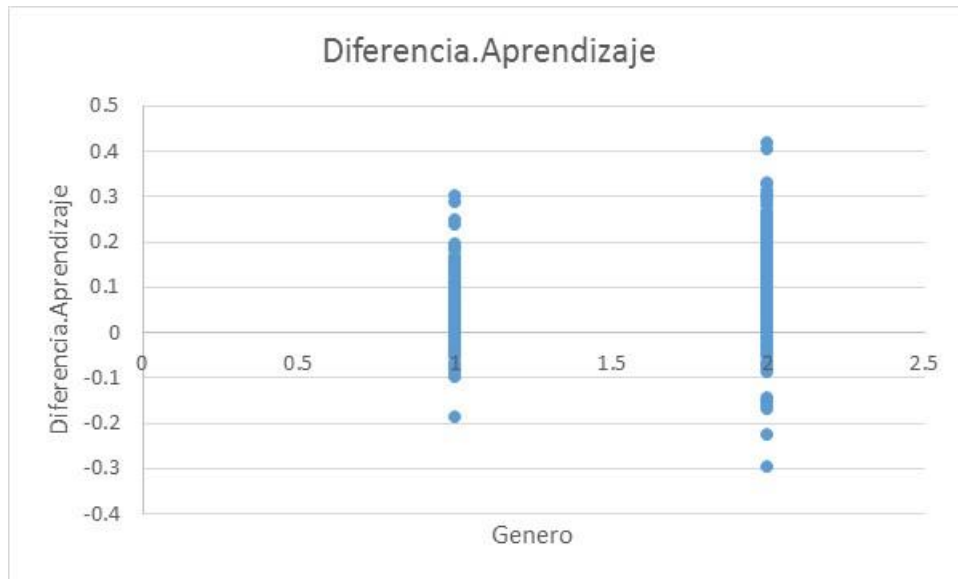


Figura 32. La diferencia de aprendizaje por género

En la figura 32 parece haber más variabilidad en la diferencia de aprendizaje de los hombres. La prueba de hipótesis para dos muestras muestra que, de hecho, hay una diferencia estadística en la media. Las mujeres tienen una media de 0.05 y los hombres de

0.102. Con un valor muy bajo de p , debemos rechazar la hipótesis de que ambos medios son iguales y, por lo tanto, los varones tienen la mayor diferencia en el aprendizaje.

Otra característica importante es la distribución de la calificación final y de la diferencia en el aprendizaje. El histograma de la calificación final como se muestra en la variable Definitiva se muestra en la Figura 33. Parece mostrar que el curso es muy indulgente, ya que las calificaciones están muy sesgadas hacia una calificación alta. Debemos recordar aquí, que la calificación aprobatoria es 0.6 con un mínimo de 0 y un máximo de 1.

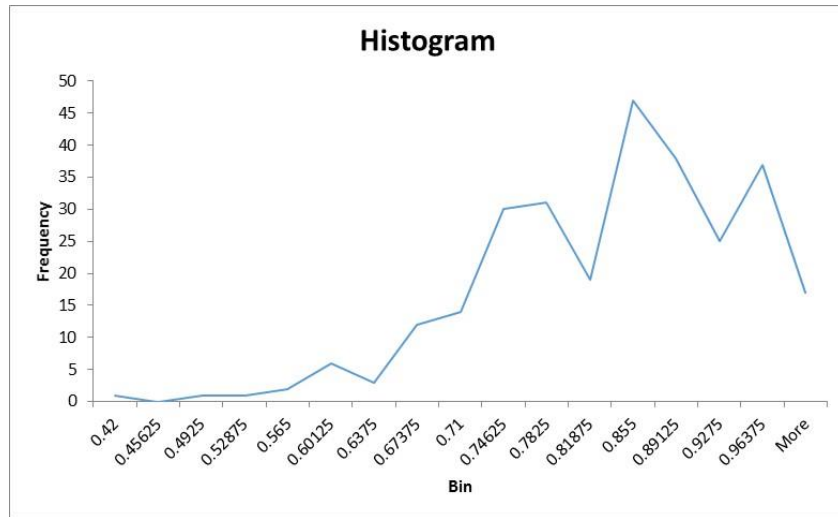


Figura 33. Histograma del rendimiento académico

Una prueba de normalidad que usa las técnicas de Shapiro-Wilk y Anderson-Darling, de hecho, muestra que la calificación final no sigue una distribución normal. En ambos casos, el valor p estaba por debajo de 0.0001, que es mucho menor que el alfa de 0.05.

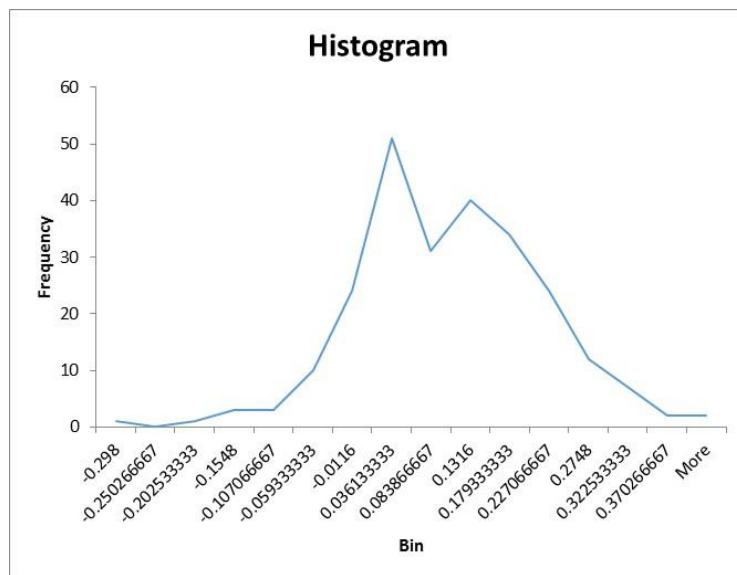
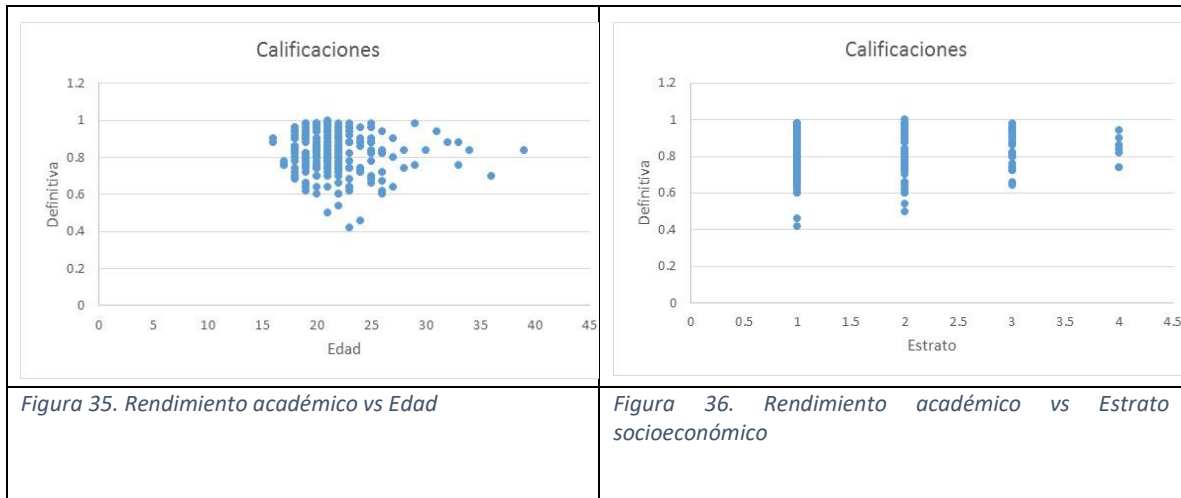


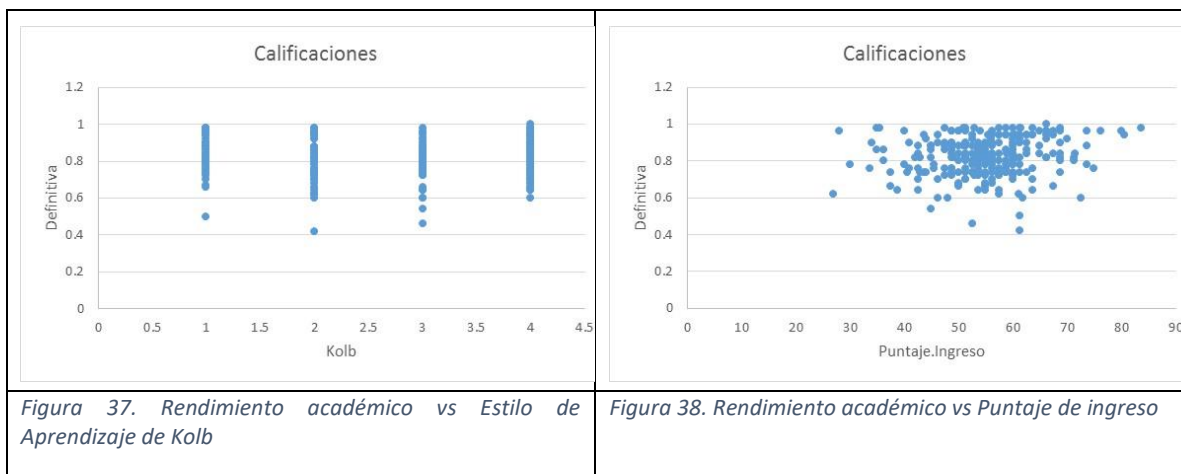
Figura 34. Diferencia en el histograma de la diferencia de aprendizaje

La Figura 34 muestra el histograma para la diferencia en el aprendizaje. Las pruebas de normalidad usando las mismas técnicas dieron valores de p de 0.269 y 0.158 respectivamente, lo que indica que la diferencia en el aprendizaje sí sigue una distribución de probabilidad normal. Este hecho es importante porque las pruebas normales llevadas a cabo en el tamaño de las muestras y las pruebas de hipótesis solo son válidas si los datos siguen una distribución de probabilidad normal. Resulta que la nota final no. Por lo tanto, se convierte en una medida poco confiable del rendimiento estudiantil.

En las figuras 35 y 36 vemos la calificación final en comparación con la edad y los estratos socioeconómicos.

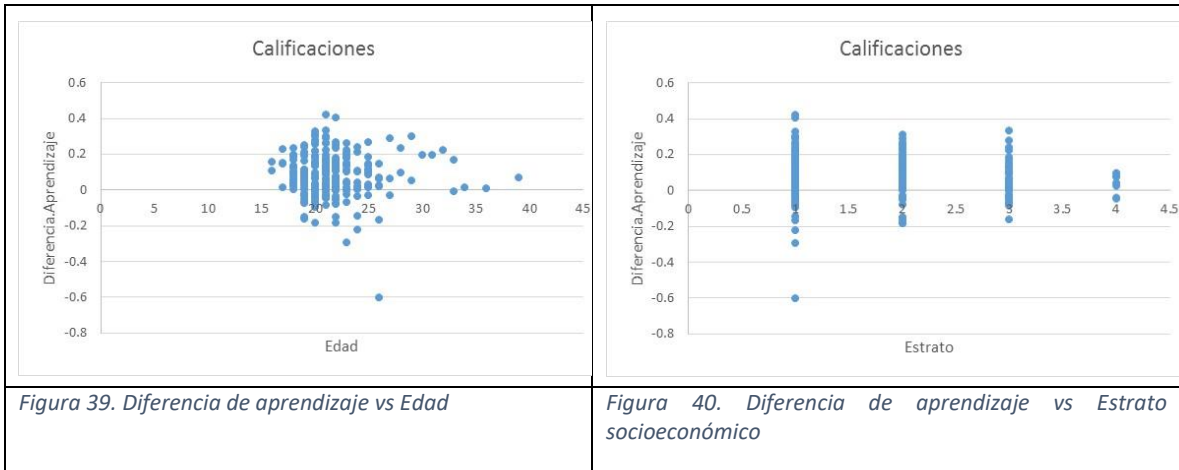


En las figuras 37 y 38 vemos que la calificación final comparó el estilo de aprendizaje de Kolb y el puntaje del examen de ingreso.

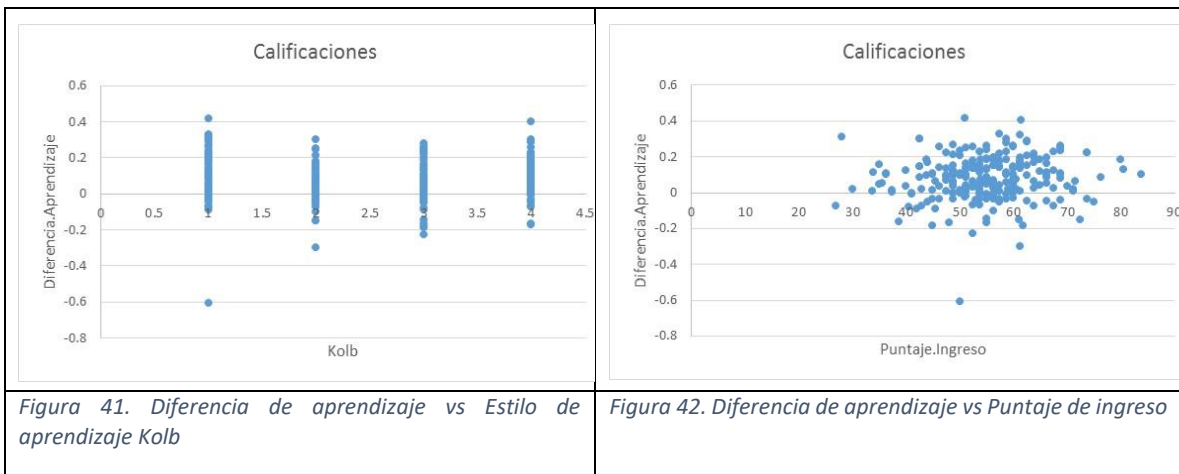


En todos los casos, no hay diferencias claras que no puedan explicarse por la diferencia en el tamaño de la muestra, por ejemplo, estratos o edad. La calificación final va de 0.6 a 1 en todos los estilos de aprendizaje de Kolb y parece que no hay diferencia en el examen de nivel de entrada.

En las figuras 39 y 40 mostramos la diferencia en el aprendizaje frente a la edad y los estratos.



En las figuras 41 y 42 vemos la diferencia en el aprendizaje comparado con el estilo de aprendizaje de Kolb y el puntaje del examen de ingreso. La Figura 38b muestra que la diferencia en el aprendizaje no es la misma para todos los estilos de aprendizaje. Recordando que 1 = Acomodando, 2 = Asimilativo, 3 = Convergente y 4 = Divergente, vemos menos diferencia de aprendizaje para asimilativo y convergente. Eso no quiere decir que lo hicieron mal, solo significa que no ganaron tanto. Podría ser que comenzaron desde un alto nivel de experiencia en comparación con otros perfiles.



4.1.2 Reducción de la dimensionalidad

Se requiere determinar qué variables no brindan información útil y pueden eliminarse. Hacemos esto al revisar en PCA1, PCA2 y PC3 los grupos de variables (resultados de la evaluación o actividades prácticas de los estudiantes) que brindan la misma información. Esto se muestra en los biplots para F1 y F2. Estos biplots muestran la correlación entre las variables y dos ejes de PCA, principalmente F1 y F2. Son equivalentes al trazado de las cargas factoriales. Mostramos esto en las figuras 15, 16 y 17. Los biplot sticks que son bastante horizontales están correlacionados con F1 y los verticales están correlacionados con F2. Los palos diagonales están correlacionados con ambos. Los palos cortos son más prominentes en otras dimensiones (F3, F4, etc.).

Las variables con barras agrupadas están altamente correlacionadas independientemente de su orientación. Por lo tanto, en cualquier grupo cercano de variables, todas pueden descartarse excepto una. Dado que cada variable representa una acción o un grado, podemos reducir el número de evaluaciones o descartar actividades que no son relevantes (y posiblemente dejar espacio para otras actividades).

La Figura 43, usando todas las variables, muestra dos grupos diferenciados. Uno es sobre actividades prácticas (se muestra en azul) y el otro sobre datos de resultados de la evaluación. También muestra que no parece haber correlación entre las actividades de aprendizaje práctico y la calificación final (Definitiva) y la diferencia en el aprendizaje. Vamos a estudiar esto más adelante a continuación. Como la calificación final es una combinación lineal de todas las calificaciones, obviamente la calificación final se correlaciona con las variables de calificación. También parece que estas variables también están correlacionadas con la variable de diferencia de aprendizaje.

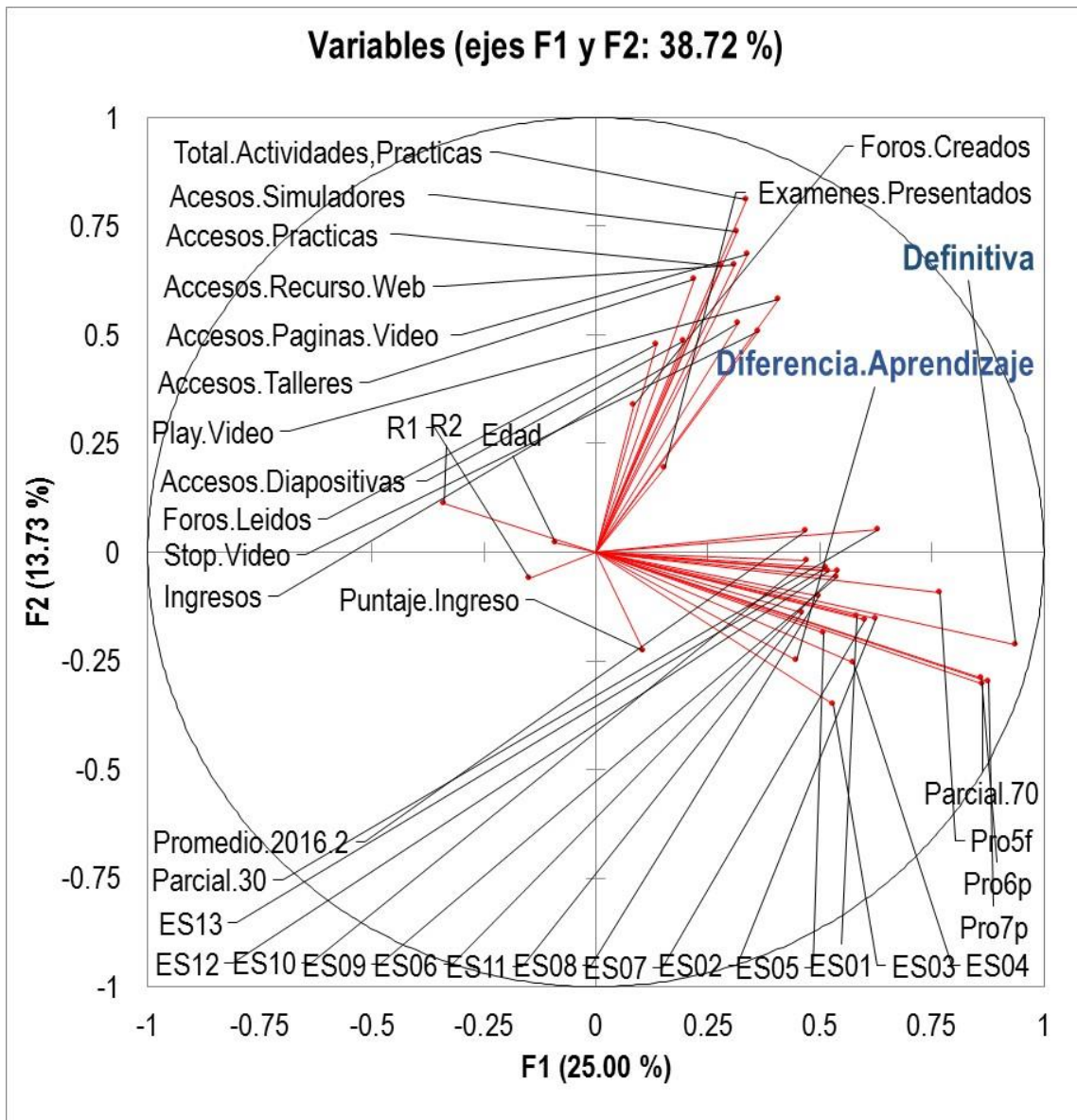


Figura 43. Biplot de PCA 1 que muestra grupos de variables. El clúster azul trata sobre los resultados de la evaluación.

Para determinar es posible reducir el número de evaluaciones de cursos, se llevó a cabo PCA 2, que contiene solo información sobre las evaluaciones. El biplot se muestra en la Figura 44. En esta figura, vemos dos grupos de evaluación cerrados como se muestra en la Tabla 6. El primer clúster cerrado que se muestra en verde está formado por las Evaluaciones 1 a 8. El segundo grupo en azul consiste en evaluaciones de 9 a 13. Los óvalos púrpuras también muestran que el grupo verde está relacionado con los resultados de la evaluación Pro7p, Pro6p y Parcial.70, mientras que el grupo azul está relacionado con los resultados de la evaluación Prof5f y Parcial.30. Significa que el curso está dividido en dos secciones. También significa que el curso está sobre evaluado. De la evaluación 1 a 8, 7 pueden descartarse dado que brindan la misma información. De las evaluaciones 9 a 13 4 pueden descartarse dado que también proporcionan la misma información. La evaluación del curso debe consistir en dos evaluaciones parciales y una evaluación global final.

La Figura 44 también muestra que el nivel socioeconómico no está relacionado con las notas, ya que las variables cortas indican prominencia en otras dimensiones y, en general, estas palancas están orientadas ortogonalmente a los resultados de la evaluación. Sin embargo, quedan algunas correlaciones que exploraremos más adelante.

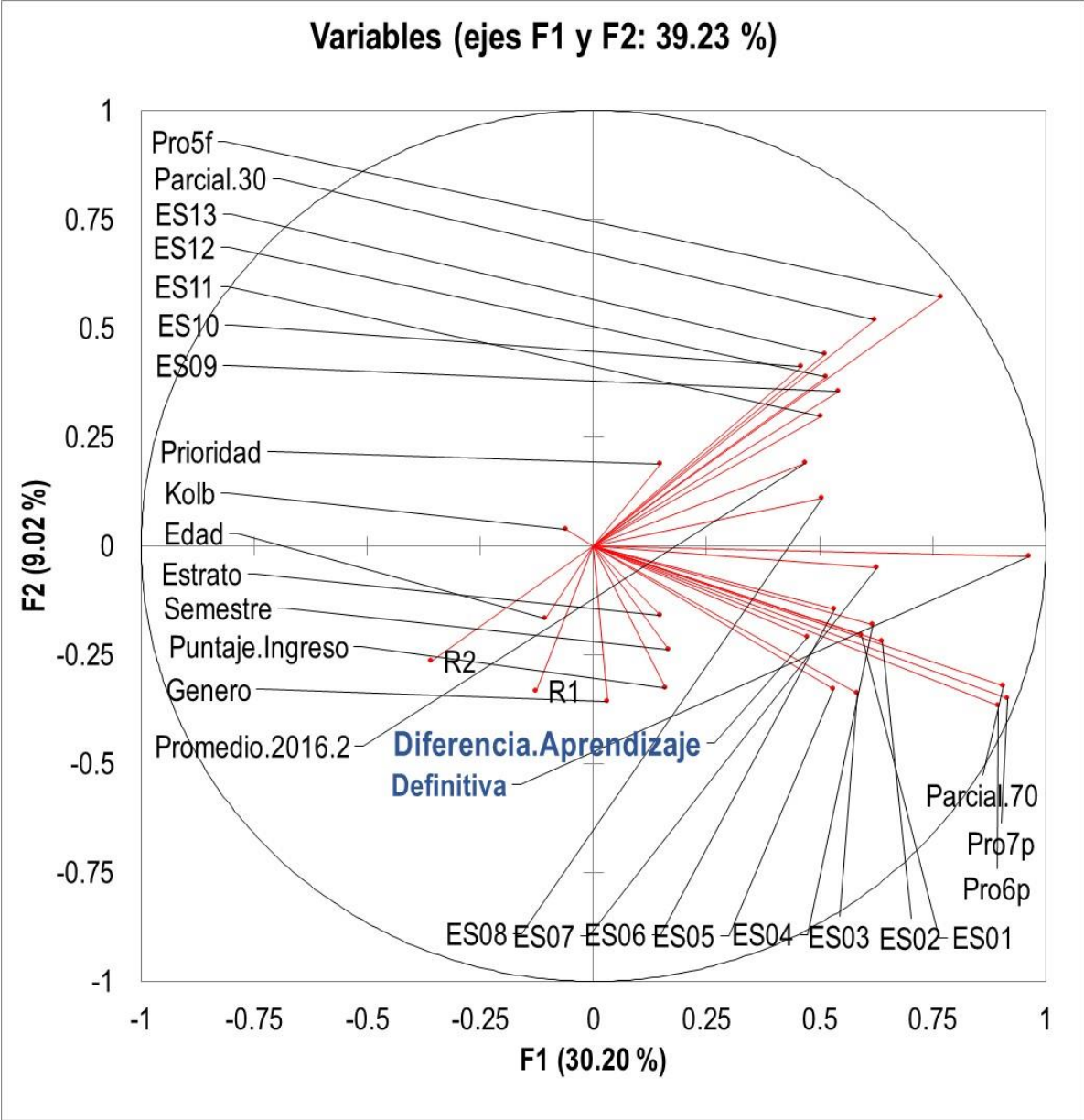


Figura 44. Biplot de PCA 2 que muestra grupos de variables que indican diferentes evaluaciones

Cluster	Members
Verde	ES01, ES02, ES03, ES04, ES05, ES06, ES07, ES08
Azul	ES09, ES10, ES11, ES12, ES13
Violeta	Green + Pro6p + Pro7p + Parcial.70 Blue + Pro5f + Parcial.30

Tabla 6. PCA 2 Biplot Clusters

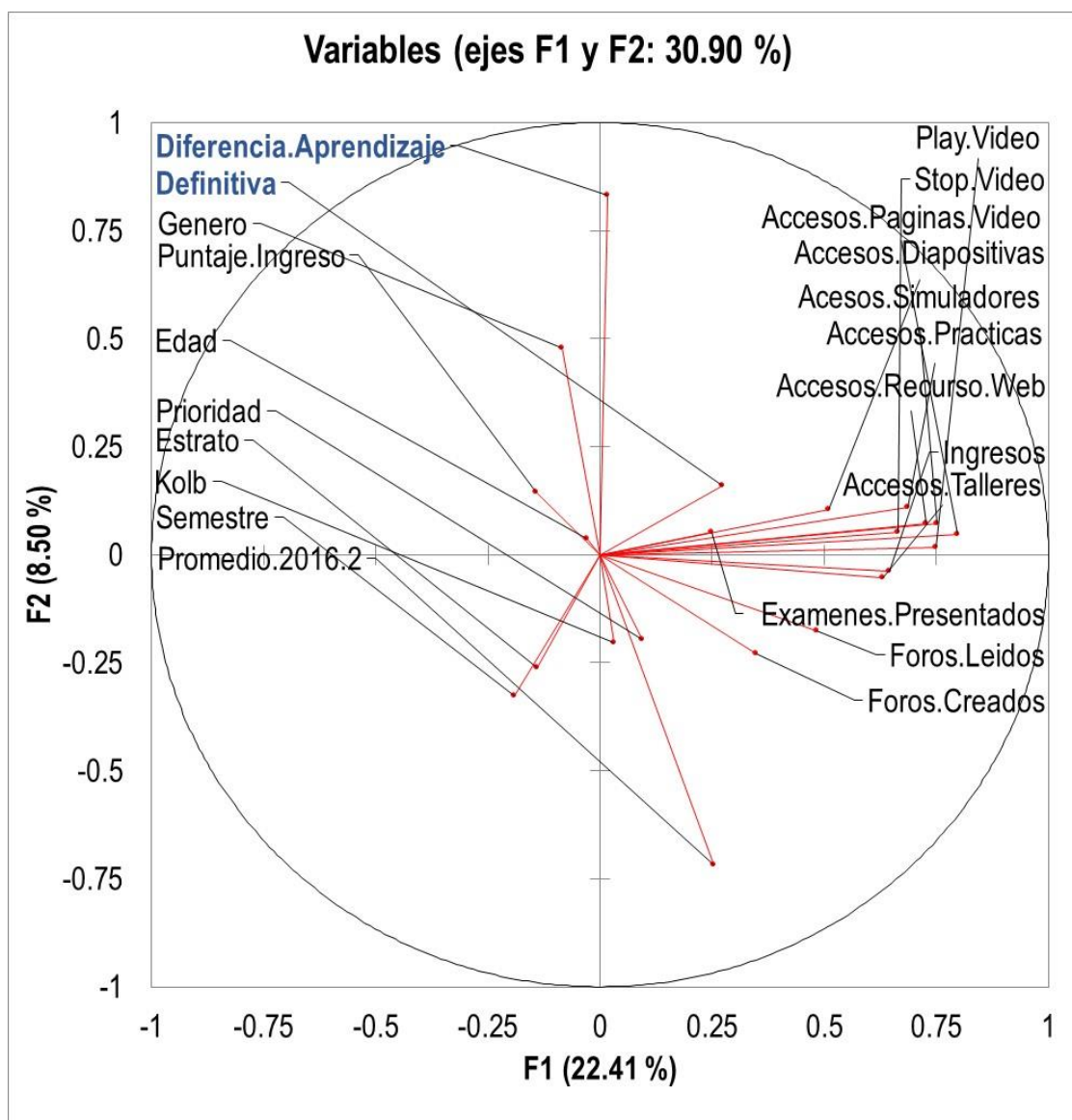


Figura 45. Biplot de PCA 3 que muestra grupos de variables que indican actividades que dan resultados similares.

La Figura 45 muestra el biplot para PCA3 que está relacionado con las acciones de los estudiantes. Vemos tres miniclusters en la Tabla 7. El clúster azul se trata de acceder a los

recursos. El grupo verde se trata de acceso general a la plataforma. El grupo púrpura se trata de acceder a foros en línea. En el grupo verde, se pueden descartar varias variables. En el grupo azul, se pueden descartar hasta 5 variables.

Cluster	Members
Verde	Ingresos, Accesos.Talleres
Azul	Play.Video, Stop.Video, Accesos.Paginas.Video, Accesos. Simuladores, Accesos.Practicas, Accesos.Recurso.Web
Violeta	Foros.Leidos, Foros.Creados

Tabla 7. PCA 3 Biplot Clusters

Sin embargo, una vez que se elimina la información del grado, vemos una imagen un tanto diferente en la Figura 42, donde se muestra el biplot para PCA 4 (solo las actividades prácticas incluidas sin ningún grado de datos).

En la Figura 46 se muestran tres grupos de actividades (ver Tabla 8). El clúster 1 en azul se refiere al acceso de video. El grupo 2 en verde se refiere al acceso a los recursos en línea. El grupo 3 en rojo se refiere al acceso a los laboratorios de práctica. Y el grupo 4 en morado se relaciona con el acceso a los foros de discusión.

Cluster	Members
Azul	Play video, stop video
Verde	Ingresos, Acceso.Paginas.Web, Acceso.Paginas.Video
Rojo	Acceso.Practicas, Acceso.Simuladores, Total.Actividades
Violeta	Foros.Creado, Foros.Leidos

Tabla 8. PCA 4 Biplot Clusters

Por lo tanto, una variable puede eliminarse del clúster 1, dos del clúster 2, dos de los clústeres 3 y uno del clúster 4.

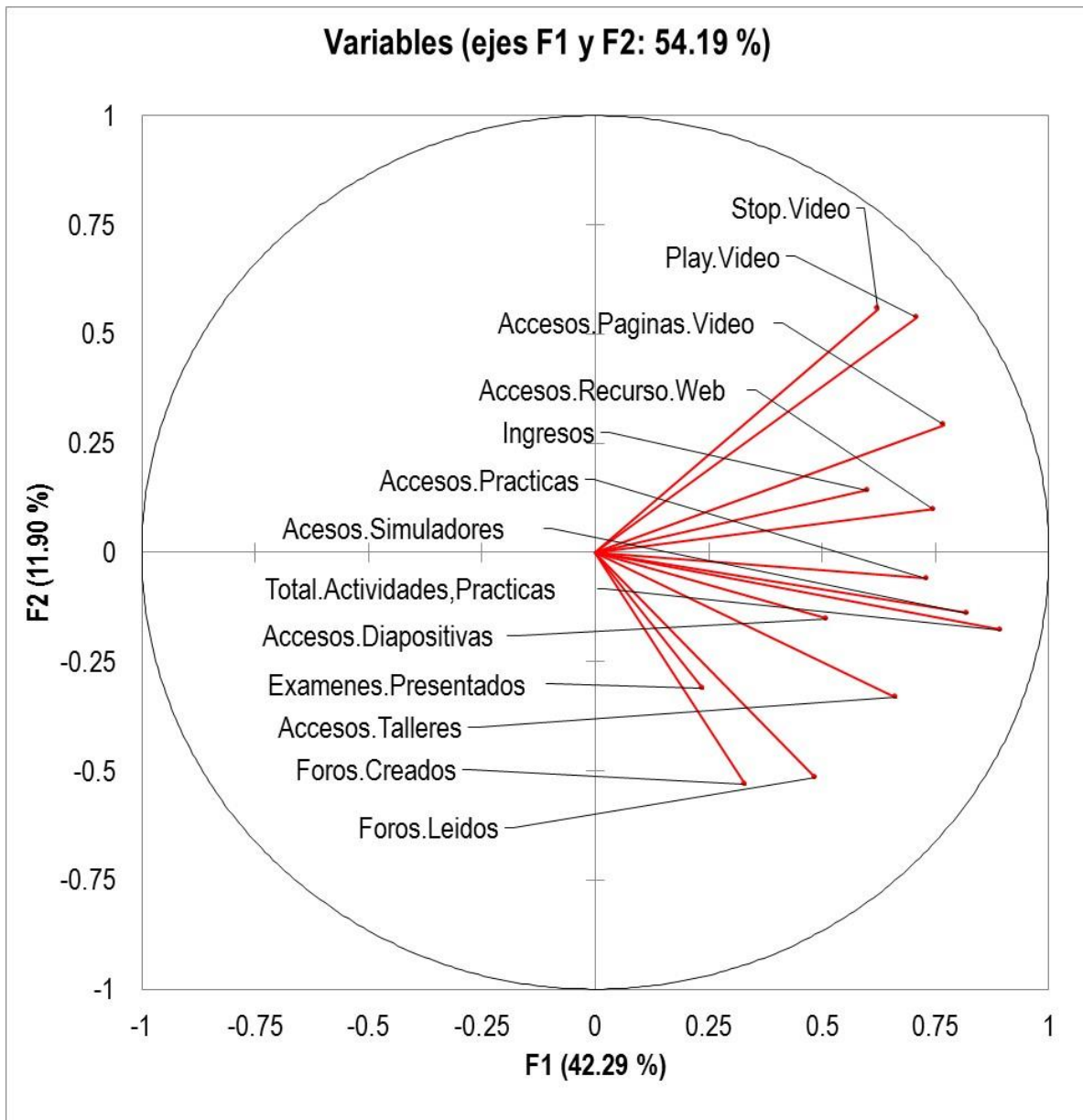


Figura 46. Biplot de PCA 4 que muestra grupos de variables que indican actividades que dan resultados similares

4.2 Clasificación y clustering

Mientras que la reducción de dimensión se logra agrupando cargas de factores, o proyecciones variables hacia el eje componente, las agrupaciones de estudiantes pueden descubrirse trazando las proyecciones de las muestras, denominadas puntuaciones de observación. Si apropiadamente etiquetamos esas proyecciones, aparecen grupos basados en diferentes características.

4.2.1 Agrupación por Rendimiento Académico

PCA2 usa datos de grado estudiantil. En la Figura 43 mostramos un diagrama de dispersión en el que el eje horizontal corresponde a las puntuaciones de F1 y el eje vertical corresponde a las puntuaciones de muestra de F2 de PCA2, que utiliza información de grado. El color se basa en la clasificación del estilo de aprendizaje de Kolb (los estudiantes toman encuestas para determinar esta clasificación). En la figura 20 mostramos el mismo diagrama de dispersión que en la Figura 42 con los colores correspondientes a la diferencia de aprendizaje.

En la Figura 47 vemos que parece haber una línea divisoria que separa el cuadrante en el que tanto F1 como F2 están por encima de cero, en donde los perfiles de estilo de aprendizaje C y S parecen dominantes. Como tanto F1 como F2, como se ve en la estadística de los cosenos cuadrados, están relacionados con el rendimiento general de los estudiantes, los estudiantes para los que se coloca la muestra en este cuadrante serían estudiantes de mayor rendimiento.

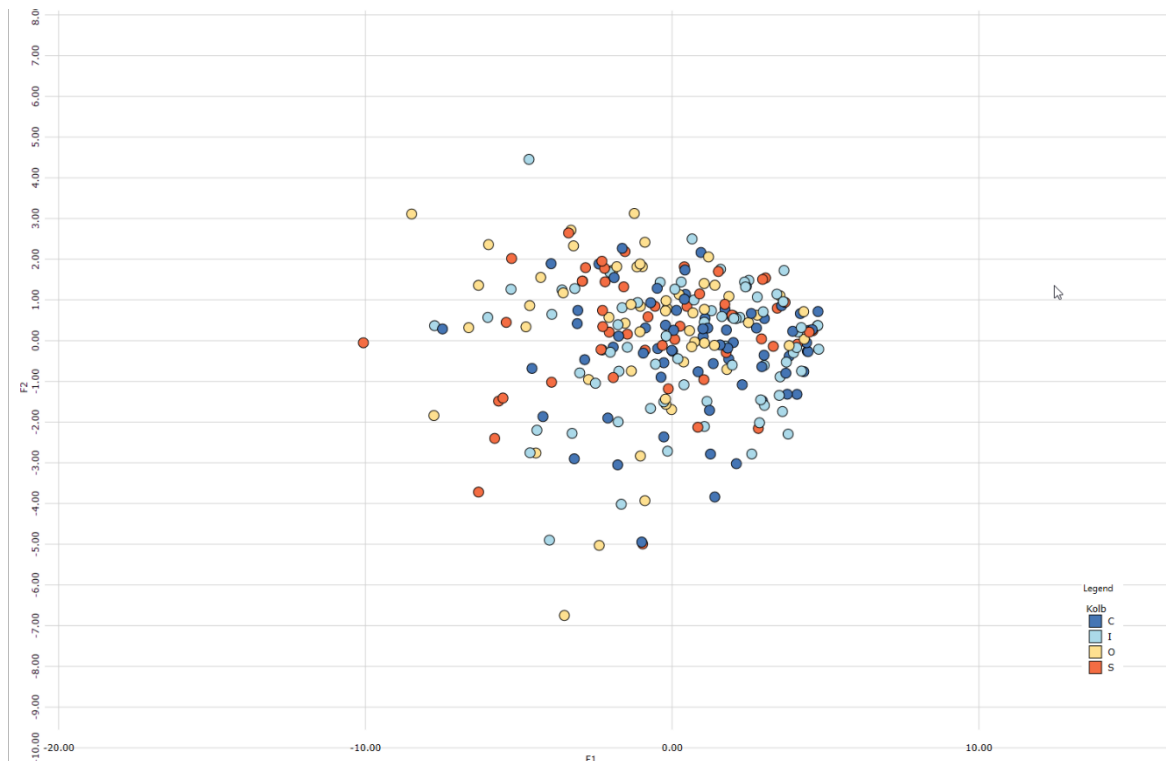


Figura 47. Diagrama de dispersión PCA 2 (Grades) de las puntuaciones F1 y F2. El color se basa en los estilos de Kolb

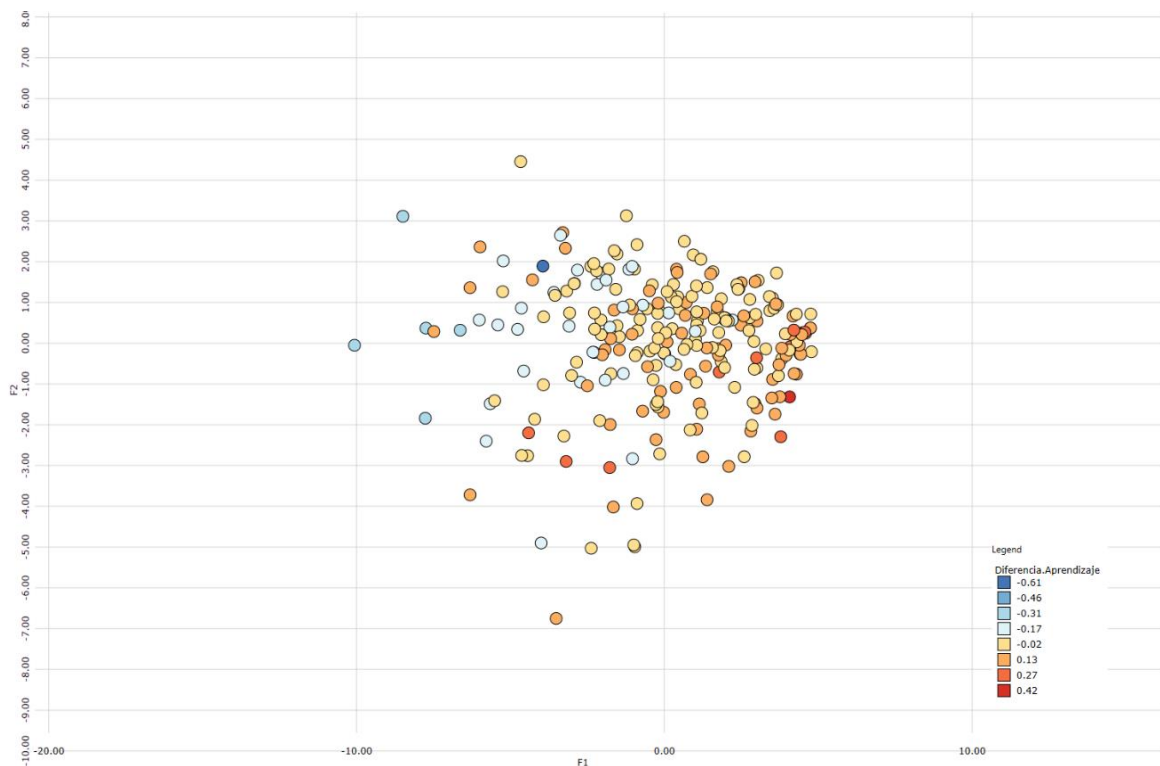


Figura 48. Diagrama de dispersión PCA 2 (Rendimiento académico) de las puntuaciones F1 y F2. El color se basa en la diferencia de aprendizaje

En la figura 48, parece haber una línea divisoria que depende solo de F1, que separa la diferencia de aprendizaje principalmente positiva de la diferencia de aprendizaje, en su mayoría negativa.

Para investigar lo que las figuras 47 y 48 parecen mostrar, necesitamos ver los datos en la Tabla 9 en los que se muestra la distribución de los estilos de aprendizaje.

Kolb Labels	Count	% of Total
C	73	29
I	72	29
O	58	23
S	49	19
Grand Total	252	

Tabla 9. Distribución de los estilos de aprendizaje de Kolb

En la Tabla 10, se muestra el recuento de estudiantes en dicho cuadrante por estilo de aprendizaje. Muestra que, aunque los perfiles de recuento C y I son dominantes (23 ambos),

en porcentajes por estilo, no lo son. También vemos que el 23% de todas las muestras de estilo C están en el cuadrante, mientras que la cifra para los estilos I, O y S es del 32%, 24% y 31%. Por lo tanto, solo podemos concluir que el estilo O es débil en el rendimiento general de la calificación medida por F1 y que todos los demás estilos son iguales.

Kolb Style	Count	% of Total	% of Kolb Style
C	23	0.09	0.32
I	23	0.09	0.32
O	14	0.06	0.24
S	15	0.06	0.31

Tabla 10. Recuento de muestras en el cuadrante de mayor rendimiento y porcentajes

En la Tabla 11 mostramos los promedios en F1 por estilo de aprendizaje y por género. Vemos que los hombres superan a las mujeres y que los perfiles C e I tienen promedios en el cuadrante (valores F1 positivos), mientras que los perfiles O y S no lo hacen.

Avg of F1	Column Labels		
Row Labels	1	2	Grand Total
C	0.7031	0.8605	0.8217
I	0.2447	0.7250	0.4849
O	-0.9254	-	1.0681
S	-0.3492	-	1.0731
Grand Total	-0.1007	0.0685	0

Tabla 11. Promedios F1 por estilo de aprendizaje y género

Sin embargo, las pruebas estadísticas muestran con $\alpha = 0.05$ y $p = 0.001$ que los medios para aprender estilos son de hecho estadísticamente diferentes, mientras que con $p = 0.661$, los promedios entre géneros no son diferentes con ninguna significación estadística. Por lo tanto, los hombres superan a las mujeres en los resultados de la evaluación, pero solo ligeramente. Mientras que los Perfiles C y I claramente superan los perfiles O y S.

La Tabla 12 muestra el promedio en la diferencia de aprendizaje considerando tanto el perfil como el género. Los resultados coinciden con nuestras conclusiones del cuadro 8. En este caso, la diferencia entre la diferencia de aprendizaje en mujeres y hombres es estadísticamente significativa con $p = 0,003$ y entre los estilos de aprendizaje con $p = 0,013$. Por lo tanto, los hombres tienen una diferencia de aprendizaje más amplia ya que las

mujeres y los perfiles C y I tienen una diferencia de aprendizaje más amplia que los perfiles O y S.

Avg of Dif.Aprend	Column Labels		
Row Labels	1	2	Grand Total
C	0.0800	0.1127	0.1046
I	0.0764	0.1132	0.0948
O	0.0212	0.0776	0.0523
S	0.0274	0.0661	0.0487
Grand Total	0.0524	0.0969	0.07889

Tabla 12. Diferencia de aprendizaje por estilo de aprendizaje y género

La Tabla 13 muestra la diferencia en el aprendizaje por estratos socioeconómicos. El análisis ANOVA con $p = 0,433$ muestra que la diferencia en el promedio no es estadísticamente significativa.

Row Estrato	Labels Average of Dif.Aprend
1	0.0793
2	0.0915
3	0.0605
4	0.0314
Total	0.0789

Tabla 13. Diferencia de aprendizaje vs estrato socioeconómico

La Figura 49 muestra un diagrama de caja para aprender las diferencias por edad. Con $p = 0.314$, las diferencias percibidas en los medios no son estadísticamente significativas. Se muestra que 32 años tienen la media más grande, pero el tamaño de la muestra hace que el intervalo de confianza sea demasiado amplio.

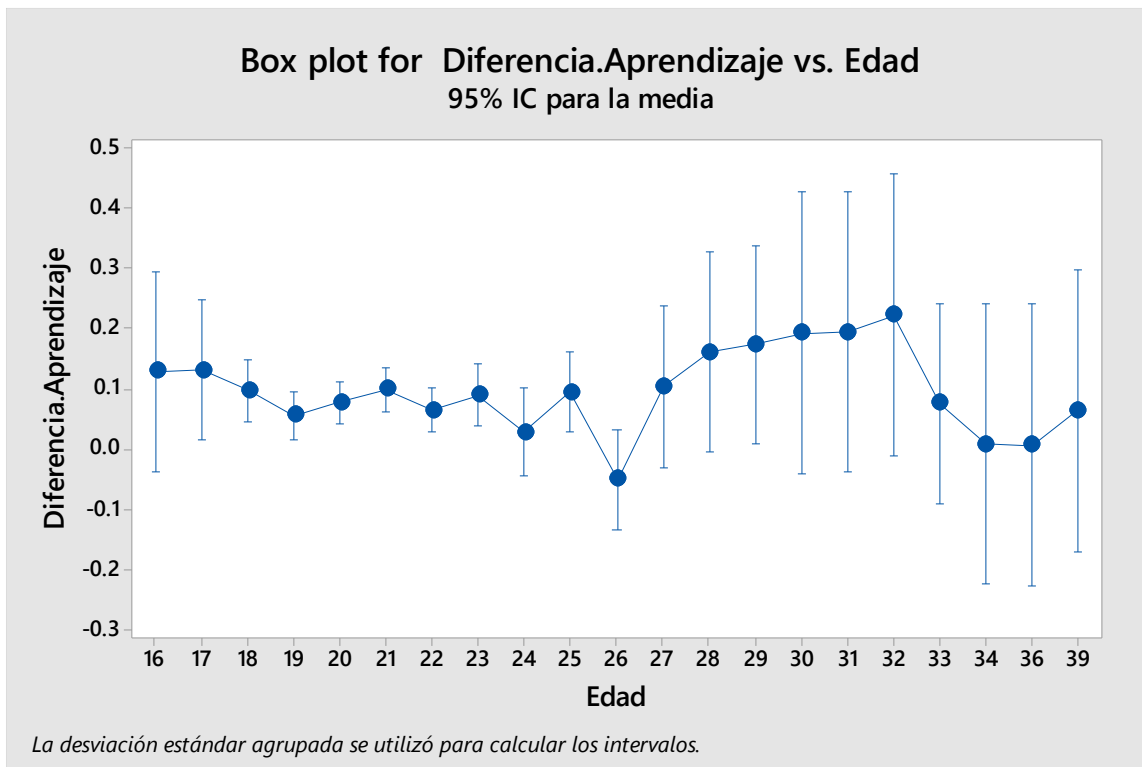


Figura 49. Diagrama de caja para el promedio de la diferencia de aprendizaje por edad

4.2.2 Agrupación por actividades de aprendizaje

Las actividades se analizan en PCA4. PCA4 incluye solo actividades prácticas. Según las estadísticas de los cosenos al cuadrado, F1 se puede asociar a 9 de 13 actividades y F2 a solo dos, que son el acceso a los foros. Por lo tanto, F1 sirve como una medida global de las actividades de la plataforma de los estudiantes y F2 como complemento de la actividad de los foros.

En la Figura 50 hay una gráfica de dispersión de F1 frente a F2 por PCA 4, es decir, actividades según lo registrado por la plataforma. Los colores son por el perfil de estilo de aprendizaje de Kolb. Parece haber una línea para $F1 > 0$ y $F2 > 0$ que parece estar dominada principalmente por el perfil C, con presencia de I y S. Esto indica que diferentes perfiles llevarán a cabo diferentes actividades en la plataforma. Además, hay un grupo (círculo azul) en $F1 < 0 << 1$ en el que el perfil I es dominante. Es decir, los estudiantes de perfil I son más propensos a llevar a cabo actividades de aprendizaje en absoluto según lo registrado por la plataforma.

En la Figura 51 mostramos el mismo diagrama de dispersión que en la Figura 50, pero los colores se basan en la diferencia de aprendizaje. Parece que no hay una agrupación obvia de acuerdo con las diferencias de aprendizaje que muestran las actividades de aprendizaje.

La Figura 50 parece indicar que los diferentes perfiles de estilo de aprendizaje se comportarán de manera diferente cuando se lleven a cabo las actividades registradas por la plataforma. Pero, la Figura 51 parece indicar que esas actividades de aprendizaje no

tienen influencia en los resultados del aprendizaje. Esto se investigará más a fondo en la próxima sección de modelos predictivos.

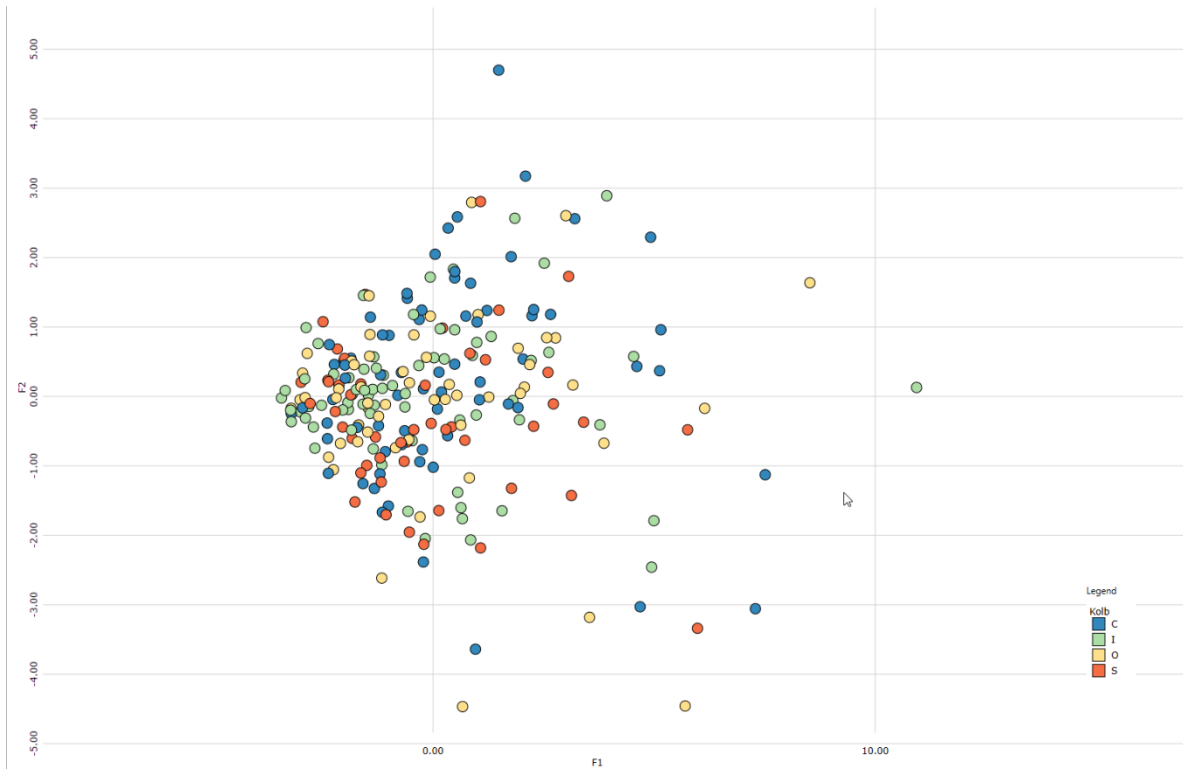


Figura 50. Diagrama de dispersión PCA 4 (Actividades) de las puntuaciones F1 y F2. El color se basa en los estilos de Kolb

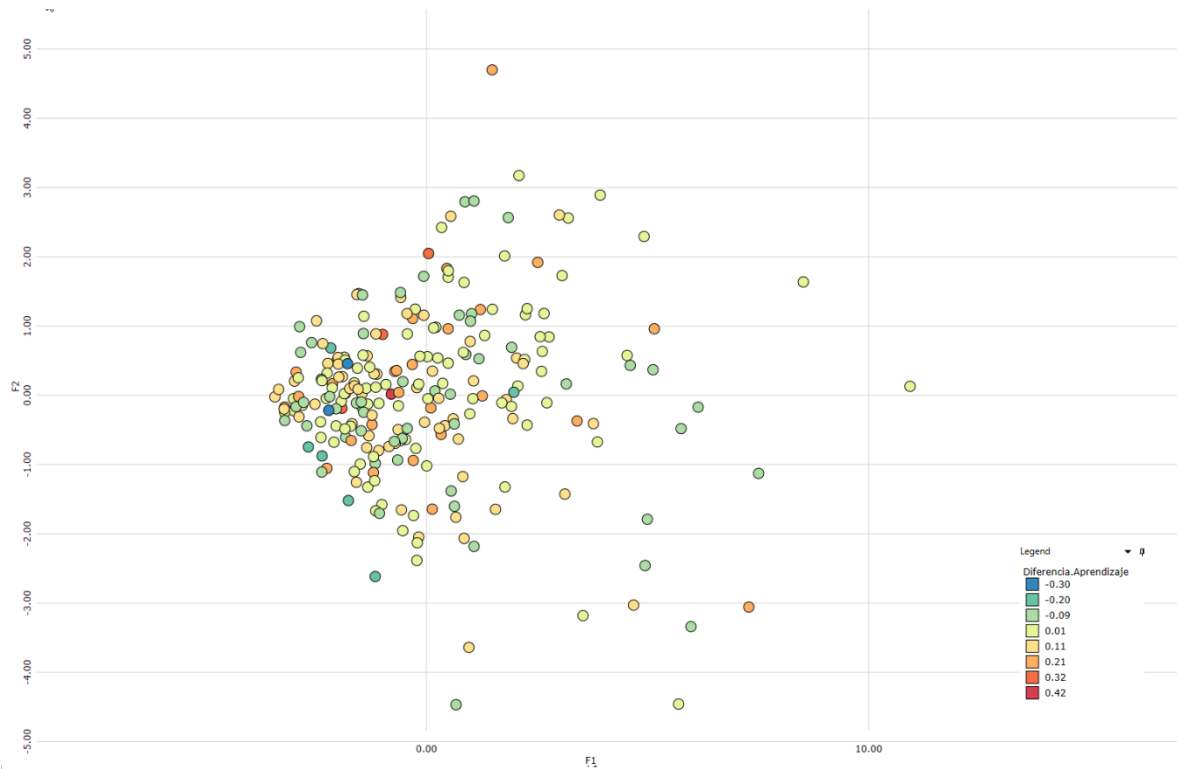


Figura 51. Diagrama de dispersión PCA 4 (Actividades) de las puntuaciones F1 y F2. El color se basa en la diferencia de aprendizaje

Para investigar más estos asuntos, presentamos la Tabla 14, que indica qué proporciones de los estilos de aprendizaje de Kolb tienen $F1 > 0$ y $F2 > 0$ indicando un nivel de actividad de la plataforma más alto.

Kolb Style	Count	% of Total	% of Kolb Style
C	25	10	35
I	15	06	21
O	13	05	24
S	7	03	15

Tabla 14. Conteo de muestras en el cuadrante de nivel de actividad más alto y porcentajes

Vemos que el perfil C tiene el 35% de todos los participantes en el cuadrante de alta actividad. El perfil O ocupa el segundo lugar con un 24% seguido del perfil I en el tercer lugar con un 21%.

En la Tabla 15 se muestra los promedios de F1 (medida de actividad general) por estilo de aprendizaje y género. Parece mostrar que las mujeres tienden a tener más actividad en la plataforma que los hombres y que el perfil C es con diferencia el perfil más activo de todos. El análisis ANOVA muestra que con $p = 0.483$ con $\alpha = 0.05$ las diferencias percibidas en la actividad por estilo de aprendizaje no son estadísticamente significativas, tampoco lo son las diferencias por género con $p = 0.5$

Avg of F1	Column Labels		
Row Labels	1	2	Grand Total
C	-0.2164	0.4912	0.3118
I	-0.0711	-0.5015	-0.2894
O	0.4303	-0.2376	0.0660
S	0.3731	-0.4836	-0.1088
Grand Total	0.1233	-0.0836	0

Tabla 15. Promedios F1 (medida de actividad general) por estilo de aprendizaje y género

¿Es esto relevante? Desafortunadamente la Figura 51 parece indicar que las actividades de aprendizaje -según lo registrado por la plataforma- no son relevantes para la diferencia de aprendizaje o los resultados de aprendizaje.

4.3 Rediseño del MPOC teniendo en cuenta los principios del Modelo de Estilos de Aprendizaje de Kolb

Teniendo en cuenta los resultados obtenidos con la versión original del curso “Astronomía cotidiana”, y buscando comprobar las consecuencias de la neutralidad pedagógica, se realizó un rediseño de algunos contenidos y actividades de aprendizaje que tuvieran en cuenta los principios del Modelo de Estilos de Aprendizaje de Kolb y su Modelo de Aprendizaje Experiencial descritos en el Capítulo 3. El curso se ofreció durante el primero periodo académico de 2017 nuevamente ofrecido con un cupo de 400 estudiantes.

Se aplicaron los mismos procedimientos para la construcción de un conjunto de datos con la información de interacción de los estudiantes con la plataforma Selene, información demográfica y académica obtenida a través de SIMCA, estilos de aprendizaje y los resultados de las evaluaciones semanales.

La Figura 52 muestra los resultados de una prueba ANOVA aplicada a los 251 estudiantes con un estilo de aprendizaje definido y que concluyeron el curso. Se concluye que no existe diferencia significativa en el rendimiento académico (expresado como el resultado acumulativo de las evaluaciones semanales) y el estilo de aprendizaje según Kolb. La diferencia encontrada con los resultados del curso anterior puede justificarse por el enfoque en el rediseño del curso para alcanzar su neutralidad pedagógica (Smith, Reed, & Jones, 2008).

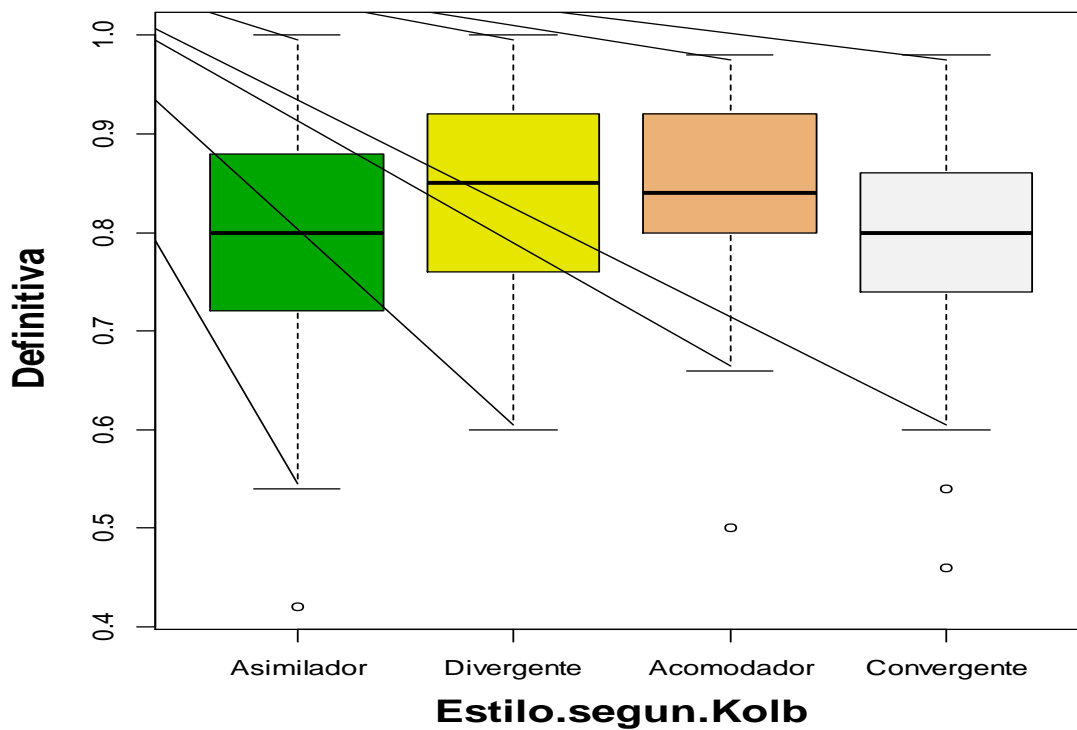


Figura 52. Rendimiento académico por estilo de aprendizaje según Kolb

Anova Table (Type III tests)

Response: datos2\$valores

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	257.321	1	327.2217	< 2e-16 ***
datos2\$F1	5.318	3	2.2541	0.08477 .
Residuals	110.093	251		

Esta situación puede verificarse en la Figura 53.

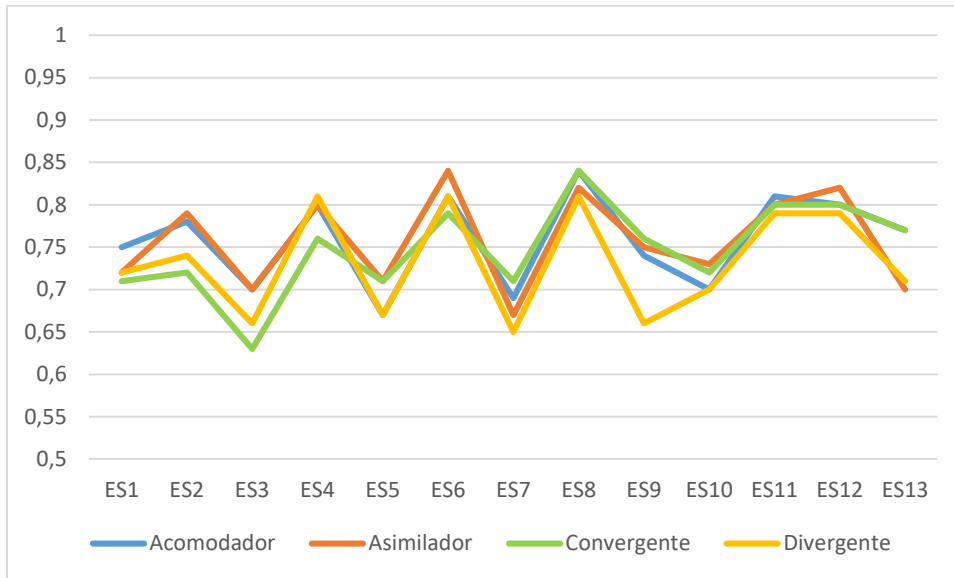


Figura 53. Rendimiento académico semanal por estilo en el MPOC rediseñado

Como se aprecia en la Figura 54, en el MPOC original –diseñado sin tener en cuenta los preceptos del Modelo de Estilos de Aprendizaje de Kolb-, los estilos Acomodador y Divergente obtuvieron consistentemente mejores calificaciones que los estilos Asimilador y Convergente en los exámenes semanales, excepto la semana 9 (donde los Divergentes bajaron su promedio) y la semana 10 (donde los Asimiladores obtuvieron el mejor promedio). Este comportamiento no se observa en el MPOC rediseñado.

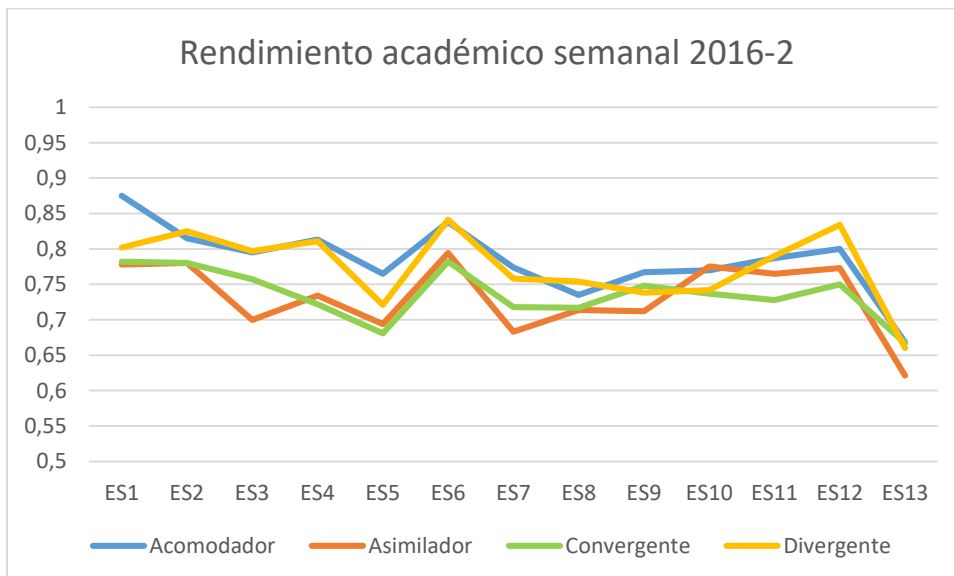


Figura 54. Rendimiento académico semanal por estilo en el MPOC inicial

4.4 Analíticas del aprendizaje como aproximación a un análisis de comportamientos de los estudiantes

La Figura 55 ilustra la cantidad de visualizaciones por día de los videos (el tipo principal de recurso educativo del MPOC) del curso Astronomía cotidiana (dataset con información del segundo periodo académico de 2016). Dado que todos los videos propios del curso están alojados en Youtube, se usó Google Analytics para la captura y generación de los reportes. Con claridad se observa un patrón de ingreso consistente en dos picos de visualizaciones a la semana. El máximo por semestre fue 1071 visualizaciones lo cual da una idea del nivel de consulta de los contenidos.



Figura 55. Visualizaciones videos del curso segundo semestre de 2016 (Fuente: Google Analytics)

La figura muestra además un periodo de baja actividad, aproximadamente de una semana a mediados de semestre. Este periodo corresponde a una semana a mediados de octubre cuando la instancia de Open edX de la Universidad del Cauca estuvo fuera de servicio por problemas índole técnica, lapso en cual no se programó el examen calificable respectivo.

Es claro que los estudiantes demuestran mayor actividad en el curso el día de la aplicación de los exámenes en línea, pues al menos, deben ingresar una vez de forma obligatoria para presentarlo. Google Analytics solamente da cuenta de las visualizaciones de video, no incluye otros servicios de Open edX empleados en el curso como la consulta a otros materiales ni los foros de discusión. Además, muestra resultados por día y no por periodos de tiempo más cortos.

Para subsanar esta dificultad se implementó una estrategia de seguimiento a las actividades de aprendizaje de los estudiantes cuyos resultados sobre el tiempo de ingreso a la plataforma se presentan en las Figuras 56 y 57. Para ello, se analizaron los datos de acceso al curso teniendo en cuenta no la fecha sino la hora y el día de la semana, para facilitar la búsqueda de los hábitos de ingreso. En el curso de 2016 se registraron 15.266 accesos de los estudiantes, por 15.719 en el curso del 2017.

El curso de segundo semestre de 2016 definió que los exámenes semanales podían presentarse los lunes y los jueves. Mientras los jueves los cuestionarios estuvieron

habilitados entre las 20:30 y las 21:30, los lunes estuvieron habilitados entre las 15:30 y la 16:30 pm si eran días laborales y entre las 12:30 y 13:30 si eran días festivos.

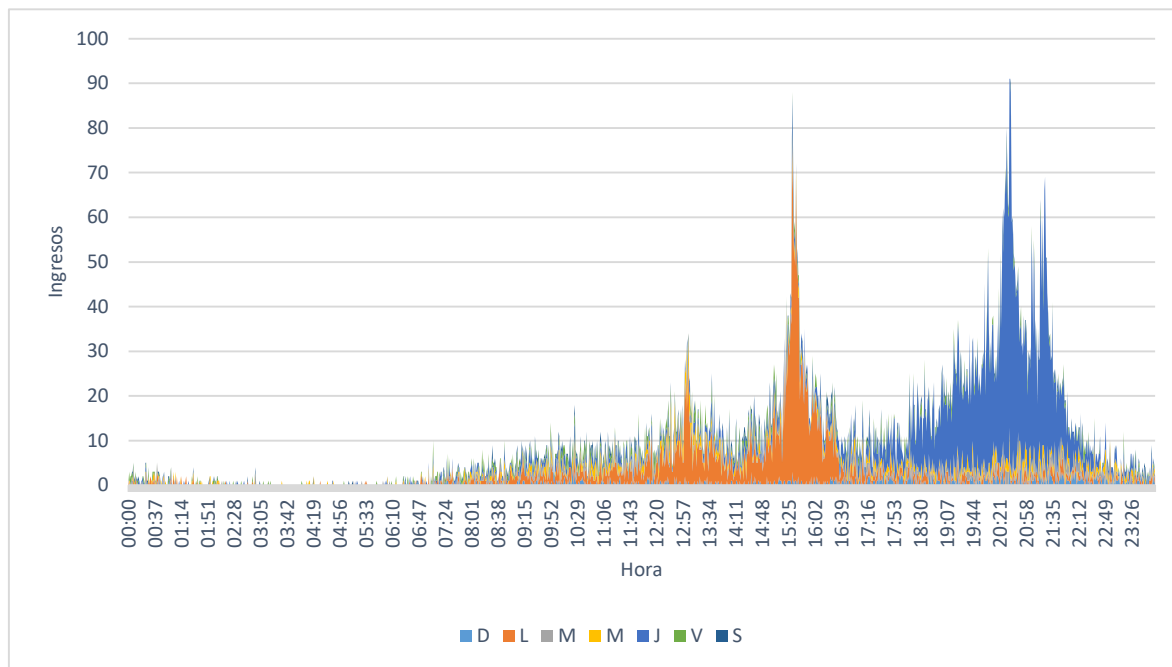


Figura 56. Ingresos al curso por estudiantes en el segundo semestre de 2016

Los datos evidencian picos de actividad que se corresponden con los horarios de presentación de los exámenes así: lunes 13:02 y 13:05 con 26 ingresos, lunes 15:31 con 72 ingresos y los jueves a las 20:36 con 88 ingresos con un segundo pico a las 21:25 horas con 62 ingresos. El total de ingresos a la plataforma para los horarios de los exámenes fue de 553, 1.272 y 2.424 respectivamente (3,63%, 8,38% y 15,91% de todos los ingresos al curso). Al consultarse a los estudiantes por sus motivos de conexión al curso antes de la finalización del tiempo para presentar el cuestionario de los jueves (21:25 horas), última oportunidad para presentar la evaluación del tema semanal, la justificación obtenida se relaciona con la verificación de que la plataforma del curso hubiera recibido correctamente las respuestas enviadas.

El curso de primer semestre de 2017 definió que los exámenes semanales podían presentarse los días martes jueves. Mientras los martes los cuestionaron estuvieron habilitados entre las 12:30 y las 13:30 horas, los jueves estuvieron habilitados entre las 20:30 y las 21:30 horas; decisión que facilitó la gestión de las evaluaciones y atención de los estudiantes por parte del equipo docente.

Los datos evidencian picos de actividad que se corresponde con los horarios de presentación de los exámenes así: lunes 12:31 con 51 ingresos y el jueves 20:31 con 195 ingresos. El total de ingresos a la plataforma para los horarios de los exámenes fue de 892 y 3463 respectivamente (5,67 % y 22,03% de todos los ingresos al curso).

Los datos analizados no reflejan un patrón de conexión asociado con la apertura de los contenidos de los temas semanales, como en el caso reportado en (Bachelet y Chaker, 2017). En un curso en línea masivo con reconocimiento académico, parece no ser prioridad en los estudiantes conectarse justo al momento cuando son liberados nuevos contenidos y actividades.

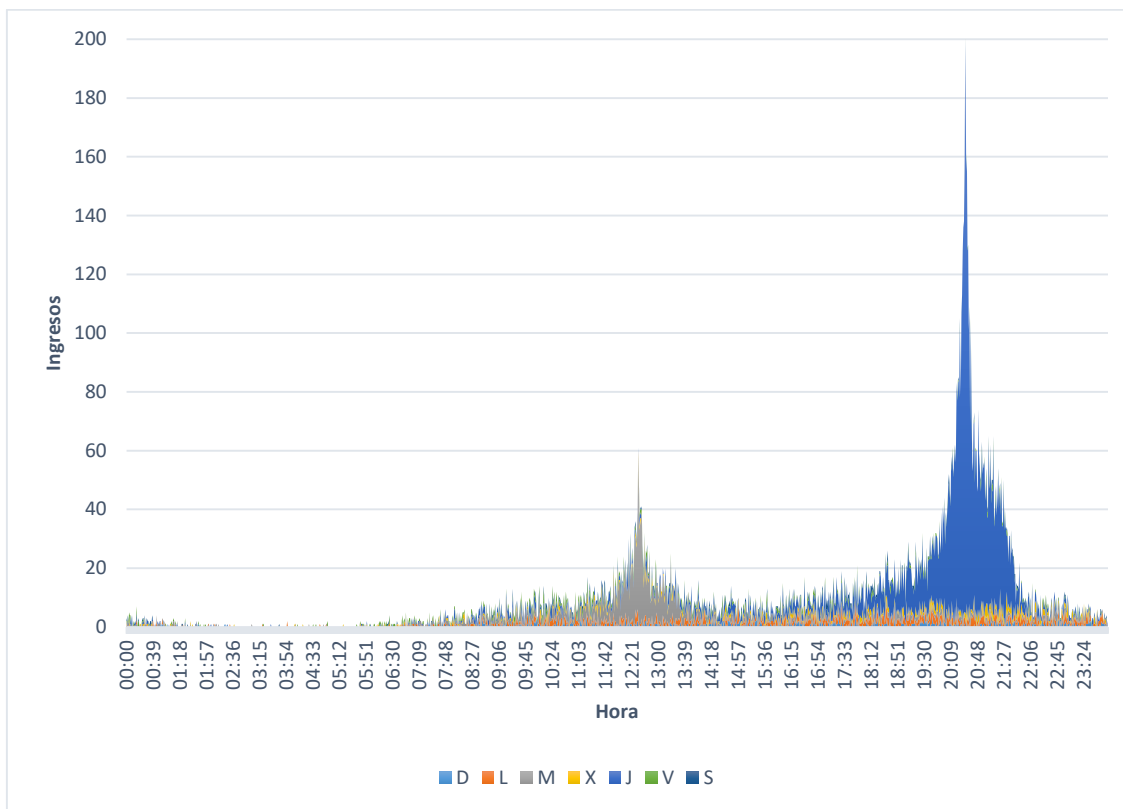


Figura 57. Ingresos al curso por estudiantes en el primer semestre académico de 2017

Que más estudiantes del curso de 2017 prefieran presentar el examen en el horario nocturno está relacionado con los horarios específicos de los estudiantes que se matricularon dicho semestre.

Ahora bien, conociendo los hábitos de ingreso y habiendo determinado que más de la cuarta parte de todos los accesos se realizan durante el lapso en el cual se encuentran habilitados los exámenes calificables, surge una pregunta básica por resolver: ¿los estudiantes que tienen más ingreso en horarios diferentes a los de las evaluaciones –generalmente previos a una prueba- demuestran mejor rendimiento académico?

Para resolver la pregunta se analizaron tres exámenes, uno por cada unidad temática del curso del primer semestre académico de 2017, codificados como U1T1, U2T4 y U3T3 donde el último dígito indica el número de la semana en la respectiva unidad temática. El rendimiento académico estuvo constituido por los resultados de la participación en la prueba en línea en una escala de 0 a 1, los estudiantes se clasificaron en cuartiles (QX) según la diferencia del número de ingresos realizados por cada tema en horario fuera de exámenes respecto a los ingresos durante del examen, intentando que las frecuencias de individuos en cada cuartil fueran aproximadas, como queda en evidencia en la Tabla 16.

Examen	Ingresos fuera de examen			
	QA	QB	QC	QD
U1T1	>8	8-6	5-3	<2
U2T4	>3	3-2	1	<0
U3T3	>4	4-3	2-1	<0

Tabla 16. Organización por cuartiles según ingresos previos a un examen (Fuente propia)

Para los análisis sólo se tuvieron en cuenta a los estudiantes que presentaron los exámenes en línea. Los resultados se exponen las Figura 58, 59, y 60.

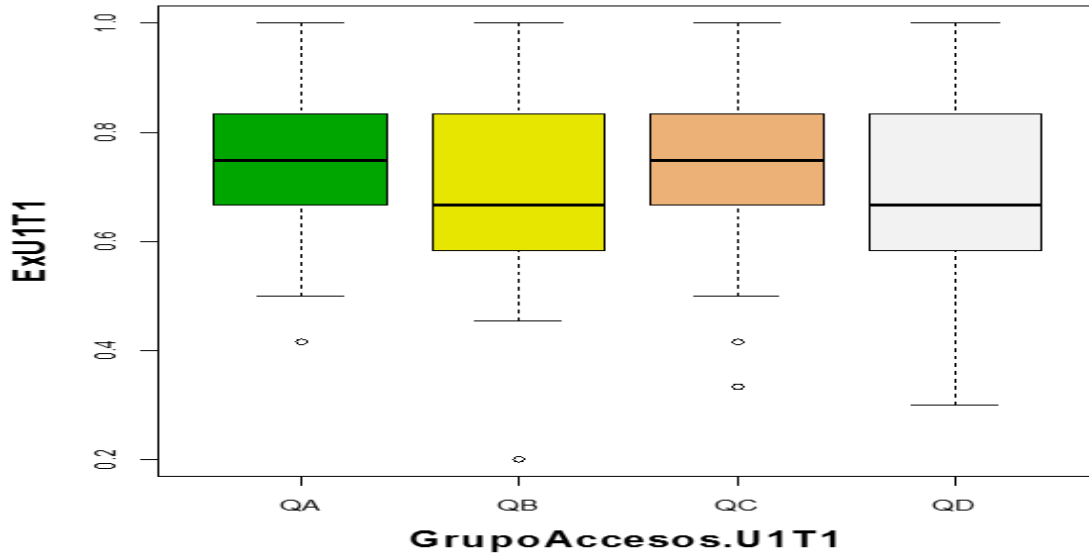


Figura 58. Resultados de evaluación U1T1 e ingresos fuera del examen

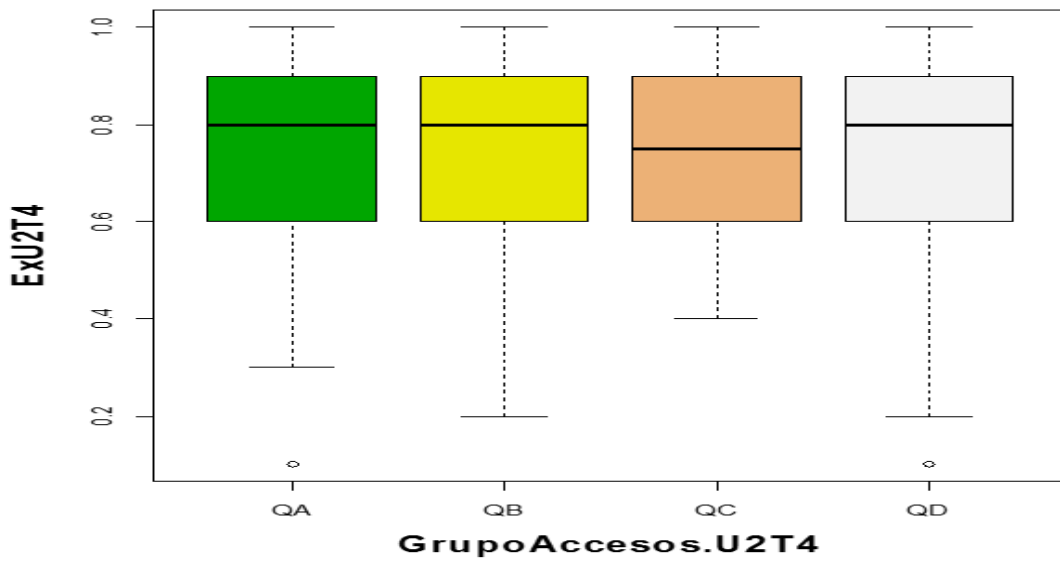


Figura 59. Resultados de evaluación U2T4 e ingresos fuera del examen

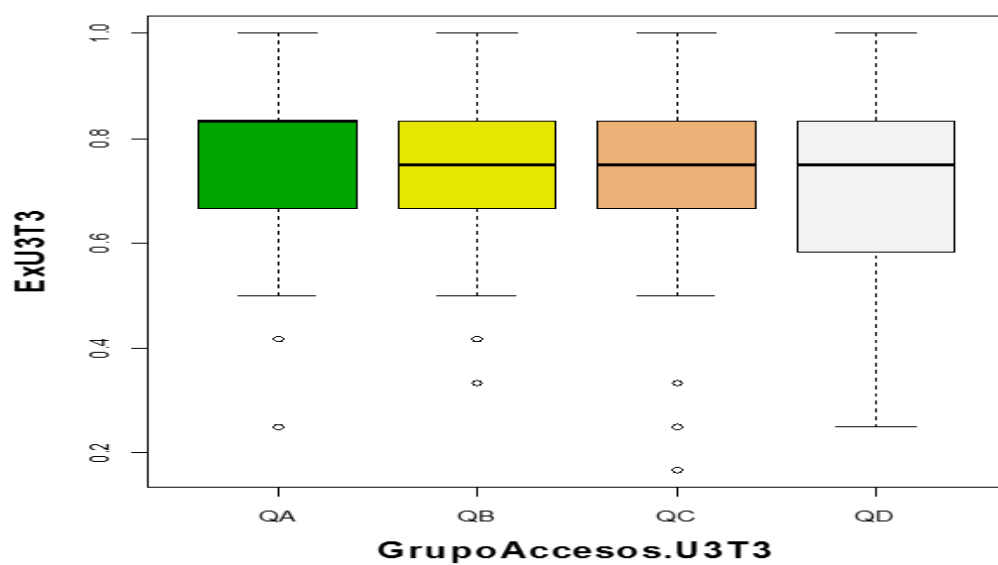


Figura 60. Resultados de evaluación U3T3 e ingresos fuera del examen

Con valores de p iguales a 0.3276, 0.998 y 0.09382 los análisis de varianzas demostraron sin lugar a dudas que no existe una relación directa entre el rendimiento académico y el número de ingresos a la plataforma web del MPOC Astronomía cotidiana previo al tiempo del examen, las medias de los grupos organizados en cuartiles según la Tabla 16 se encuentra entre el 80% y el 65% de la máxima calificación posible. Conectarse un mayor número de veces al curso antes de un examen no necesariamente implica un mejor rendimiento académico.

Una posible explicación a este hallazgo radica en que, por la modalidad del curso, se hace principalmente una evaluación de habilidades de conocimiento según la Taxonomía de Bloom (Ferraz & Belhot, 2010), de tal forma que, estando conectado al examen, los estudiantes pueden encontrar las respuestas con relativa facilidad navegando y buscando en Internet, sin la obligatoriedad de haber estudiado previamente los materiales del curso.

Finalmente, luego del resultado anterior, se pueden formular otras preguntas, entre ellas ¿qué variable de las medidas en la investigación tiene mejor correlación con el rendimiento académico?

Bajo el paraguas “actividad fuera del lapso del examen” teniendo datos acerca de: ingresos al curso, exámenes presentados (incluyendo auto-evaluaciones no calificables), visitas a páginas de videos, interacción con las funciones VCR del visor de videos, visitas a páginas de diapositivas, visita a contenidos externos, uso de simuladores, foros leídos y foros creados; los análisis de varianza aplicados al Tema 1 de la Unidad 1 (U1T1), indica que la cantidad de PlayVideo (clic en el botón para empezar a ver un video) es la variable que mejor correlación demuestra con el rendimiento académico. Como se evidencia en la ANOVA realizado a través de la aplicación R y en la Figura 30.

Para ello, se organizaron los estudiantes en seis grupos acorde a la Tabla 17.

Grupo	A	B	C	D	E	F
Resultado evaluación	1-0,9	0,9-0,8	0,8- 0,7	0,7- 0,6	0,6- 0,5	0,5- 0,1

Tabla 17. Organización de estudiantes según los resultados de la evaluación

En la Universidad del Cauca, un examen se aprueba con el 60% de aciertos, por tanto, los grupos A, B, C y D aprobaron el curso. Los estudiantes del grupo E estuvo a un 10% de aciertos de aprobar el examen y el grupo F estuvo a más del 10% de aciertos para aprobarlo.

Anova Table (Type III tests)

Response: datos2\$valores

	Sum Sq	Df	F value	Pr(>F)
(Intercept)	2027.37	1	164.9101	< 2e-16 ***
datos2\$F1	218.11	5	3.5482	0.00407 **
Residuals	3048.86	248		

4.5 Discusión y Conclusiones

El análisis estadístico realizado al primer dataset construido para curso virtual Astronomía cotidiana arrojó los siguientes resultados:

- Los resultados de la evaluación no se distribuyen normalmente. Están sesgados hacia un alto grado, muestra que el curso virtual no es particularmente difícil de aprobar.
- La Diferencias de aprendizaje sí se distribuyen normalmente.
- Los hombres superan ligeramente a las mujeres en el rendimiento académico.
- Los hombres tienen una Diferencia de aprendizaje mayor que las mujeres.
- Los perfiles Acomodador y Divergente del modelo de estilos de aprendizaje de David Kolb superan los perfiles Convergente y Asimilador en el rendimiento académico y en diferencia de aprendizaje
- No hay diferencia estadística en el rendimiento general del rendimiento académico o diferencia de aprendizaje por edad o estratos socioeconómicos.
- No hay diferencia estadística en el uso de la plataforma entre los diferentes estilos de aprendizaje del modelo de Kolb, con lo cual no se pueden esperar resultados significativos para la detección automática del estilo de aprendizaje en función den entrenamiento de un algoritmo de aprendizaje de máquina supervisado.
- No hay diferencias estadísticas en el uso de la plataforma según el género.

El diseño tradicional de MOOC –al menos en los datos compendiados en el Caso de Estudio- parece favorecer el aprendizaje expresado como resultados de las evaluaciones a unos estilos de aprendizaje más que otros. Con el rediseño del MPOC, teniendo en cuenta los principios del Modelo de Estilos de Aprendizaje según Kolb, no se apreciaron diferencias significativas en el rendimiento académico de los estudiantes, lo cual puede atribuirse a las nuevas características de los contenidos y actividades de aprendizaje.

El reconocimiento académico de cursos masivos en línea es un aspecto estratégico que debe abordar las instituciones de educación superior. Con un equipo docente limitado en número, conocer los hábitos de ingreso de una gran cantidad de estudiantes es vital para mejorar la eficacia de estas innovadoras alternativas educativas.

Para el caso de estudio que una cantidad significativa de los accesos a un curso en línea (27%) se realizan durante el tiempo que estuvieron habilitados los cuestionarios de evaluación, apenas dos horas en toda la semana. Esta información es importante, entre otras cosas, para poder dosificar adecuadamente el esfuerzo del personal docente y de apoyo que requieren los estudiantes de un curso masivo.

Los análisis realizados descubrieron también que no existe correlación entre el rendimiento académico y el hábito de tener una mayor cantidad de ingresos a un curso previo a un examen en línea, explicado por el tipo de evaluación realizada. Este hallazgo permite sugerir que la evaluación de este tipo de cursos debe realizarse en función de examinar habilidades de comprensión y aplicación según la taxonomía de Bloom que de conocimiento.

La investigación sí encontró diferencia estadística significativa entre el rendimiento académico y la cantidad de Plays a los videos de la unidad ejecutados previa a un examen, aunque la relación no es lineal. Una posible explicación es la existencia de conjuntos de estudiantes que con un menor esfuerzo de estudio pueden obtener mejores resultados en las evaluaciones que otros.

La investigación pone de manifiesto la importancia del diseño curricular, de contenidos, de actividades de aprendizaje y de la evaluación en relación con las características intrínsecas y extrínsecas de los estudiantes. Aunque se encontró que un importante porcentaje de estudiantes tienen acceso a Internet desde su casa alrededor del 90% (Figuroa, Solarte, Ramírez, 2018) las instituciones de educación superior y los propios gobiernos deben asegurar las condiciones de acceso y conectividad para la totalidad de la población, incluyendo la menos favorecida.

Capítulo 5

5.1 Analíticas del aprendizaje para predicción

En este apartado se intenta crear modelos para predecir las dos variables de interés que son los resultados del aprendizaje (Definitiva) y la Diferencia entre el desempeño académico en el curs en línea y el promedio académico de los estudiantes (Diferencia.Aprendizaje). El PCA 1 muestra que las variables se dividen en dos grupos, uno de ellos está relacionado con el desempeño de las calificaciones y el otro está relacionado con las actividades de aprendizaje registradas por la plataforma.

5.2 Relaciones causales y modelos predictivos

En este apartado se intenta crear modelos para predecir las dos variables de interés que son los resultados del aprendizaje (Definitiva) y la Diferencia entre el desempeño académico en el curs en línea y el promedio académico de los estudiantes (Diferencia.Aprendizaje). El PCA 1 muestra que las variables se dividen en dos grupos, uno de ellos está relacionado con el desempeño de las calificaciones y el otro está relacionado con las actividades de aprendizaje registradas por la plataforma.

En el Capítulo 4 se verificó que las variables de interés están relacionadas principalmente con el desempeño de las calificaciones globales. Se determinó que las variables de interés no están correlacionadas con las actividades de aprendizaje registradas por la plataforma. No obstante, se intentó crear modelos utilizando varias técnicas para gestionar la incertidumbre, que son el uso de modelos probabilísticos y el uso de los principales componentes como sustituto de los modelos multivariantes.

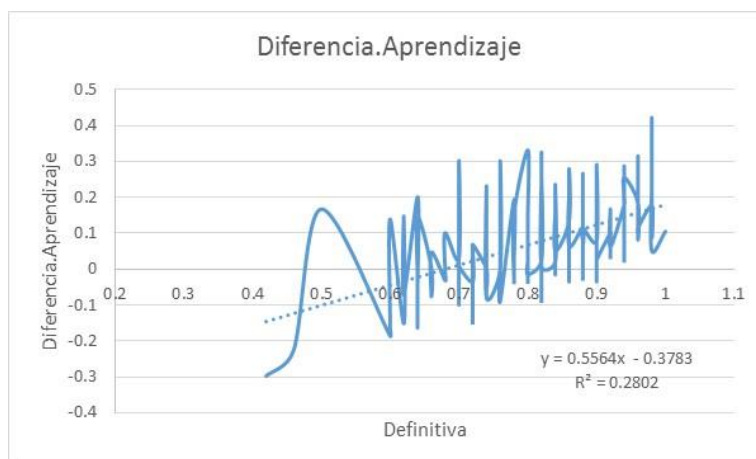


Figura 61. Desempeño académico vs Diferencia de Aprendizaje (en línea - presencial)

Como se verifico en el capítulo anterior, la calificación final está correlacionada con las variables de desempeño de la calificación. En este caso, no es necesario crear un modelo para ello, ya que la calificación final (Definitiva) es una combinación lineal de calificaciones anteriores, por lo tanto, este modelo es bien conocido. Lo interesante es la relación entre las dos variables de interés. Es decir, ¿el rendimiento académico en el curso en línea está relacionado con cierta diferencia de aprendizaje? La respuesta es sí, pero con una alta variabilidad como se muestra en la Figura 60. La estadística R^2 del modelo no es muy alta, sólo 0,2802, pero la tendencia lineal es clara.

En las figuras 62, 63, 64 y 65 se muestran los modelos que relacionan la calificación final con los accesos a la plataforma global (Accesos), los exámenes en línea presentados (Exámenes), el acceso a recursos web en línea (Acceso.Recursos.Web) y el acceso a páginas de vídeo (Acceso.Paginas.Video). Las estadísticas R^2 son todas muy bajas como se esperaba, con los valores $R^2=0,0341, 0,0131, 0,0199, 0,0217$ respectivamente.

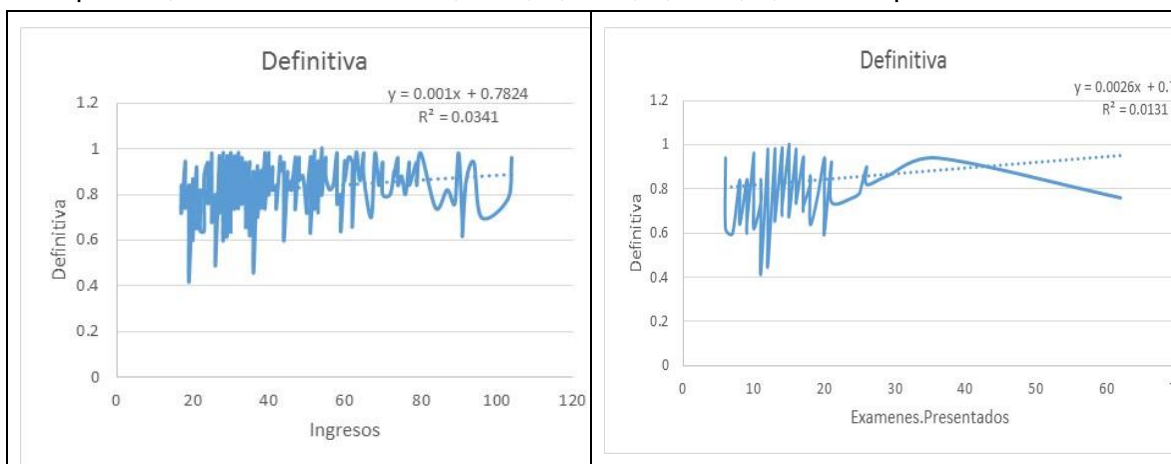


Figura 62. Desempeño académico vs Accesos a la plataforma

Figura 63. Desempeño académico vs Exámenes presentados

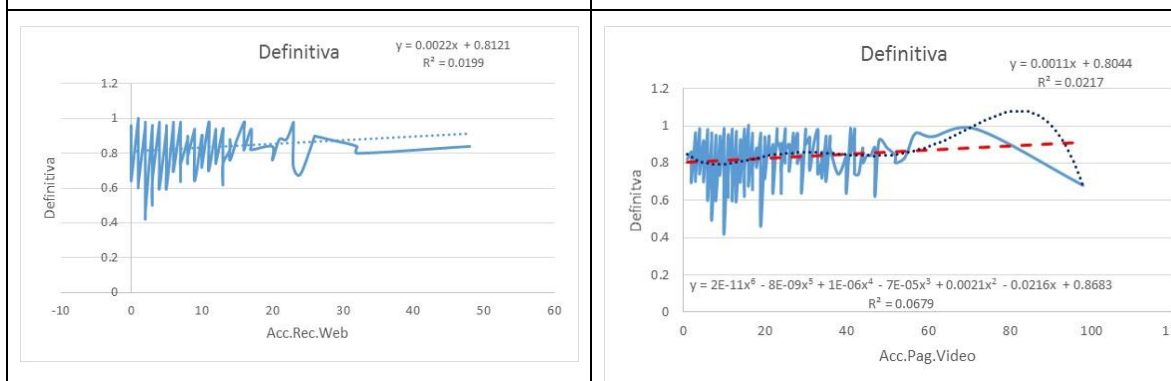


Figura 64. Desempeño académico vs Acceso a recursos en línea

Figura 65. Desempeño académico vs Acceso a recursos en video

En las figuras 66 y 67 se muestran modelos para la calificación final en función del acceso a talleres (Acceso.Practicas) y del total de actividades prácticas (Total.Actividades.Practicas). En ambos casos la estadística R^2 es baja $R^2=0.0194, 0.0234$ respectivamente.

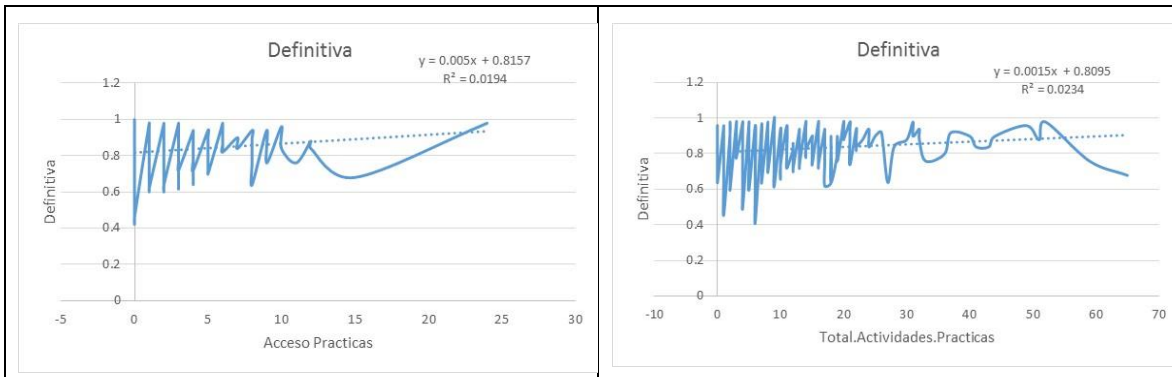


Figura 66. Desempeño académico vs Acceso a prácticas web

Figura 67. Desempeño académico vs Total actividades prácticas

Se concluye que no se puede derivar ningún modelo efectivo que nos ayude a predecir la calificación final en función de las actividades de la plataforma.

Se procedió posteriormente a encontrar modelos para predecir la variable de Diferencia en el aprendizaje. En las figuras 68, 69, 70 y 71 se muestran modelos de dicha variable (Diferencia.Aprendizaje) frente a exámenes presentados, acceso a actividades prácticas, videos reproducidos y actividades prácticas totales. Las estadísticas de R^2 también son aún más bajas: $R^2=0,0071$, $0,0014$, $0,0002$, $0,0006$ respectivamente.

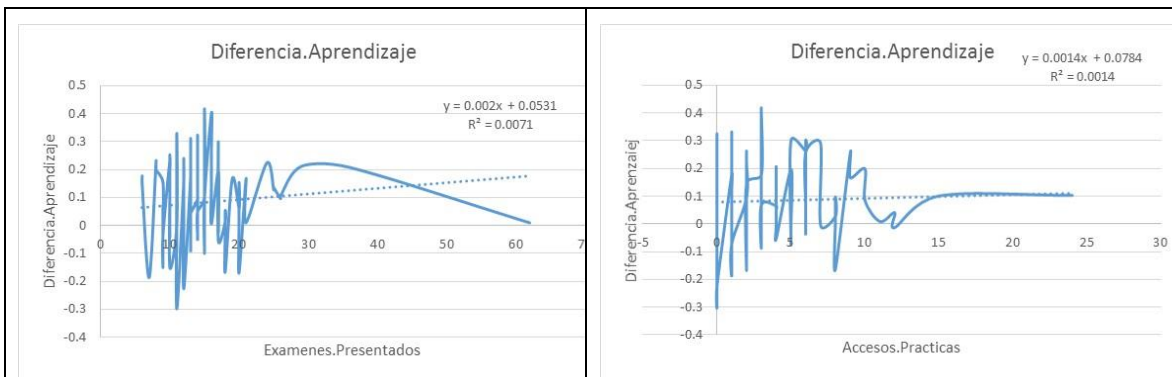


Figura 68. Diferencia de aprendizaje vs Exámenes presentados

Figura 69. Diferencia de aprendizaje vs Acceso a prácticas web

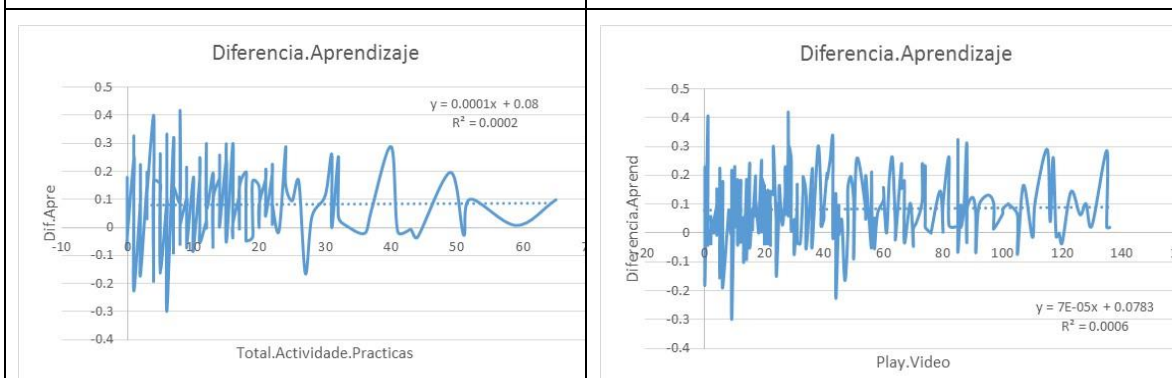


Figura 70. Diferencia de aprendizaje vs Total actividades prácticas

Figura 71. Diferencia de aprendizaje vs Videos vistos

Aunque algunas tendencias son evidentes en algunas de las figuras, la cantidad de variabilidad capturada por el modelo expresada por las estadísticas R^2 es demasiado baja. Para reducir la incertidumbre utilizamos la primera de dos técnicas: Modelos probabilísticos.

5.1.2 Gestión de la incertidumbre con modelos probabilísticos

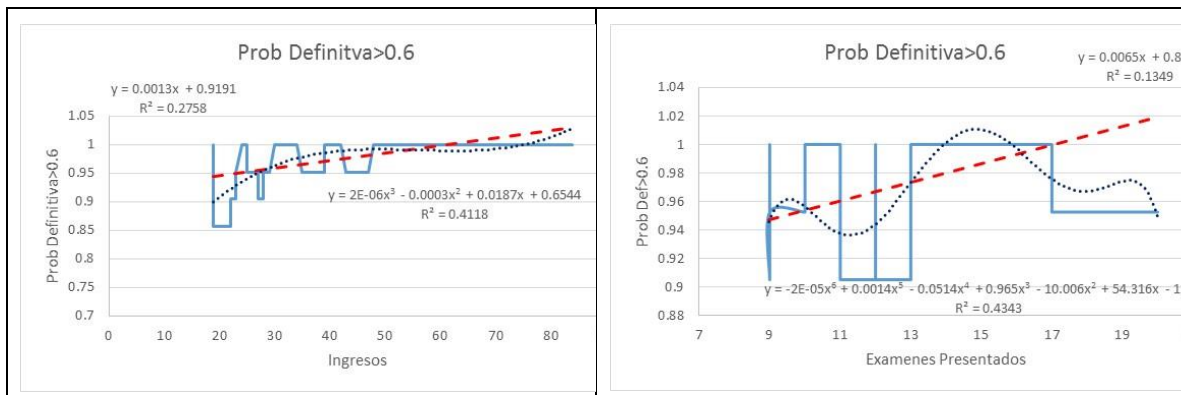
Cuando la estadística R^2 es baja, indica que el modelo sólo puede explicar una pequeña fracción de la variabilidad observada y, por lo tanto, que la relación entre las variables es débil. Por ello, para lograr una mejor visualización de las relaciones se introduce una nueva medida llamada **Probabilidad de éxito** o en este caso la probabilidad de que un estudiante tenga una calificación final superior a 0,6 (aprobar el curso) y que la diferencia de aprendizaje sea positive (que el rendimiento académico de un estudiante en cursos en línea sea mayor que su promedio académico en cursos presenciales).

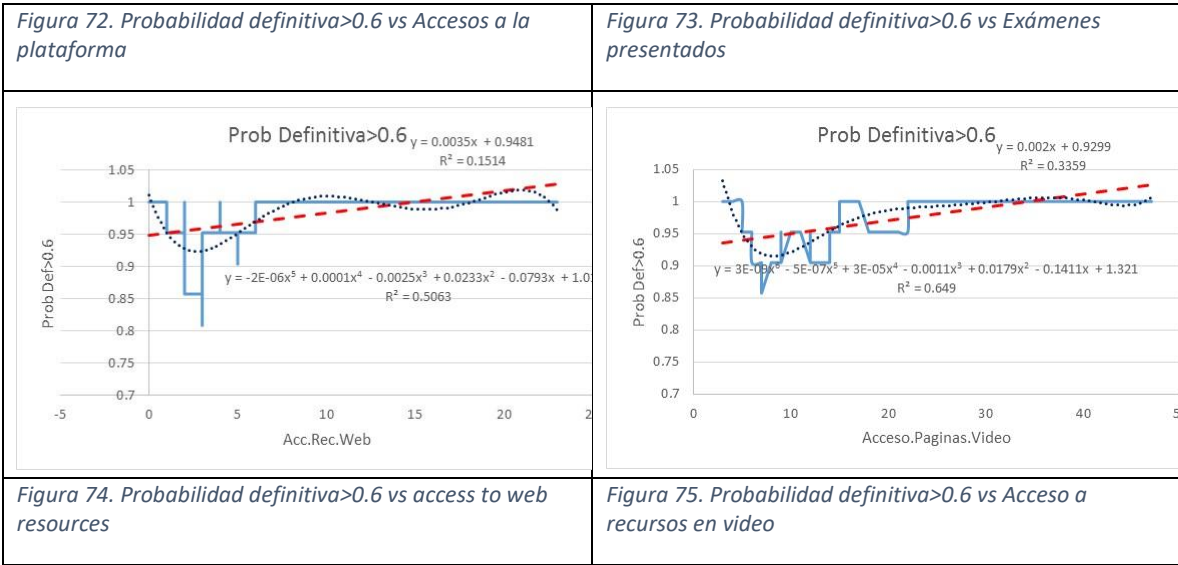
Para estimar esta probabilidad, después de ordenar los datos por eje de abscisas, se toma una muestra de n ítems de las variables de éxito por encima y por debajo de un valor dado. En general, se pierden n puntos de datos de los grupos, $n/2$ puntos al principio de conjunto de datos y $n/2$ al final. Se usó $n=21$. Se tomaron las muestras y luego se contó el número de aciertos. Se supuso que $A(i)$ es una medida que indica que el punto horizontal x_i es un éxito ($A(i)=1$) o no ($A(i)=0$). La probabilidad de éxito $P_s(i)$ en el punto horizontal i es estimada por:

$$P_s(i) = \frac{1}{n} \sum_{j=i-\frac{n}{2}}^{j=i+\frac{n}{2}} A(j)$$

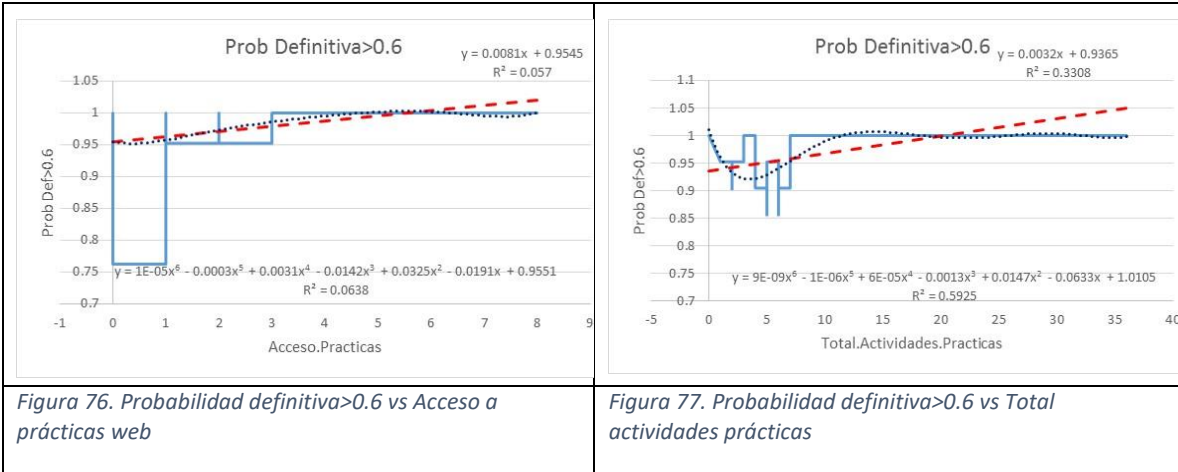
El tamaño de la muestra utilizada para estimar la probabilidad de éxito es pequeño, por lo que se realizaron pruebas con diferentes valores de n tales como 21, 31 y 41, y se encontró que el comportamiento de la estimación de probabilidad era aproximadamente el mismo, mostrando estabilidad en la estimación. Naturalmente, cuanto mayor sea la muestra, más precisa será la estimación.

Las Figuras 72, 73, 74 y 75 muestran la probabilidad de que la calificación final sea mayor o igual a 0.6 vs la frecuencia de accesos a la plataforma, exámenes en línea, acceso a recursos web y acceso a páginas de video.





Las Figuras 76 y 77 muestran la probabilidad de que la calificación final sea mayor o igual a 0,6 frente al acceso a las practices web y al acceso total a las actividades prácticas.



En las figuras 78, 79, 80 y 81 se muestran modelos de la probabilidad de que la diferencia de aprendizaje sea positiva frente a los exámenes en línea, el acceso a las practices web, los vídeos reproducidos y el total de actividades prácticas.

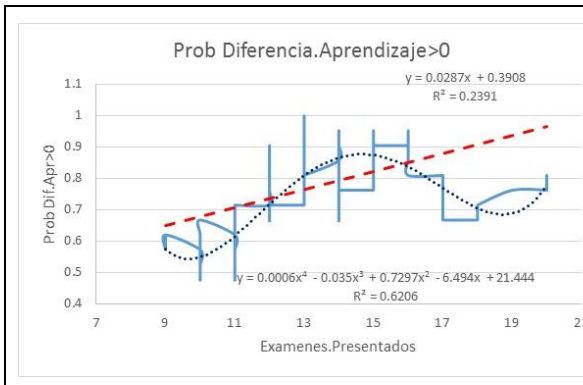


Figura 78. Diferencia de aprendizaje vs Exámenes presentados

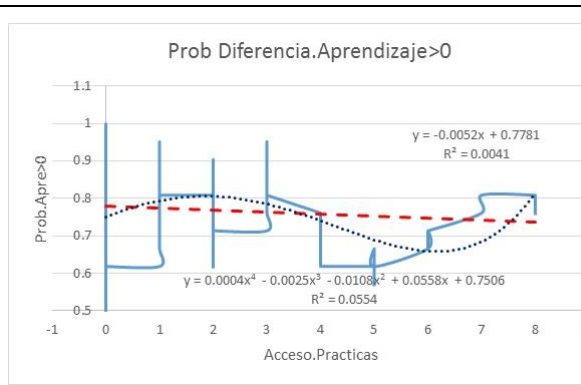


Figura 79. Diferencia de aprendizaje vs Acceso a prácticas web

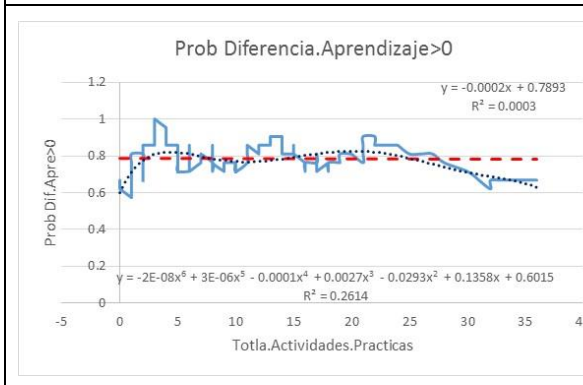


Figura 80. Diferencia de aprendizaje vs Total actividades prácticas

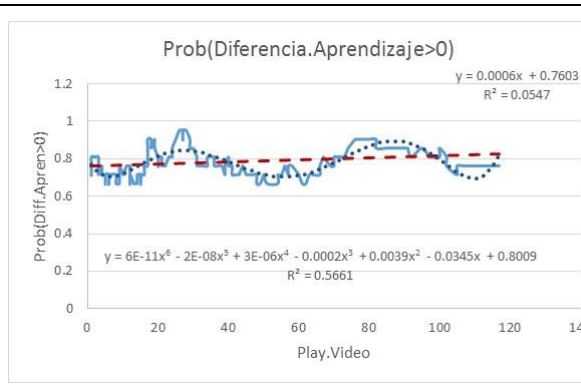


Figura 81. Diferencia de aprendizaje vs Videos vistos

Se puede observar como la estadística R^2 mejoró significativamente incluso cuando se utilizó un ajuste no lineal; como se puede ver en las Tablas 18 y 19.

Variable (R^2)	Desempeño académico	Probabilidad Desempeño académico > 0.6 (lineal)	Probabilidad Desempeño académico > 0.6 (non-linear)
Accesos a la plataforma	0.0341	0.2758	0.4118
Exámenes presentados	0.0131	0.1349	0.4343
Access to web resources	0.0199	0.1514	0.5063
Acceso a recursos en video	0.0217	0.3359	0.6490
Acceso a prácticas web	0.0184	0.0570	0.0638
Total actividades prácticas	0.0234	0.3308	0.5925

Tabla 18. Estadística R^2 para los modelos determinísticos de la Calificación final en comparación con los modelos probabilísticos

Variable (R^2)	Diferencia de aprendizaje	Probabilidad Diferencia de aprendizaje > 0 (Linear)	Probabilidad Diferencia de aprendizaje > 0 (Non-linear)

Exámenes presentados	0.0071	0.2391	0.6206
Acceso a prácticas web	0.0014	0.0041	0.0554
Total actividades prácticas	0.0002	0.0003	0.2614
Videos vistos	0.0006	0.0547	0.5661

Tabla 19. Estadística R^2 para los modelos determinísticos de la Diferencia de aprendizaje en comparación con los modelos probabilísticos

En las tablas anyeriores se destacan los modelos que lograron una R^2 superior a 0,5. Sólo los modelos no lineales probabilísticos fueron capaces de alcanzar $R^2 > 0,5$.

5.1.3 Gestión de la incertidumbre con los componentes principales

Otra manera de manejar la incertidumbre de las estadísticas bajas de R^2 es usar las puntuaciones de F1 como variable de entrada. Por lo tanto, se manejó la incertidumbre usando la combinación lineal de varias variables que es F1 del Análisis de Componentes Principales.

Los componentes principales se asocian con variables mediante la estadística de senos cuadrados. F1 captura la mayor variabilidad y, por lo tanto, está asociada con la mayor cantidad de variables. Así, la F1 de un PCA puede considerarse una medida global de las actitudes del estudiante (motivación, confianza), es decir, una combinación de las variables más importantes. Sin embargo, si algunas variables tienen patrones de medida que difieren significativamente de las variables asociadas principalmente con F1, pueden agruparse en otro componente principal, por ejemplo F2, con una correlación significativa. Por lo tanto, normalmente tomar el Componente Principal asociado con una variable particular en el eje horizontal mejorará la precisión de los modelos porque se toma en consideración más de una variable a través del componente principal.

En la figura 82 se muestra un modelo para la calificación final en función de las puntuaciones F1 para PCA 2, es decir, los datos de calificación. El modelo es muy preciso. Esto no es ninguna sorpresa ya que la calificación final es una combinación lineal de las calificaciones parciales del estudiante.

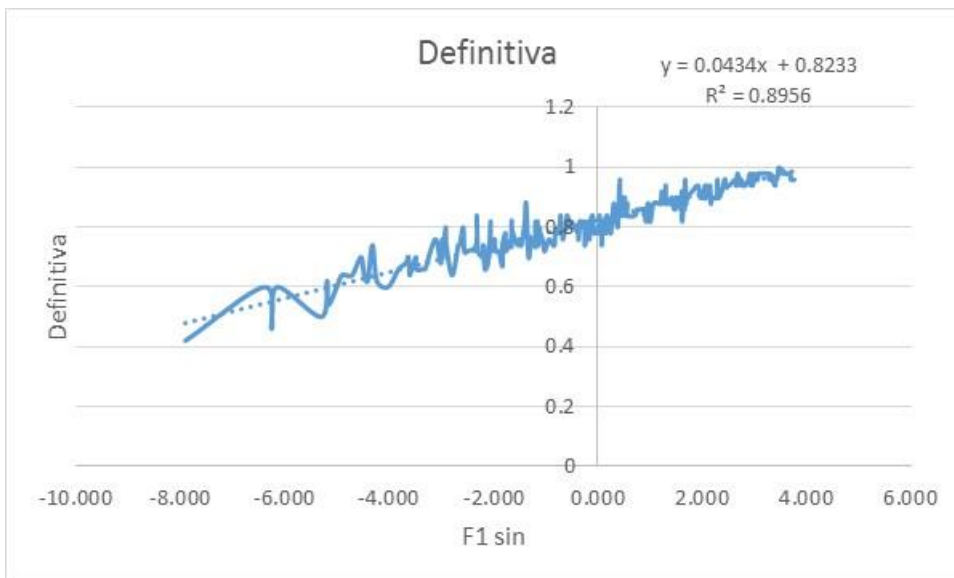
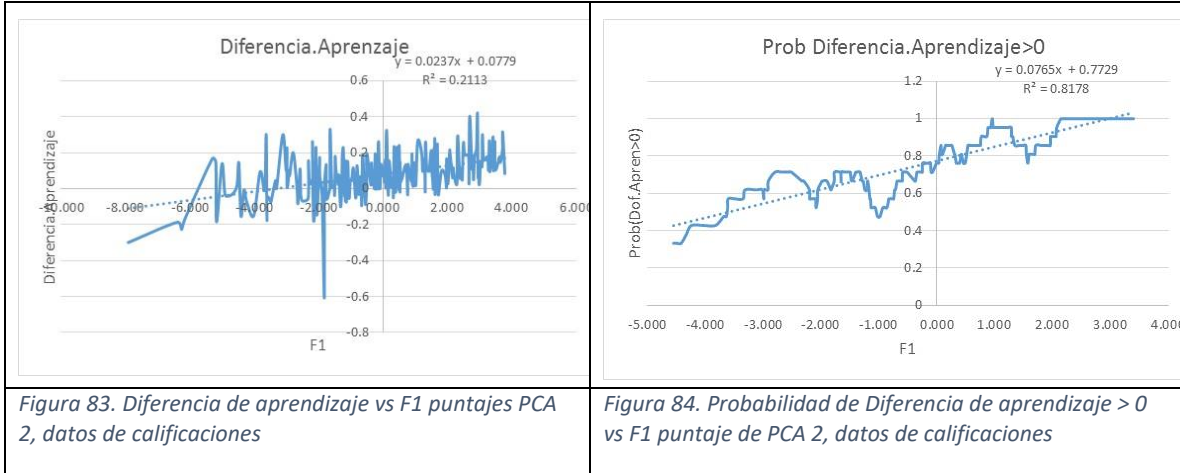
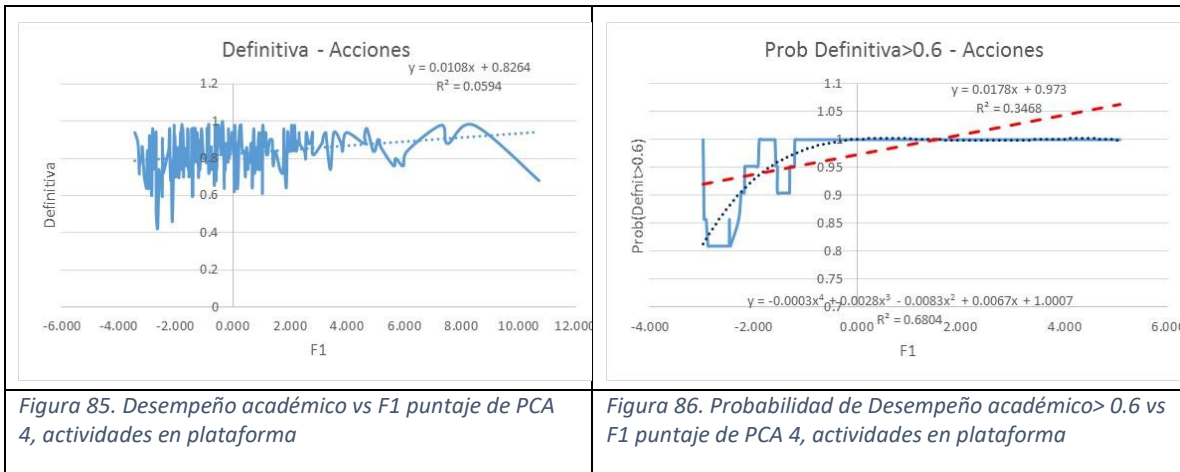


Figura 82. Desempeño académico vs F1 puntajes de PCA 2, datos de calificaciones

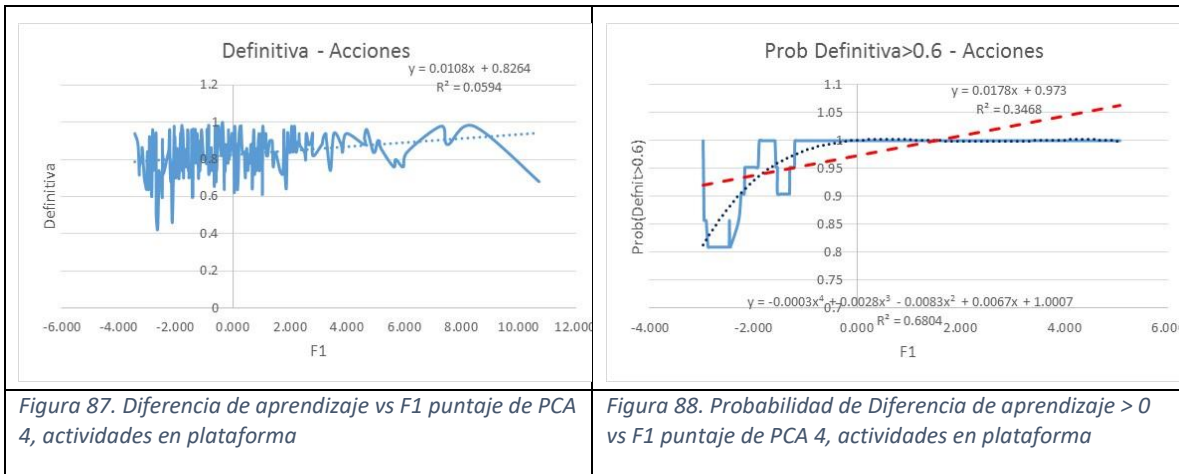
En la figura 83 se muestra un modelo de cambio en el aprendizaje frente a F1 de PCA 2. F1 no es un buen predictor del cambio en el aprendizaje. Sin embargo, el modelo probabilístico mostrado en la figura 84 mejora significativamente la variabilidad explicada por la estadística R^2 . Eso va de 0,2113 a 0,8178.



En las figuras 85 y 86 se intenta construir un modelo en el que se utiliza F1 de PCA 4, es decir, la actividad de la plataforma, para predecir la calificación final. El R^2 va de 0,0594 a 0,3468.



En las figuras 87 y 88 se intenta construir un modelo en el que F1 de PCA 4, es decir, la actividad de la plataforma, se utiliza para predecir la diferencia de aprendizaje. El R^2 no mejora significativamente al pasar de 0,0003 a 0,0359.



Con un proceso más refinado tanto en el diseño, ejecución y evaluación del curso, como en la recolección de información, se pueden conseguir niveles de predicción de las variables educativas con alto nivel de acierto. La Figura 89 ilustra la Probabilidad de aprobar el curso versus la Primera Componente Principal en la tercera ejecución del curso “Astronomía cotidiana” alcanzando un 94,6% de predicción acertada.

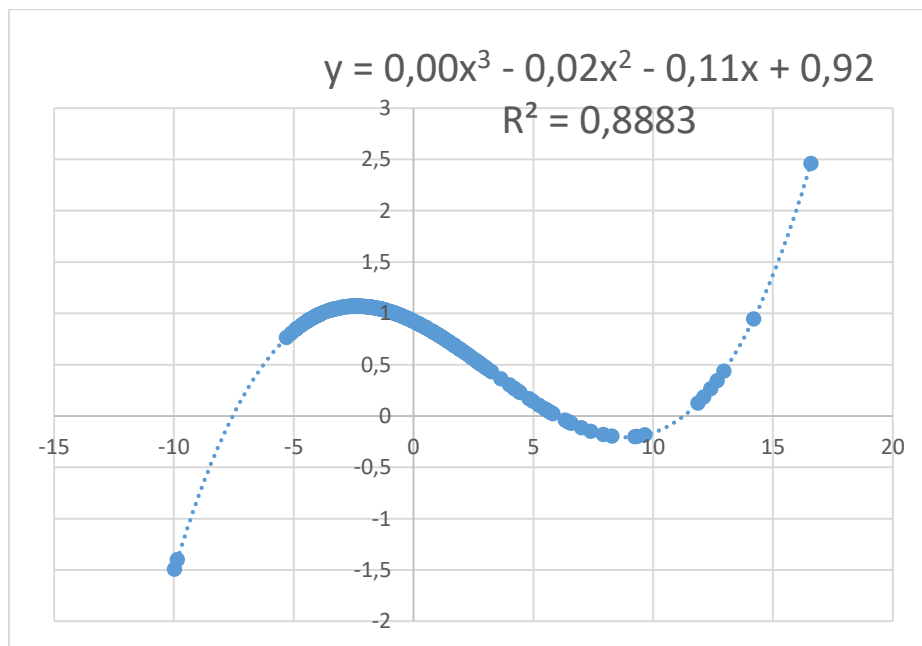


Figura 89. Probabilidad Aprobación del curso vs F1 en el Dataset 2

Con ese mismo conjunto de datos, la Primera Componente Principal es capaz de predecir con un 69,3% la aprobación del examen final del curso y un 76.1% Diferencias de aprendizaje negativas (quienes obtienen mejores resultados en las evaluaciones en cursos en modalidad virtual que en modalidad presencial). La Segundo Componente Principal es capaz de predecir Ganancias de aprendizaje positivas con el 84,4% de acierto.

El Análisis de Componentes Principales es una técnica tan potente que, en este mismo caso, la Decimosexta Componente Principal es capaz de predecir el 69,4% de aciertos respecto a los estudiantes que crean al menos tres foros a lo largo del curso.

5.2 Analítica del aprendizaje para la identificación automática del estilo de aprendizaje según Kolb

Para el entrenamiento de un algoritmo de máquina que permita la identificación automática del estilo del aprendizaje de los estudiantes, se dividió el segundo conjunto de datos construido en dos componentes:

- Dataset de Entrenamiento: un conjunto de datos balanceado con 36 instancias de cada estilo de aprendizaje según el Modelo de Kolb para un total de 144 instancias
- Dataset de Prueba: un conjunto de datos desbalanceado con el resto de instancias no usadas en el dataset de entrenamiento.

En el Anexo 2 se detalla el proceso de selección de las instancias que hacen parte del Dataset de Entrenamiento siendo que el que mejores resultados dio fue una selección de los mejores 36 resultados de aprendizaje por estilo, organizado de mayor a menor mediante el denominado Tfin, un examen realizado en la semana 15 del curso, que abarcaba todo el temario del curso.

La Tabla 20 resume los mejores resultados conseguidos a través del Entrenamiento de diversos tipos de algoritmos.

	Rules	Bayes	Trees	Meta	Lazy	Functions	Misc
	Decision Table	Bayes Net	J48	Interactive Classifier Optimizer	KStar	Multiplayer Perceptron	Imput Mapped Classifier
Acomodador	100%	100%	47%	94%	50%	36%	100%
Asimilador	0%	0%	25%	0%	31%	17%	0%
Convergente	47%	28%	33%	25%	17%	25%	47%
Divergente	8%	19%	31%	11%	31%	14%	8%
Total instancias correctas	56	53	49	47	46	33	30

Tabla 20. Resultado del entrenamiento de distintas categorías de algoritmos

Como era de esperarse por los resultados obtenidos a través del Análisis de Componentes Principales, los porcentajes de instancias correctamente clasificadas son bajos respecto al total del conjunto de datos, lo cual es una muestra clara de la dispersión del comportamiento de los diferentes estilos de aprendizaje según Kolb en las actividades y los servicios del caso de estudio.

Si bien Decision Table y Bayes Net alcanzaron el total de instancias correctamente clasificadas mayor, se puede apreciar que la clasificación realizada es muy desbalanceada,

propiciando situaciones en donde un estilo de aprendizaje (en este caso el Asimilador) no pudo clasificarse adecuadamente. Por ende, se decidió trabajar con los algoritmos basados en árbol para realizar la predicción del estilo de aprendizaje en el Dataset de Prueba, pues, además requirieron menos tiempo de procesamiento de máquina para entrenar el algoritmo.

La Tabla 21 ilustra que el algoritmo J48 responde mejor al entrenamiento que otros algoritmos similares.

	J48	Random Forest	Random Tree
Acomodador	47%	75%	53%
Asimilador	25%	17%	28%
Convergente	33%	19%	19%
Divergente	31%	28%	31%
Total instancias correctas	49	47	47

Tabla 21. Mejor entrenado mecanismo árbol

Weka (Hall et al, 2009) permite crear el modelo de entrenamiento del algoritmo J48 al cual se aplicó el Dataset de Prueba, cuyos resultados se aprecian en el siguiente modelo.

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
 Relation: DatasetPrueba36InstanciasPorEstilo-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Remove-R1-5,7-8
 Instances: 144
 Attributes: 14

- Grade
- Access
- Submitted.Exams
- Slides
- Page.Teacher
- Play.Video
- Stop.Video
- Completed.Videos
- Page.Web
- Practices
- Forums.Read
- Forums.Created
- Page.Support
- class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree

```

-----
Grade <= 0.77
| Page.Web <= 25
| | Grade <= 0.67
| | | Forums.Created <= 4
| | | | Page.Web <= 6: Assimilator (4.0/1.0)
| | | | Page.Web > 6: Converger (3.0)
| | | Forums.Created > 4: Assimilator (3.0)
| | Grade > 0.67
| | | Grade <= 0.72
| | | | Submitted.Exams <= 25
| | | | Grade <= 0.69: Converger (10.0/3.0)
| | | | Grade > 0.69
  
```

```

| | | | Page.Teacher <= 5: Converger (2.0)
| | | | Page.Teacher > 5: Diverger (12.0/1.0)
| | | Submitted.Exams > 25: Assimilator (2.0)
| | Grade > 0.72
| | | Foruns.Created <= 7
| | | | Access <= 40
| | | | Page.Web <= 5: Converger (6.0/1.0)
| | | | Page.Web > 5: Assimilator (3.0)
| | | | Access > 40
| | | | Foruns.Read <= 6: Diverger (2.0)
| | | | Foruns.Read > 6: Converger (6.0)
| | | Foruns.Created > 7
| | | | Foruns.Created <= 12: Assimilator (4.0)
| | | | Foruns.Created > 12: Diverger (2.0)
| Page.Web > 25: Diverger (5.0)
Grade > 0.77
| Page.Web <= 27
| | Play.Video <= 86
| | | Foruns.Created <= 5
| | | | Stop.Video <= 11
| | | | | Grade <= 0.8: Assimilator (4.0/1.0)
| | | | | Grade > 0.8
| | | | | Foruns.Read <= 12: Accomodator (3.0)
| | | | | Foruns.Read > 12: Diverger (2.0)
| | | | Stop.Video > 11
| | | | | Completed.Videos <= 7
| | | | | | Slides <= 18
| | | | | | | Slides <= 9: Accomodator (2.0)
| | | | | | | Slides > 9
| | | | | | | | Access <= 41: Accomodator (2.0)
| | | | | | | | Access > 41: Diverger (5.0)
| | | | | | | Slides > 18: Converger (4.0/1.0)
| | | | | Completed.Videos > 7
| | | | | | Foruns.Created <= 2: Accomodator (3.0)
| | | | | | Foruns.Created > 2
| | | | | | | Page.Teacher <= 21: Accomodator (2.0)
| | | | | | | Page.Teacher > 21: Converger (4.0)
| | | Foruns.Created > 5
| | | | Page.Web <= 20
| | | | | Page.Web <= 2: Assimilator (3.0/1.0)
| | | | | Page.Web > 2
| | | | | Submitted.Exams <= 22: Accomodator (12.0)
| | | | | Submitted.Exams > 22
| | | | | | Page.Teacher <= 17: Assimilator (2.0)
| | | | | | Page.Teacher > 17: Accomodator (4.0/1.0)
| | | | Page.Web > 20: Diverger (3.0)
| | Play.Video > 86
| | | Completed.Videos <= 13: Assimilator (7.0/1.0)
| | | Completed.Videos > 13: Accomodator (5.0)
| Page.Web > 27
| | Practices <= 25: Accomodator (2.0)
| | Practices > 25
| | | Page.Web <= 54
| | | | Play.Video <= 50: Assimilator (2.0)
| | | | Play.Video > 50: Converger (5.0)
| | | Page.Web > 54: Assimilator (4.0/1.0)

```

Number of Leaves : 35

Size of the tree : 69

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	49	34.0278 %
Incorrectly Classified Instances	95	65.9722 %
Kappa statistic	0.1204	

Mean absolute error 0.3311
 Root mean squared error 0.5382
 Relative absolute error 88.2256 %
 Root relative squared error 124.2036 %
 Total Number of Instances 144

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,472	0,222	0,415	0,472	0,442	0,240	0,665	0,403	Accomodator
	0,250	0,194	0,300	0,250	0,273	0,059	0,538	0,275	Assimilator
	0,333	0,287	0,279	0,333	0,304	0,044	0,550	0,258	Converger
	0,306	0,176	0,367	0,306	0,333	0,138	0,591	0,303	Diverger
Weighted Avg.	0,340	0,220	0,340	0,340	0,338	0,120	0,586	0,310	

=== Confusion Matrix ===

a b c d <-- classified as
 17 5 6 8 | a = Accomodator
 11 9 12 4 | b = Assimilator
 6 11 12 7 | c = Converger
 7 5 13 11 | d = Diverger

Estilo de aprendizaje según Kolb	Acierto en el clasificador	Total instancias
Acomodador	47%	36
Asimilador	25%	36
Convergente	33%	36
Divergente	31%	36

Tabla 22. Acierto en la predicción Dataset de Entrenamiento

Estilo de aprendizaje según Kolb	Acierto en la predicción	Total instancias
Acomodador	25%	64
Asimilador	0%	9
Convergente	57%	14
Divergente	25%	20

Tabla 23. Acierto en la predicción Dataset de Prueba

Total de instancias correctamente predichas: 27%.

El uso de técnicas de balanceo de conjuntos de datos de prueba (como por ejemplo el submuestreo) no mejoró significativamente los porcentajes de predicción.

Si bien el mecanismo construido alcanza un 57% de predicción correcta en el estilo de aprendizaje Convergente, en total la predicción del 27% mejora en apenas un 2% la probabilidad de acierto al azar.

Teniendo en cuenta que el tamaño del conjunto de datos es relativamente pequeño, se consideró la alternativa de crear subconjuntos en función del diseño curricular del curso. De esta manera, y centrando la prueba en la interacción semanal de los estudiantes en la Unidad Temática I donde el subconjunto de datos de una semana se constituyó como dataset de entrenamiento y el subconjunto de datos de la semana siguiente hizo las veces de dataset de prueba, se obtuvieron los resultados ilustrados en la Tabla 24.

J48 – Unidad Temática 1		Prueba				
		T2	T3	T4	T5	
Entrenamiento	T1	Acomodador	46%	46%	56%	57%
		Asimilador	33%	39%	30%	11%
		Convergente	21%	25%	24%	22%
		Divergente	20%	11%	11%	14%
		total	34%	33%	35%	32%
	T2	Acomodador		60%	52%	54%
		Asimilador		26%	17%	24%
		Convergente		16%	12%	12%
		Divergente		38%	38%	27%
		total		41%	35%	34%
	T3	Acomodador			61%	61%
		Asimilador			33%	13%
		Convergente			12%	14%
		Divergente			21%	20%
		total			38%	34%
	T4	Acomodador				42%
Asimilador					35%	
Convergente					14%	
Divergente					29%	
total					32%	

Tabla 24. Predicción del estilo de aprendizaje por semana de la Unidad Temática 1

El porcentaje de acierto de esta técnica (usando J48 para el clasificador) predijo acertadamente el estilo de aprendizaje entre un 29% y un 41%, lo cual es una mejora respecto a la predicción realizada con el conjunto de datos que comprendió las tres Unidades Temáticas de la asignatura.

5.3 Discusión y Conclusiones

En función de los modelos predictivos y el Análisis de Componentes Principales, se puede concluir lo siguiente:

- La frecuencia de las actividades de aprendizaje registradas sobre los servicios ofrecidos por instancia de Open edX en la Universidad del Cauca no puede ser utilizadas para crear modelos predictivos determinísticos para el Rendimiento académico o la Diferencia de aprendizaje.
- Las variables Acceso a recursos web, Acceso a recursos de vídeo, Acceso a talleres y Total actividades prácticas sí pueden ser utilizadas para crear modelos predictivos probabilísticos para la calificación final, con ello se pueden convertir en un insumo

importante para implementar un sistema de alertas tempranas respecto al rendimiento académico de los estudiantes que requieren mayor apoyo y acompañamiento en un curso virtual.

- Las variables Los exámenes en línea y los videos reproducidos pueden utilizarse para crear modelos predictivos probabilísticos para la Diferencia de aprendizaje.
- El primer componente principal de las actividades de aprendizaje realizadas en la plataforma, considerado una combinación lineal de varias variables, no puede utilizarse para crear modelos deterministas o probabilísticos para la calificación final o la diferencia de aprendizaje de los estudiantes.

A nivel general, no se encontraron diferencias significativas en el porcentaje de instancias correctamente clasificadas por los diversos tipos de algoritmos de aprendizaje de máquina probados (rules, bayes, tres, meta, Lazy, Funtions) según los organiza Weka, aplicando con conjunto de datos de entrenamiento obtenido a partir del segundo Dataset, el que incluye el rediseño del MPOC teniendo en cuenta el Modelo de Estilos de Aprendizaje de Kolb. Pero el porcentaje de aciertos es bajo, no supera el 39% en el mejor de los casos.

Se seleccionó el algoritmo J48 perteneciente a los métodos de árbol, pues fue el que tuvo un comportamiento más equilibrado respecto a la correcta clasificación de los cuatro estilos de aprendizaje del Modelo en mención. También un muy ligero mejor desempeño que otros algoritmos tipo árbol como Random Forest y Ramdom Tree. Pero definitivamente la predicción alcanzada es baja luego del entrenamiento del algoritmo. 57% de predicción correcta para el estilo de aprendizaje Convergente (la mejor entre los cuatro estilos) es baja respecto a lo encontrado en la literatura (Abdullah et al, 2015) (Hamtni, & Ateia, 2015).

Las características de la predicción están limitadas, de una parte, por la construcción del dataset que incluye aspectos tales como los servicios que ofrece la plataforma web que aloja el curso en línea y, de otra parte, la metodología de desarrollo de cada curso, propia de cada orientador y de cada institución. En los MPOC con reconocimiento académico, el limitado número de estudiantes admitidos es un factor en contra para alcanzar mejores niveles de precisión en la predicción. En tal sentido, no es sencillo mejorar la calidad de los dataset de entrenamiento y prueba pues cada cohorte tiene una ejecución propia no necesariamente consecuentes con las cohortes anteriores.

La Primera Componente Principal es buen predictor para:

- Aprobación de la calificación final del curso

Es un predictor aceptable para:

- Aprobación del examen final
- Ganancias de aprendizaje positivas
- Diferencia de aprendizaje positivas

Otras Componentes Principales pueden ser mejores predictores que F1

Como trabajo a futuro se propone incluir más variables de la interacción del estudiante con Open edX como por ejemplo número de sesiones por periodo académico o el tiempo conectado a plataforma. Dichas variables son particularmente difíciles de extraer o calcular en función de la información que se pudo extraer de la plataforma.

Otra alternativa es asumir un enfoque de Minería de Procesos como el utilizado en (Kizilcec, Pérez-Sanagustín, & Maldonado, 2017) para la identificación de patrones de navegación en la plataforma web y sus correlaciones con los estilos de aprendizaje.

Referencias bibliográficas

- Abdullah, M., Alqahtani, A., Aljabri, J., Altowirgi, R. & Fallatah, R. (2015) Learning style classification based on student's behavior in moodle learning management system. *Transactions on Machine Learning and Artificial Intelligence*, 3(1), 28.
- Adamopoulos, P. (2013). What makes a great MOOC? An interdisciplinary analysis of student retention in online courses. *AIS Electronic Library*.
- Ahmad, N. B., Ishak, M. K., Alias, U. F., & Mohamad, N. (2015). An Approach for E-Learning Data Analytics using SOM Clustering. *International Journal of Advances in Soft Computing & Its Applications*, 7(3).
- Alario Hoyos, C., Pérez Sanagustín, M., Cormier, D., & Delgado Kloos, C. (2014). Proposal for a conceptual framework for educators to describe and design MOOCs.
- Alario-Hoyos, C., Delgado Kloos, C., & Rayyan, S. (2016). Describing MOOC-based Hybrid initiatives: The H-MOOC Framework. European MOOCs Stakeholders Summit EMOOCs.
- Alshammari, M., Anane, R., & Hendley, R. (2014, July). Adaptivity in E-learning systems. In *Complex, Intelligent and Software Intensive Systems (CISIS)*, 2014 Eighth International Conference on (pp. 79-86). IEEE.
- Alshammari, M., Anane, R., & Hendley, R. J. (2015). Students' Satisfaction in Learning Style-Based Adaptation. In *Advanced Learning Technologies (ICALT)*, 2015 IEEE 15th International Conference on (pp. 55-57). IEEE.
- Alshammari, M., Anane, R., & Hendle, R. J. (2015). An e-learning investigation into learning style adaptivity. In *System Sciences (HICSS)*, 2015 48th Hawaii International Conference on (pp. 11-20). IEEE
- Armendariz, D., McHardy, Z. & García, D. (2014). OCTAL: Online Course Tool for Adaptive Learning. *ACM L@S'2014*.
- Bachelet, R., & Chaker, R. (2017, May). Toward a Typology of MOOC Activity Patterns. In *European Conference on Massive Open Online Courses* (pp. 134-139). Springer, Cham.
- Baer, L., Norris, D., Hill, A. & Brodnick, R. (2013). Crafting Transformative Strategies for Personalized Learning/Analytics. *ACM LAK'2013*.
- Bansal, N. (2013). *Adaptive recommendation system for MOOC*. Tesis de maestría Instituto Tecnológico de la India.
- Baraniuk, R. (2007). *Challenges and Opportunities for the Open Education Movement: A Connexions Case Study*. In T. Liyoshi & M.S. Vijay-Kumar (Eds.), *Opening up Education: The Collective Advancement of Education through Open Technology, Open Content, and Open Knowledge* (pp. 116-132). Cambridge: MIT Press
- Berlanga, A. & García F. (2004). *Sistemas hipermedia adaptativos en el ámbito de la educación*. Disponible en: <http://es.youscribe.com/catalogue/tous/recursos->

profesionales/sistemas-de-informacion/sistemas-hipermedia-adaptativos-en-el-ambito-de-la-educacion-1800783

Cabero-Almenara, J., Llorente-Cejudo, M. D. C., & Vázquez-Martínez, A. I. (2014). *Las tipologías de MOOC: Su diseño e implicaciones educativas*.

Campbell, L. (2004). Engaging with the Learning Object Economy. In A. LITTLEHORN (Ed.), *Reusing online resources: A Sustainable Approach to E-learning* (pp. 35-45). London: Routledge.

Castleberry, D., & Brandt, S. R. (2016). The Effect of Question Ordering Using Bloom's Taxonomy in an e-Learning Environment. In International Conference on Computer Science Education Innovation & Technology (CSEIT). Proceedings (p. 22). Global Science and Technology Forum.

Clark, D. (2013). *Adaptive MOOCs*. Disponible en: <http://colearnr-media.s3.amazonaws.com/education/moocs/adaptiveMOOC.pdf>.

Chang, R. I., Hung, Y. H., & Lin, C. F. (2015). Survey of learning experiences and influence of learning style preferences on user intentions regarding MOOCs. *British Journal of Educational Technology*.

Creed-Dikeogu, G. & Clark, C. (2013) Are You MOOC-ing Yet? A Review for Academic Libraries. *Kansas Library Association College and University Libraries Section Proceedings*.3 (1). Disponible en: <http://dx.doi.org/10.4148/culs.v1i0.1830>

Conole, G. (2013). Las pedagogías de los entornos personales de aprendizaje. *En L. Castañeda y J. Adell (Eds.), Entornos Personales de Aprendizaje: Claves para el ecosistema educativo en red* (pp. 185-188). Alcoy: Marfil.

Conole, G. (2014). A new classification schema for MOOCs. *The international journal for Innovation and Quality in Learning*, 2(3), 65-77.

Dans, E. (2009). *Online Education: Educational Platforms and the Openness Dilemma*. RUSC, 6(1), 22-30.

Daradoumis, T., Bassi, R., Xhafa, F. & Caballé, S. (2013). A review on massive e-learning (MOOC) design, delivery and assessment. In *P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), 2013 Eighth International Conference on* (pp. 208-213). IEEE

Delgado-Kloos, C., Muñoz-Merino, P., Muñoz-Organero, M., Alario-Hoyos, C., Pérez-Sanagustín, M., Parada, G. & Sanz, J. L. (2014). Experiences of running MOOCs and SPOCs at UC3M. In *Global Engineering Education Conference (EDUCON), 2014 IEEE* (pp. 884-891).IEEE.

Díaz, R., Solarte, M., Delgado, C. & Ramírez, G. (2016, September). Caracterización de videos para cursos en línea, abiertos y masivos. Conferencia TICAL 2016.

Dissanayake, D., Perera, T., Elladeniya, C., Dissanayake, K., Herath, S., & Perera, I. (2018). Identifying the learning style of students in MOOCs using video interactions. *International Journal of Information and Education Technology*, 8(3).

Downes, S. (2011). *MOOC 2011: The Massive Open Online Course in Theory and in Practice*. Disponible en: <http://www.slideshare.net/Downes/xmooc-the-massive-open-online-course-in-theory-and-in-practice>.

Duval, E. (2011). Attention please!: learning analytics for visualization and recommendation. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. 9-17

Ehlers, U. (2011). Extending the Territory: From Open Educational Resources to Open Educational Practices. *Journal of Open, Flexible and Distance Learning*, 15(2), 1-10.

Fasihuddin, H., Skinner, G. & Athauda, R. (2014). *Personalizing Open Learning Environments through the adaptation to Learning Styles*.

Fasihuddin, H., Skinner, G., & Athauda, R. (2015). Towards adaptive open learning environments: Evaluating the precision of identifying learning styles by tracking learners' behaviours. *Education and Information Technologies*, 1-19.

Ferraz, A. P. C. M., & Belhot, R. V. (2010). Taxonomia de Bloom: revisão teórica e apresentação das adequações do instrumento para definição de objetivos instrucionais. *Gest. Prod., São Carlos*, 17(2), 421-431.

Fidalgo, A., García, F. & Seins, M. (2013). A methodology proposal for developing *Adaptive cMOOC*. *ACM TEEM'13*.

Figuerola, C., Solarte M. & Ramírez-Gonzalez, G. (2018). Evolución del perfil del estudiante de pregrado de la Universidad del Cauca que matricula cursos en línea con reconocimiento académico.

FORBES (2012). *Coursera's Huge Online Classes Roar Into Brazil, India and China*. Disponible en: <http://www.forbes.com/sites/georgeanders/2012/08/09/courserashuge-online-classes-roar-into-brazil-india-and-china>.

Fournier, H., Kop, R., & Sitlia, H. (2011). The value of learning analytics to networked learning on a personal learning environment. In *Proceedings of the 1st International Conference on Learning Analytics and Knowledge* (pp. 104-109). ACM.

Fox, A. (2013). From MOOCs to SPOCs. *Communications of the ACM*, 56(12), 38-40.

Ghazali, A., Noor, S. & Saad, S. (2015). Review of personalized learning approaches and methods in e-learning environment. In *Electrical Engineering and Informatics (ICEEI)*, 2015 International Conference on (pp. 624-627). IEEE.

Guo, P. J., Kim, J., y Rubin, R. (2014). How video production affects student engagement: An empirical study of mooc videos. *Memorias Proceedings of the first ACM conference on Learning@ scale conference*, ACM, pp.41-50.

Guo, W. (2014, October). From SPOC to MPOC--The Effective Practice of Peking University Online Teacher Training. In *Educational Innovation through Technology (EITT)*, 2014 International Conference of (pp. 258-264). IEEE.

Grünewald, F., Meinel, C., Totschnig, M. & Willems, C. (2013). Designing MOOCs for the

support of multiple learning styles. In *Scaling up learning for sustained impact* (pp. 371-382). Springer Berlin Heidelberg.

Guárdia, L., Maina, M. & Sagrá, A. (2013). MOOC Design Principles. *A Pedagogical Approach from the Learner's Perspective*.

Hernández, S. (2013). *SCOPEO informe No. 2: MOOC Estado de la situación actual, posibilidades, retos y futuro*. Disponible en: <http://scopeo.usal.es/wp-content/uploads/2013/06/scopeoi002.pdf>.

Hamtini, T., & Ateia, H. (2015). A proposed dynamic technique for detecting learning style using literature based approach. In *Applied Electrical Engineering and Computing Technologies (AEECT)*, 2015 IEEE Jordan Conference on (pp. 1-6). IEEE.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.

Honey, P. & Mumford, A. (2000). *The learning styles helper's guide*. Maidenhead, Berkshire: Peter Honey.

Hung, Y., Chang, R. & Lin, C. (2016). Hybrid learning style identification and developing adaptive problem-solving learning activities. *Computers in Human Behavior*, 55, 552-561.

Jaramillo, D. & Solarte, M. (2016). Architectural approach for automatic follow up of learning activities in massive open online courses. *Sistemas & Telemática*, 14(37), 57-72.

Jaramillo, D. (2017). Seguimiento automático de actividades de aprendizaje en cursos en línea abiertos y masivos. Tesis de Maestría, Universidad del Cauca.

Jolliffe, I.T.: *Principal Component Analysis*. Springer, Heidelberg (2002).

Keefe, J. (1988). *Profiling & Utilizing Learning Style*. NASSP, 1904 Association Drive, Reston, VA 22091-1578.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & education*, 104, 18-33.

Kolb, D. A. (2014). *Experiential learning: Experience as the source of learning and development*. FT press.

Kruchten, P., Obbink, H., & Stafford, J. (2006). The past, present, and future for software architecture. *IEEE software*, 23(2), 22-30.

Letón, E., Luque, M. & Molanes, E. (2013). *Cómo diseñar un MOOC basado en mini-videos docentes modulares*. Disponible en: http://www.ia.uned.es/minivideos/publicaciones/2013_el_etal_CIE_v2.pdf

Liyanagunawardena, T., Adams, A. & Williams, S. (2013). MOOCs: A systematic study of the published literature 2008-2012. *The International Review of Research in Open And Distributed Learning*.

Mackness, M. Williams, R. (2010). *The ideals and reality of participating in a MOOC*. Disponible en: <http://www.lancs.ac.uk/fss/organisations/netlc/past/nlc2010/abstracts/PDFs/Mackness.pdf>

Mahraj, K. (2012). Using information expertise to enhance massive open online courses. *Public Services Quarterly*, 8, 360-368.

Manochehr, N. N. (2006). The influence of learning styles on learners in e-learning environments: An empirical study. *Computers in Higher Education Economics Review*, 18(1), 10-14.

Martin, F. (2012). Education: will massive open online courses change how we teach?. *Communications of the ACM*, 55(8), 26-28.

McAuley, A., Stewart, B., Siemens, G. & Cormier, D. (2010). *The MOOC model for digital practice*. Disponible en: https://oerknowledgecloud.org/sites/oerknowledgecloud.org/files/MOOC_Final_0.pdf

Morales, M., Hernández, R. & Gutl, C. Telescope, a MOOCs initiative in Latin America: Infrastructure, best practices, completion and dropout analysis. *Frontiers in Education Conference (FIE), 2014 IEEE, España*, pp. 1-7, 2014.

Neubaum, G., Wichmann, A., Eimler, S. C., & Krämer, N. C. (2014, August). Investigating Incentives for Students to Provide Peer Feedback in a Semi-Open Online Course: An Experimental Study. In *Proceedings of The International Symposium on Open Collaboration* (p. 19). ACM.

Onah, D. F., & Sinclair, J. (2015). Learners expectations and motivations using content analysis in a MOOC. In *EdMedia 2015-World Conference on Educational Media and Technology* (Vol. 2015, No. 1, pp. 185-194). Association for the Advancement of Computing in Education (AACE).

Olson, D. & Delen, D. (2008). Advanced data mining techniques. *Springer Science & Business Media*.

Pappano, L. (2012). The Year of the MOOC. *The New York Times*, 2(12), 2012.

Parrish, P. E. (2004). The trouble with learning objects. *Educational Technology Research and Development*, 52(1), 49-67.

Possos, W. & Cruz, R. (2017). Dataset para la Clasificación de Comportamientos Sedentarios en Entornos Cerrados. Trabajo de grado. Universidad del Cauca.

Proyecto Erasmus+ MOOC-Maker Construction of Management Capacities of MOOCs in Higher Education. <http://mooc-maker.org>

Rimbaud, Y., McEwan, T., Lawson, A. & Cairncross, S. (2013). *Adaptive Learning in Computing for non-native Speakers*.

Roig, R., Mengual, S. & Suarez, C. (2014). Evaluación de la calidad pedagógica de los MOOC. *Profesorado*, 18(1) 27-41.

Sánchez, M. & Prendes, M. (2015). Más allá de las pruebas objetivas y la evaluación por pares: alternativas de evaluación en los MOOC. *Universities and Knowledge Society Journal* (pp. 119-131)

Siemens, G. (2004). *Connectivism: A Learning Theory for the Digital Age*. Disponible en: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.87.3793&rep=rep1&type=pdf>

Siemens, G., Gasevic, D., Haythornthwaite, C., Dawson, S. P., Shum, S., Ferguson, R., & Baker, R. (2011). Open Learning Analytics: an integrated & modularized platform.

Smith, B., Reed, P., & Jones, C. (2008). 'Mode Neutral' pedagogy. *European Journal of Open, Distance and E-learning*, 11(1).

Solarte, M. (2015). *Adaptación y personalización en MOOC: una revisión*. (Proyecto Doctoral Contribuciones a la identificación automática del estilo de aprendizaje del estudiante en procesos formativos en línea masivos) Popayán: Universidad del Cauca.

Solarte, M. (2016). Proyecto MOOC Maker Informe sobre reconocimiento de créditos aplicable a MOOC entre IES en América Latina y Europa.

Solarte, M. & Ramírez, G. (2016,). ¿Cómo evalúan los estudiantes la experiencia de un curso virtual de la Universidad del Cauca? VII Coloquio Internacional de Educación

Solarte, M., Ramírez-Velarde, R., Alario-Hoyos, C., Ramírez-González, G., & Ordóñez-Eraso, H. (2018). Kolb's Learning Styles, Learning Activities and Academic Performance in a Massive Private Online Course. In *Mexican International Conference on Artificial Intelligence* (pp. 327-341). Springer, Cham.

Sonwalkar, N. (2013). *The First Adaptive MOOC: A Case Study on Pedagogy Framework and Scalable Cloud Architecture—Part I*. MOOCs FORUM. 22-29. Disponible en: <http://online.liebertpub.com/doi/abs/10.1089/mooc.2013.0007>

Sunar, A. S., Abdullah, N. A., White, S., & Davis, H. C. (2015). *Personalisation of MOOCs: the state of the art*.

Torres, J., Infante, A. & Díaz, P. (2014). MOOCs and personalized massification. *Profesorado*.18(1).63-72.

Tovar, E. & Dimovska, A. (2013). *OCW-S: enablers for building sustainable Open Education*. Madrid: Universidad Politécnica de Madrid.

Wiley, D. (2012). *The MOOC Misnomer*. Disponible en <http://opencontent.org/blog/archives/2436>

Williams, S. & Adams, A. (2013). The Impact and Reach of MOOCs: A Developing Countries' Perspective. *eLearningpapers*33.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. In *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining* (pp. 29-39). Citeseer.

Xu, D., Huang, W. W., Wang, H., & Heales, J. (2014). *Enhancing e-learning effectiveness using an intelligent agent-supported personalized virtual learning environment: An empirical investigation*. *Information & Management*, 51(4), 430-440.

Yan-hong, L., Bo, Z., & Jian-hou, G. (2015, July). Make adaptive learning of the MOOC: The CML model. *In Computer Science & Education (ICCSE), 2015 10th International Conference on* (pp. 1001-1004). IEEE

Yin, R. K. (2003). *Designing case studies*.

Yuan, L. & Powell, S. (2013). MOOCs and Open Education: Implications for Higher Education. *Centre for Education Technology & Interoperability Standards*.

Zapata, M. (2012). *Enseñanza Universitaria en línea, MOOC y aprendizaje divergente*. Disponible en: <http://eprints.rclis.org/22634/>

Zapata, M (2013a), *MOOCs, una visión crítica y una alternativa complementaria: La individualización del aprendizaje y de la ayuda pedagógica*. E-LiS e-prints in library & information science. Disponible en: <http://eprints.rclis.org/18658/>