



UNIVERSIDAD DE CUENCA

Facultad de Ingeniería

Carrera de Sistemas

“Detección de secuencias de aprendizaje en cursos abiertos masivos y en línea”

Trabajo de titulación previo
a la obtención del título de
Ingeniero de Sistemas

Autor:

Franklin Bolívar Bernal Montenegro

CI: 0302195177

Director:

Ing. Jorge Javier Maldonado Mahauad, PhD.

CI: 1102959051

Cuenca, Ecuador

24-octubre-2019



Resumen

El uso de la tecnología en los contextos educativos actuales, por ejemplo, presencial, blended y online han cambiado la forma de enseñar y aprender. En estos contextos los cursos abiertos masivos y en línea (MOOCs), son una de las tendencias más disruptivas de los últimos 8 años, teniendo entre 2012 y 2018 aproximadamente 101 millones de personas inscritas en alguno de los más de 11.400 MOOCs ofertados por alguna de las plataformas existentes como por ejemplo Coursera. Sin embargo, la tasa de aprobación de los estudiantes en este tipo de cursos apenas bordea el 5%. Esto ha originado un gran interés en los investigadores por estudiar el comportamiento de los estudiantes en este tipo de cursos; y además el poder develar por qué un estudiante termina o no un MOOC. El objetivo de este trabajo de tesis es explorar el comportamiento de los estudiantes en un curso MOOC. Específicamente se busca estudiar las secuencias de aprendizaje y extraer los patrones de comportamiento en las sesiones de estudio que estos realizan en un MOOC, y la relación con su desempeño académico. Para lograr el objetivo propuesto en este trabajo, haciendo uso técnicas de aprendizaje máquina y minería de procesos, se obtuvieron logs de eventos que permiten modelar el comportamiento de los estudiantes en un curso MOOC en la plataforma Coursera (N=1.550). Este log de eventos permite extraer las características de las sesiones de estudio de los estudiantes y evidenciar las interacciones de los estudiantes con los recursos del curso por cada semana del curso. Como resultado, se obtuvieron dos grupos de estudiantes basados en sus sesiones de estudio, donde se encontraron diferencias estadísticamente significativas entre las sesiones de estudio de ambos grupos. Además, se encontraron diferencias entre las interacciones de los estudiantes con los recursos del MOOC y la forma de abordar las lecciones por semanas entre ambos grupos. Los resultados encontrados contribuyen a la caracterización de los estudiantes de un MOOC en base a sus sesiones de estudio. Además, se identificó un comportamiento intenso por parte de los estudiantes que aprobaron el curso, en el que repitieron varias veces las evaluaciones hasta conseguir aprobarlas, sin regresar a una video-lectura actividad anterior, es decir; los estudiantes fueron perseverantes y repitieron las evaluaciones hasta conseguir aprobarlas.

Palabras Claves: MOOC. Coursera. Minería de procesos. Analíticas de aprendizaje.



Abstract

The use of technology in current educational contexts, for example, face-to-face, mixed and online have changed the way of teaching and learning. In these contexts, the massive and online open courses (MOOCs) are one of the most disruptive trends of the last 8 years, with approximately 101 million people enrolled in some of the more than 11,400 MOOCs offered by some of 2012-2018. The right platforms such as Coursera. However, the approval rate of students in these types of courses barely borders 5%. This has caused a great interest in researchers to study the behavior of students in this type of courses; and also the power to reveal why a student ends or not a MOOC. The objective of this thesis work is to explore the behavior of students in a MOOC course. Specifically, it seeks to study the learning sequences and extract the behavior patterns in the study sessions they perform in a MOOC, and the relationship with their academic performance. To achieve the objective proposed in this work, make use of machine learning and process mining techniques, obtain records of events that allow modeling the behavior of students in a MOOC course on the Coursera platform (N = 1,550). This event log allows you to extract the characteristics of the students' study sessions and to demonstrate the interactions of the students with the resources of the course for each week of the course. As a result, we obtained two groups of students identified in their study sessions, where we found statistically specific differences between the study sessions of both groups. In addition, differences were found between student interactions with MOOC resources and how to address the lessons per week between both groups. The results found according to the characterization of the students of a MOOC based on their study sessions. In addition, intense behavior was identified by the students who passed the course, in which he repeated the evaluations several times until he was able to pass them, without returning to a previous activity video-reading, that is; the students were persevering and repeated the evaluations until they were approved.

Keywords: MOOC. Coursera. Process mining. Learning analytics.



Índice del Trabajo

Resumen	2
Abstract	3
Índice de Figuras	6
Índice de Tablas	8
Dedicatoria	12
Agradecimientos.....	13
Capítulo 1: Introducción	14
1.1. Panorama general y justificación	14
1.2. Estado del Arte.....	16
1.2.1. Estudio de las Estrategias de Aprendizaje en MOOCs.....	16
1.2.2. Estudio de las Sesiones de Aprendizaje en MOOCs.....	18
1.3. Objetivos del Trabajo	19
1.3.1. Objetivo general.....	19
1.3.2. Objetivos específicos.....	19
1.4. Metodología	19
1.5. Estructura de la Tesis.....	21
Capítulo 2: Cursos abiertos masivos y en línea (MOOCs)	22
2.1. Introducción.....	22
2.2. El inicio y la evolución de los MOOCs.....	22
2.3. Definición de los cursos abiertos masivos y en línea	23
2.4. Características de los cursos abiertos masivos y en línea	24
2.5. Clasificación de los cursos abiertos masivos y en línea.....	25
2.5.1. xMOOC.....	25
2.5.2. cMOOC.....	28
2.6. Recapitulación.....	31
Capítulo 3: Plataforma Coursera	32
3.1. Introducción.....	32
3.2. Principales características de los cursos MOOC en Coursera.....	32
3.3. Arquitectura de los datos en Coursera	33
3.4. Principales tablas de la base de datos	37
3.5. Recapitulación.....	39
Capítulo 4: Minería de Procesos	40



4.1.	Introducción.....	40
4.2.	Minería de Procesos	40
4.3.	Tipos de Minería de Proceso	41
4.3.1.	Descubrimiento.....	42
4.3.2.	Verificación de Conformidad.....	42
4.3.3.	Mejoramiento	42
4.4.	Algoritmos de Descubrimiento en MP	43
4.4.1.	Heuristic Miner.....	43
4.4.2.	Fuzzy Miner	44
4.4.3.	Multi Phase Miner	45
4.5.	Principales herramientas para Minería de Procesos	46
4.5.1.	ProM	46
4.5.2.	DISCO	47
4.6.	Recapitulación.....	51
5.	Caso de Estudio.....	52
5.1.	Introducción.....	52
5.2.	Tipo y enfoque del caso de estudio.....	52
5.3.	Contexto: MOOC y participantes.....	52
5.4.	Metodología PM2 para Minería de Procesos	54
5.4.1.	Objetivos.....	55
5.4.2.	Extracción de datos	55
5.4.3.	Generación del log de eventos	61
5.4.4.	Descubrimiento y análisis de los modelos	65
5.5.	Recapitulación.....	82
6.	Conclusiones y trabajos futuros	85
6.1.	Conclusiones.....	85
6.2.	Trabajo Futuro.....	89
	Bibliografía	91
	Anexos	97
	Anexo 1: Tabla de ejemplos de proveedores xMOOC.....	97
	Anexo 2: Script de Python – Procesamiento de sesiones	98
	Anexo 3: Script de Python – Agregar estado en las actividades.....	103
	Anexo 4: Script de Python – Agregar el número de interacciones con las actividades	108

Índice de Figuras

Figura 1: Etapas para realizar el estudio	20
Figura 2: Características de los MOOCs (Yousef, Chatti, Schroeder, Wosnitza, & Jakobs, 2014)	24
Figura 3: xMOOC según Littlejohn (2013)	26
Figura 4: Características del xMOOC (Yousef et al., 2014)	27
Figura 5: cMOOC según Littlejohn (2013)	29
Figura 6: Características del cMOOC (Yousef et al., 2014)	30
Figura 7: Interfaz principal y organización de los cursos en Coursera (Pérez-Álvarez, Maldonado-Mahauad, & Pérez-Sanagustín, 2018)	33
Figura 8: Arquitectura de Datos Coursera (Coursera, 2016)	34
Figura 9: Modelo de datos simplificado para el desarrollo del caso de estudio	37
Figura 10: Posicionamiento de PM (Rudnitckaia, 2017)	41
Figura 11: Secuencia para la aplicación del MP	41
Figura 12: Tipos de Minería de Procesos en términos de entrada y salida (Bogarín, Cerezo, & Romero, 2018)	42
Figura 13: Resultado de un análisis con un Algoritmo de Minería Heurística ..	44
Figura 14: Resultado de un análisis con un Algoritmo de Minería Difusa	45
Figura 15: Resultado de un análisis con un Algoritmo de Minería Multi Fase ..	45
Figura 16: ProM, interfaz gráfica	47
Figura 17: Descubrimiento Automatizado de Procesos en DISCO	48
Figura 18: Animación del mapa de procesos en DISCO	48
Figura 19: Estadísticas detalladas en DISCO	49
Figura 20: Casos en Disco	49
Figura 21: Filtros en Disco	50
Figura 22: Importar y Exportar datos en Disco	50
Figura 23: Distribución porcentual de actividades en el MOOC del caso de estudio	53
Figura 24: Etapas para generar el modelo de procesos, adaptado de PM2 (Van Eck et al., 2015)	55
Figura 25: Fragmento de un archivo CSV de la tabla course_grades	56
Figura 26: Ejemplo de un fichero HTML de la tabla course_grades	56



Figura 27: Fragmento de tablas importadas del modelo de datos Coursera y auxiliares del caso de estudio 57

Figura 28: Pentaho Data Integration - Transformación "if field value is null" 58

Figura 29: Pentaho Data Integration - Transformación "Mapeo de Valores" 59

Figura 30: Fragmento de código de los scripts mostrados en los Anexos 1 y 261

Figura 31: Distribución de las variables para caracterizar los grupos del clúster 65

Figura 32: Clúster 0 (morado) y clúster 1 (amarillo) 66

Figura 33: Elbow Curve, para determinar el número ideal de grupos 66

Figura 34: Silhouette Score para determinar el número ideal de grupos 67

Figura 35: Comportamiento promedio por semana Clúster 0 67

Figura 36: Comportamiento promedio por semana de los estudiantes que aprobaron el curso del Clúster 0 68

Figura 37: Comportamiento promedio por semana Clúster 1 68

Figura 38: Comportamiento promedio por semana de los estudiantes que aprobaron el curso del Clúster 1 69

Figura 39: Ejemplo de interpretación de un modelo de procesos 70

Figura 40: Clúster_0, modelo de procesos a nivel micro 72

Figura 41: Clúster_1, modelo de procesos a nivel micro 73

Figura 42: Modelo de procesos en la primera semana del MOOC 74

Figura 43: Modelo de procesos, segunda semana del MOOC 74

Figura 44: Modelo de procesos, tercera semana del MOOC 75

Figura 45: Modelo de procesos, cuarta semana del MOOC 76

Figura 46: Modelo de procesos, quinta semana del MOOC 77

Figura 47: Modelo de procesos, sexta semana del MOOC 78

Figura 48: Modelo de procesos, séptima semana del MOOC 78

Figura 49: Modelo de proceso, nivel micro de los estudiantes que aprueban el curso 80

Figura 50: Modelo de proceso, nivel micro de los estudiantes que no aprueban el curso 80

Figura 51: Comportamiento de los que aprueban vs no aprueban 81

Figura 52: Promedio de sesiones realizadas por semana de los estudiantes del MOOC 82



Índice de Tablas

Tabla 1: Tipos de actividades de un MOOC en Coursera (Pérez-Álvarez et al., 2018)	36
Tabla 2: Distribución de Género	53
Tabla 3: Distribución de Edades	54
Tabla 4: Distribución de Nivel Académico	54
Tabla 5: Definición de las interacciones con los tipos de actividades del curso	61
Tabla 6: Fragmento de los datos usados para generar el clúster	63
Tabla 7: Fragmento del log de eventos utilizado para el análisis a nivel micro	64
Tabla 8: Fragmento del log de eventos utilizado para el análisis a nivel macro	64
Tabla 9: Prueba t de Student aplicada a los datos del Clúster 0 y el Clúster 1	69
Tabla 10: Ejemplos de Proveedores xMOOC (Littlejohn, 2013)	97



Cláusula de licencia y autorización para publicación en el Repositorio Institucional

Franklin Bolívar Bernal Montenegro en calidad de autor y titular de los derechos morales y patrimoniales del trabajo de titulación "Detección de secuencias de aprendizaje en cursos abiertos masivos y en línea", de conformidad con el Art. 114 del CÓDIGO ORGÁNICO DE LA ECONOMÍA SOCIAL DE LOS CONOCIMIENTOS, CREATIVIDAD E INNOVACIÓN reconozco a favor de la Universidad de Cuenca una licencia gratuita, intransferible y no exclusiva para el uso no comercial de la obra, con fines estrictamente académicos.

Asimismo, autorizo a la Universidad de Cuenca para que realice la publicación de este trabajo de titulación en el repositorio institucional, de conformidad a lo dispuesto en el Art. 144 de la Ley Orgánica de Educación Superior.

Cuenca, 24 de octubre de 2019

Franklin Bolívar Bernal Montenegro

C.I: 0302195177



Cláusula de Propiedad Intelectual

Franklin Bolívar Bernal Montenegro, autor del trabajo de titulación ““Detección de secuencias de aprendizaje en cursos abiertos masivos y en línea”, certifico que todas las ideas, opiniones y contenidos expuestos en la presente investigación son de exclusiva responsabilidad de su autor/a.

Cuenca, 24 de octubre de 2019

A handwritten signature in blue ink, consisting of stylized letters and a large circular flourish, positioned above a horizontal line.

Franklin Bolívar Bernal Montenegro

C.I: 0302195177

Certifico

Que el presente trabajo de titulación: "Análisis del comportamiento de los estudiantes en un curso MOOC basado en sus sesiones de estudio", fue dirigido y revisado por mi persona.

Atentamente,



Ing. Jorge Maldonado Mahauad, PhD.
C.I: 1102959051
Director



Dedicatoria

Dedico este trabajo a toda mi familia, especialmente a mi esposa María Inés y mis amados hijos; Esteban Daniel, Franklin Matías y Juan Manuel, que han sido pilar fundamental y fuerza en cada uno de los escalones que he avanzado. Además, Franklin y Luci, mis padres que con su ejemplo de lucha, responsabilidad y perseverancia han guiado mis pasos toda la vida, esta tesis es para todos ustedes.

Franklin Bernal M.



Agradecimientos

El proceso que se ha seguido para realizar este trabajo de titulación, ha sido un esfuerzo compartido, un tiempo de mucho enriquecimiento personal, una etapa de la vida difícil de olvidar por el aprendizaje, frustraciones y deseos de seguir adelante. En este trabajo estuvieron involucradas muchas personas, cada una de ellas aportó de manera diferente pero muy significativa a la culminación de este trabajo. A todos ellos, mi deseo de expresar mis más sinceros agradecimientos.

Primero a mi familia, por la paciencia y apoyo incondicional para afrontar toda la carrera y las decisiones más importantes que he tomado en mi vida.

A mi director del trabajo de titulación, ingeniero Jorge Maldonado, por guiarme, darme esa motivación extra siempre necesaria, por las críticas constructivas, soluciones a los problemas que se presentaron y sobre todo por la paciencia que me tuvo a lo largo de este tiempo.

Finalmente, me gustaría agradecer a cada uno de los profesores, quienes supieron transmitir sus amplios conocimientos y experiencias a través de las aulas universitarias, que de seguro serán herramientas muy útiles en el ámbito profesional.

El Autor



Capítulo 1: Introducción

1.1. Panorama general y justificación

El uso de la tecnología en los contextos educativos actuales, como son el presencial, híbrido y online han cambiado la forma de enseñar y aprender. Por un lado, los docentes hoy en día deben ser capaces de mediar los contenidos que se imparten utilizando los recursos digitales que estén a su alcance. Por otro lado, los estudiantes cada vez tienen más acceso a estos contenidos de forma atemporal y sin importar su ubicación geográfica, lo que propicia que el aula de clase se extienda.

Una de las tendencias más disruptivas en los últimos 8 años ha sido los Cursos Masivos Abiertos y en Línea (MOOCs – Massive Open Online Courses en inglés). Estos cursos actualmente atraen a millones de estudiantes alrededor de todo el mundo (masivos) y se ofertan de forma online sobre alguna de las plataformas para cursos MOOC como son: Edx, Miríada X, Coursera. Además, sus contenidos están abiertos y están organizados bajo el formato de un curso con una intención pedagógica que los distingue de otro tipo de recursos educativos.

Entre 2012 y 2018 se registraron aproximadamente 101 millones de personas inscritas en alguno de los más de 11.400 MOOCs ofertados por alguna de las plataformas existentes, siendo Coursera el principal proveedor de este tipo de cursos con 37 millones de estudiantes registrados (Shah, 2018a). Sin embargo, y a pesar de la disrupción causada por los MOOCs, solo un porcentaje reducido de los estudiantes que se inscriben en un MOOC logran terminarlo (Wang & Baker, 2015).

La tasa de aprobación de los estudiantes en este tipo de cursos bordea apenas el 5%, esto según un estudio realizado por Kizilcec & Cohen (2017), por lo que el esfuerzo de los investigadores se ha centrado en tratar de determinar cuáles son las causas por las que los estudiantes abandonan un MOOC.

Según distintos autores como Bernal-González, Maldonado-Mahauad y demás (Bernal-González, 2015; Maldonado-Mahauad, Pérez-Sanagustín, Kizilcec, Morales, & Muñoz-Gama, 2018), algunas de las posibles causas o factores que influyen en el fracaso que influyen en abandono de los estudiantes en un MOOC son:

- (1) poco tiempo libre para abordar estos cursos,
- (2) falta de conocimientos previos para tomar los cursos,
- (3) contenido del curso deficiente,
- (4) deficiente diseño instruccional del curso,
- (5) ausencia de trabajo colaborativo,



- (6) falta de autorregulación de su aprendizaje y;
- (7) interacción insuficiente con el tutor o compañeros del curso.

En relación al diseño del curso, existen pocos trabajos que han explorado como los estudiantes interactúan con los recursos en un curso MOOC. Por ejemplo, Finchman et al. (2018) identificaron patrones en el comportamiento que tienen los estudiantes en el aprendizaje en base a las sesiones de estudio, utilizando las técnicas de aprendizaje automático no supervisado, que consistieron en Hidden Markov Models (Modelos ocultos de Markov o HMMs) y técnicas de agrupamiento (clustering), lo que ha permitido identificar la relación existente entre estrategias de aprendizaje y los resultados académicos de los estudiantes (Fincham, Gasevic, Jovanovic, & Pardo, 2018). En otro estudio Geigle y Zhai (2017), también utilizaron modelos ocultos de Markov (HMM) de dos capas, en el cual se encontraron diferencias significativas entre los estudiantes que aprobaron el curso con los que no lo hicieron. Específicamente, demostraron que los estudiantes de mejor rendimiento producen diagramas de transición HMM ampliamente diferentes que tienden a mostrar un período de concentración más largo en la realización de pruebas y una participación más activa en el foro en comparación con los estudiantes promedio (Geigle & Zhai, 2017).

Sin embargo a pesar de estos esfuerzos, se requiere de más evidencia empírica sobre (1) las interacciones de los estudiantes con los contenidos del curso, (2) el tipo de estrategias de aprendizaje que estos utilizan y (3) de cómo estas interacciones permiten determinar las estrategias de aprendizaje que los estudiantes utilizan, además de cómo por medio del diseño del curso se fomenta determinadas estrategias de aprendizaje (Davis, Seaton, Hauff, & Houben, 2018).

Para el estudio del comportamiento de los estudiantes en entornos en línea, y específicamente en cursos MOOC, técnicas computacionales como las del machine learning, la minería de datos y la minería de procesos (MP) han sido utilizadas de manera combinada para dar cuenta de las estrategias que utilizan los estudiantes en este tipo de cursos. Un ejemplo de esta combinación son el trabajo de Maldonado-Mahauad et al., (2018) y de Saint y Gašević, (2018) quienes utilizando técnicas de minería de procesos han logrado develar algunas de las estrategias de aprendizaje a partir de los grupos obtenidos a través de interacciones similares de los estudiantes en entornos online denominados como: (1) Sampling Learners: estudiantes que siguen la estructura secuencial del MOOC, (2) Comprehensive Learners: estudiantes que buscan la información requerida para aprobar las evaluaciones y (3) Targeting Learners: estudiantes de muestreo que se comportan de manera irregular y no estructurada, y en entornos blended denominados como: (1) Active Agile, (2) Efficient, (3) Summative Gamblers, (4) Active Cohesive y (5) Extreme Minimalists (Maldonado-Mahauad et al., 2018; Saint, J., Gašević, D., 2018).

La minería de procesos es una disciplina de investigación relativamente joven que se ubica entre la inteligencia computacional y la minería de datos. El punto de partida de la minería de procesos es un registro de eventos y se pueden realizar 3 tipos de PM: Descubrimiento, verificación de conformidad y el mejoramiento (W Van Der Aalst et al., 2012).

En este trabajo de tesis, se propone trabajar utilizando técnicas de MP, para analizar el comportamiento de los estudiantes en un curso MOOC. Específicamente, se buscará explorar las secuencias de aprendizaje de los estudiantes en el curso por medio de las interacciones de estos con los recursos del curso, con el propósito de descubrir patrones de aprendizaje y caracterizar sus sesiones de estudio.

1.2. Estado del Arte.

1.2.1. Estudio de las Estrategias de Aprendizaje en MOOCs

Para estudiar las estrategias de aprendizaje en los MOOC, los investigadores han seguido dos enfoques diferentes: enfoques basados en aptitudes y en procesos (Trigwell & Prosser, 1991). Por un lado, un enfoque basado en la aptitud busca medir estas estrategias por medio de instrumentos de auto-reporte cuyo objetivo es dar cuenta de las estrategias que los estudiantes creen emplear durante su aprendizaje. Por otro lado, un enfoque basado en procesos busca estudiar las estrategias de aprendizaje de los estudiantes a partir de un conjunto de eventos (es decir, acciones) que resultan de la interacción de los estudiantes con los contenidos del curso. Sin embargo, la observación de estrategias de aprendizaje, incluso cuando se manifiestan como un conjunto de eventos o acciones, implica varios desafíos, tales como:

- 1) transformar los rastros de datos de granularidad fina (p.e. clickstream) o gruesa (p.e conjunto de eventos clickstream) en un comportamiento interpretable (estrategias de aprendizaje);
- 2) identificar y observar cambios de comportamiento durante el proceso de aprendizaje; y,
- 3) entender si un comportamiento observable se relaciona con una estrategia de aprendizaje particular o con más de una (Pashler & Wagenmakers, 2012).

Los avances recientes en las disciplinas de las Analíticas de Aprendizaje (LA – Learning Analytics) y MP han contribuido para lograr superar estos desafíos (Cole, 1971). Por ejemplo, autores como Jovanović et al. (2017); utilizaron técnicas de aprendizaje automático no supervisadas con algoritmos de minería de secuencias para explorar las estrategias de aprendizaje de los alumnos a partir de la extracción de los datos de la plataforma de e-learning (Jovanović, Gašević, Dawson, Pardo, & Mirriahi, 2017). En otro trabajo realizado por Finchman y demás (2018), utilizando modelos de Markov se extrajeron tácticas de



estudio y luego las agrupó para para derivar estrategias de aprendizaje (Fincham et al., 2018).

Mukala y demás (Mukala, Buijs, Leemans, & van der Aalst, 2015) aplicaron técnicas de MP en un MOOC en Coursera con 43.218 alumnos para comprender sus procesos de aprendizaje y analizar cómo se desempeñaron después de las videoconferencias y evaluaciones. Esto se realizó por medio de grupos que se obtuvieron a partir de los diferentes atributos en los datos, dentro del primer grupo se agrupo por las siguientes variables: (1) estudiantes que no aprobaron el curso, (2) Estudiantes que aprobaron el curso, (3) estudiantes con certificado normal, y (4) estudiantes con certificado de distinción. En segundo lugar, se realizó un agrupamiento por: (1) Si el estudiante realizó un rastreo de firma o validación de identidad y (2) si el estudiante no realizó un rastreo de firma o validación de identidad. Finalmente, una tercera agrupación por la participación semanal de actividades o eventos dentro del curso: (1) eventos registrados en la semana uno y dos, (2) eventos registrados en las semanas uno, dos y tres, (3) eventos en todas las semanas de duración del curso. Mediante verificación de conformidad con el modelo que se debería seguir para aprobar el curso, se pudo observar que los estudiantes con calificaciones altas que sirven para aprobar el curso, seguían dicho modelo mientras los estudiantes de bajo nivel académico no lo hacían o se retiraban en etapas más tempranas.

Maldonado-Mahauad y demás (Maldonado-Mahauad et al., 2018) utilizaron el algoritmo de minería difusa, a partir del cual extrajeron siete tipos de estrategias de aprendizaje de los alumnos de cuatro MOOCs en Coursera, las cuales consistieron en: (1) Solo video-lecturas: sesiones que realizan solo ese tipo de actividades, (2) Solo evaluaciones: sesiones en la que los estudiantes se dedican a realizar este tipo de actividades, (3) Intentar-Evaluación a Video-Lectura: en este tipo de sesiones los estudiantes intentan aprobar una evaluación e interactúan con las video-lecturas para buscar información que los ayude a aprobar las evaluaciones (4) Explorar: un tipo de sesión en la que los estudiantes no completan las actividades ni de video-lecturas ni evaluaciones, (5) Completar video-lectura a Intentar-Evaluación: un tipo de sesión, en la que los estudiantes completan una video-lectura y proceden a rendir una evaluación sin conseguir aprobarla, (6) Video-lectura a Aprobar-Evaluación: es un tipo de sesión donde el estudiante completa una actividad de video-lectura y pasa a la evaluación y la aprueba de manera correcta (7) Otros: son secuencias de interacción largas y dispersas que no encajan en ningún patrón antes mencionado. Finalmente, Juhaňák y demás (Juhaňák, Zounek, & Rohlíková, 2017) utilizaron MP para explorar el comportamiento de los participantes a partir de un cuestionario y los patrones de interacción en un sistema de gestión de aprendizaje (Moodle), en donde se obtuvieron cuatro tipos de comportamiento, los cuales fueron: (1) comportamiento estándar: que sigue el flujo secuencial de las actividades, (2) mal uso de la retroalimentación: en el que los estudiantes confiaron en la retroalimentación con respecto a sus intentos anteriores al responder las tareas del cuestionario, (3) uso indebido de los materiales de estudio: en el cual los estudiantes ven los materiales de estudio en el curso mientras toman una



prueba, (4) comportamiento multitarea: en el que los estudiantes participan en otras actividades de aprendizaje en el curso mientras toman una prueba.

En una investigación más reciente de Van den Beemt, Buijs y Van der Aalst (2018), utilizaron un enfoque de MP y agrupación de los estudiantes por las secuencias semanales que realizaban los estudiantes, en un curso con 16.224 estudiantes, indica que los estudiantes que vieron regularmente en grupos sucesivos los videos conlleva a mejores resultados de aprendizaje, a pesar de no ser una actividad calificada, donde los estudiantes que no eran “Usuarios Verificados” son más propensos a saltarse estos videos, respecto a los que sí (Van den Beemt, Buijs, & Van der Aalst, 2018). Este último trabajo establece la base para comenzar a considerar PM como una técnica adecuada para analizar datos de rastreo.

1.2.2. Estudio de las Sesiones de Aprendizaje en MOOCs

Según autores como Romero y Usart (2014) y Sapunar y demás (2018), una sesión de estudio se puede definir como un periodo de tiempo donde el estudiante interactúa con los recursos del curso, los cuales son registrados por la plataforma de aprendizaje (Coursera, EdX, Udacity, etc.) y donde los intervalos de inactividad no son mayores a 40 minutos y permiten identificar y detectar las sesiones de estudio (Romero & Usart, 2014; Sapunar-Opazo, Pérez-Álvarez, Maldonado-Mahauad, Alario-Hoyos, & Pérez-Sanagustín, 2018). Las plataformas para MOOC capturan las interacciones de los estudiantes lo que permite poder extraer el tiempo de trabajo del estudiante dentro de la plataforma (Kovanović et al., 2015). Un método común es calcular una sesión como el tiempo de inactividad de un estudiante lo cual sirve para detectar sesiones.

Para Kovanović y demás (2015) han visto que el tiempo de inactividad de los estudiantes para detectar sesiones puede considerarse desde los 30 a los 45 minutos (Kovanović et al., 2015).

Por ejemplo, autores como Matcha y demás (2019), proponen diferentes técnicas computacionales, para detectar estrategias de aprendizaje adoptadas por los estudiantes en entornos digitales, donde las actividades que realiza el estudiante por sesión, es directamente relacionada con la estrategia de aprendizaje del estudiante

Por lo anterior, en este trabajo se buscará responder a las siguientes preguntas de investigación:

1. ¿Qué tipo de sesiones realizan los estudiantes en un MOOC?
2. ¿Qué secuencias de aprendizaje realizan los estudiantes en las diferentes sesiones?
3. ¿Qué diferencia existe entre las sesiones de los estudiantes que completan el curso y de los que no?



1.3. Objetivos del Trabajo

1.3.1. Objetivo general

Explorar (a) las secuencias de aprendizaje y (b) extraer los patrones en las sesiones de estudio de los estudiantes en un MOOC y su relación con el desempeño académico.

1.3.2. Objetivos específicos

1. Estudiar las características de los MOOCs.
2. Desarrollar un modelo de datos que permita estudiar el comportamiento de los estudiantes a partir de los datos almacenados en una plataforma MOOC.
3. Aplicar técnicas de descubrimiento de Minería de Procesos que nos permitan estudiar los patrones de interacción de los estudiantes con los contenidos del curso.
4. Estudiar las sesiones de estudio de los participantes en un MOOC mediante un caso de estudio específico.

1.4. Metodología

Para lograr los objetivos específicos y poder responder a las preguntas de investigación planteadas, este trabajo empleó una metodología cuantitativa para el análisis de los datos.

Como procedimiento general para el desarrollo de este trabajo de tesis se siguió las siguientes etapas que se presentan en la **Figura 1**.

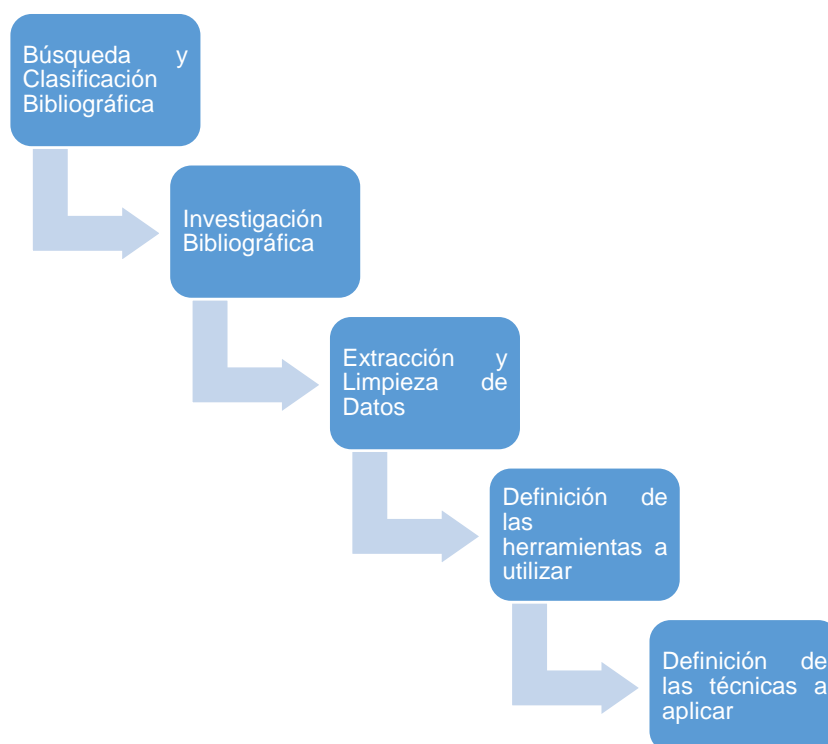


Figura 1: Etapas para realizar el estudio

- i. **Búsqueda y clasificación bibliográfica:** Se buscó artículos científicos, revistas, libros y toda aquella fuente bibliográfica de la cual se pueda obtener información verificada concerniente al tema de investigación propuesto, cabe resaltar que ya se tiene cierta información recolectada de fuentes bibliográficas verificadas. Además, se clasificó los artículos, con el fin de descartar todo aquel material que nos brinde información no relacionada con el tema propuesto.
- ii. **Revisión bibliográfica:** Con la bibliografía ya clasificada, se revisó y analizó dicha información para crear el estado del arte y saber que se ha estudiado sobre las estrategias de aprendizaje y sesiones de estudio en un MOOC.
- iii. **Extracción y limpieza de Datos:** Una vez finalizado el curso MOOC, y al haber almacenado los datos de sus estudiantes, se procederá a la extracción y limpieza de los datos, con el fin de obtener un LOG de eventos sobre el cual se aplicará las técnicas de minería de procesos.
- iv. **Definición de las herramientas a utilizar:** Al dar por finalizado el proceso de extracción y limpieza de datos se procedió a seleccionar las herramientas que nos permitieron hacer uso de los algoritmos de MP para realizar descubrimiento de procesos que dan cuenta de las estrategias que usan los estudiantes.
- v. **Definición de técnicas a aplicar:** A través de esta etapa se desea extraer conocimiento desde los registros de evento de los procesos almacenados por la plataforma, a partir de técnicas de descubrimiento de MP, que permitan develar las estrategias y las sesiones que emplean los estudiantes en un MOOC.



1.5. Estructura de la Tesis

Esta tesis se encuentra estructurada de la siguiente manera:

El capítulo dos presenta una introducción a los cursos masivos abiertos y en línea (MOOCs), para esto el capítulo dos presenta una definición de este tipo de cursos, una clasificación y sus principales características. Finalmente se presentan las conclusiones del capítulo.

El capítulo tres presenta una introducción a la plataforma Coursera, donde se describe su arquitectura para almacenar los datos. Además, se presenta una descripción general de las tablas más importantes de las cuales se recuperarán los datos que servirán para armar un log de eventos. Finalmente se presenta un modelo de datos que se utilizará para analizar el comportamiento de los estudiantes.

El capítulo cuatro presenta una introducción a la Minería de Procesos, donde se presenta un análisis de los principales algoritmos para descubrimiento y las herramientas de software que dan soporte al descubrimiento de modelos de procesos. Finalmente se presentan las conclusiones del capítulo.

El capítulo cinco se presenta el desarrollo de un caso de estudio, donde a partir del uso de la MP y utilizando los datos de un curso de la plataforma Coursera se presenta los modelos de procesos que sirven para identificar el comportamiento de los estudiantes en el curso, basado en variables de interacción por sesión de estudio.

Finalmente, el capítulo seis presenta las principales conclusiones de esta tesis y línea de trabajos futuros.



Capítulo 2: Cursos abiertos masivos y en línea (MOOCs)

2.1. Introducción

En este capítulo, el lector encontrará una revisión bibliográfica acerca de los MOOC. Se presentará una revisión de la historia e inicios y las distintas definiciones planteadas por los autores más relevantes de los últimos 10 años, los que han tomado distintas posiciones conceptuales desde lo tecnológico hasta lo didáctico y pedagógico.

La revisión de estas definiciones permitirá que el lector identifique las características más importantes que describen a un MOOC, y en base a estas, continuar en los siguientes capítulos con los conceptos previos hasta llegar al desarrollo del caso de estudio.

En este capítulo, también se realiza una revisión acerca de la clasificación principal de los MOOCs. Finalmente, el lector encontrará una recapitulación del contenido de este capítulo de la tesis.

2.2. El inicio y la evolución de los MOOCs

Antes de abordar las definiciones de los MOOC, es importante abordar los inicios y la evolución de lo que hoy se reconoce como un MOOC. Los MOOCs nacen con la idea inicial de la educación a distancia, y son producto de la evolución de la misma. En los inicios de la educación a distancia, el material o recursos (libros, cuadernos, textos guía, etc.), eran entregados a los estudiantes para que lo revisen en su hogar, sin restricción de horario con entregas periódicas y clases presenciales cada cierto tiempo obligatorias. Al continuar con esta evolución se introduce un primer avance de la educación a distancia; el audiovisual, donde los maestros grabaron audios y videos con el motivo de mejorar la experiencia de interacción, pero de la misma forma, la entrega del material o recursos era física (VHS, CD, Casete) (Liyana Gunawardena, Adams, & Williams, 2013).

Con el avance tecnológico, una nueva era digital se ve mejorada en el uso de nuevos dispositivos inteligentes. Su uso masivo abre la puerta al acceso de recursos e interacciones mejoradas que se dieron con la llegada de la Web 2.0, para que con ella se abrieran muchas puertas de las que nacerían propuestas innovadoras de interacción de los usuarios con los diferentes tipos de plataformas web y aplicaciones. Gracias a la Web 2.0 juntamente con el impacto de las redes sociales, empieza la evolución de “e-learning” para crear cursos abiertos masivos en línea MOOCs, los mismos que revolucionaron desde su aparición por el modo de impartir un curso y la forma de abordar una temática en específico.



Los MOOCs como término formal, apareció en 2008 y fue propuesto por Dave Cormier, desde entonces estos se han convertido en el catalizador del cambio del modelo tradicional de enseñanza/aprendizaje de las universidades.

El primer curso MOOC que abrió el camino de esta modalidad de aprendizaje fue denominado “Connectivism and Connective Knowledge - Conectivismo y aprendizaje conectivo”, con una duración de 12 semanas y una participación de aproximadamente 2.300 estudiantes de distintos países, quienes se registraron en forma libre. Este curso fue organizado por George Siemens y Stephen Downes en la Universidad de Manitoba (Canadá) en el 2008.

En el 2011, se lanzó el curso “Introduction to Artificial Intelligence - Introducción a la Inteligencia Artificial”, que fue organizado por Sebastián Thrun, profesor de la Universidad de Stanford y Peter Norvig, director de investigación de Google, en este curso se registraron 160.000 estudiantes de todo el mundo (Atiaja, 2016).

Desde entonces las universidades se han lanzado a una carrera respecto a la producción de MOOCs. A finales del año 2018 se registraron alrededor de 11.400 MOOCs en todo el mundo, respecto a los 2.500 registrados en el 2017 en los que participan millones de estudiantes, según el observatorio mundial para MOOCs, Class Central (Shah, 2018b).

En los 10 años de historia, los MOOCs se conceptualizan como la evolución del aprendizaje en línea. Además, su popularidad se explica a partir del aval de algunas de las universidades más prestigiosas de Estados Unidos y el mundo, como, por ejemplo; Universidad de Stanford en EEUU, Universidad de Edimburgo en Reino Unido, la Pontificia Universidad de Chile en Latinoamérica, la Universidad China de Hong Kong, entre otras las mismas que han acogido y desarrollado con entusiasmo el concepto a través de iniciativas como EdX, Coursera o Udacity (Torres, 2013).

2.3. Definición de los cursos abiertos masivos y en línea

El concepto de MOOC hace referencia a sus iniciales en inglés (Massive Open Online Course), también en el idioma español existe una referencia por su acrónimo COMA (Cursos Online Masivos Abiertos), CAEM (Curso Abierto En línea Masivo), o CALGE (Curso Abierto en Línea a Gran Escala), es decir que se los ha llamado de distintas formas para referirse al mismo concepto, donde mundialmente es más conocido como MOOC (Normandi Atiaja Atiaja & Segundo Guerrero Proenza, 2016).

Dentro de las distintas definiciones, basándose en su acrónimo, encontramos la planteada por autores; McAuley y demás (2010) quienes definen a los MOOC como un fenómeno que integra la conectividad de las redes sociales, el acceso de expertos reconocidos en un campo de estudio (tutores) y una colección de recursos abiertos e libre acceso en línea, donde varios estudiantes (cientos o miles), pueden participar de manera activa, donde ellos se ajustan a los objetivos, conocimientos previos, habilidades e intereses comunes de aprendizaje (Mcauley, Stewart, Siemens, & Cormier, 2010). Otra definición planteada por Littlejohn (2013), argumenta que un MOOC es un curso destinado

a la participación interactiva a gran escala (masiva o de muchos usuarios) y al acceso abierto a través de internet, considerando una forma relativamente nueva de aprendizaje (Littlejohn, 2013). Sin embargo, más allá de su definición, los MOOCs han sido confundidos con repositorios de recursos en línea, pero su diferencia radica en el formato de curso de los MOOCs, el cual fomenta la participación de los estudiantes en actividades de aprendizaje que permiten alcanzar los objetivos planteados por el material educativo digital.

2.4. Características de los cursos abiertos masivos y en línea

Las características de los Cursos Abiertos Masivos y en Línea, se basan en su acrónimo en inglés MOOC, donde cada letra significa una característica específica (**Figura 2**).

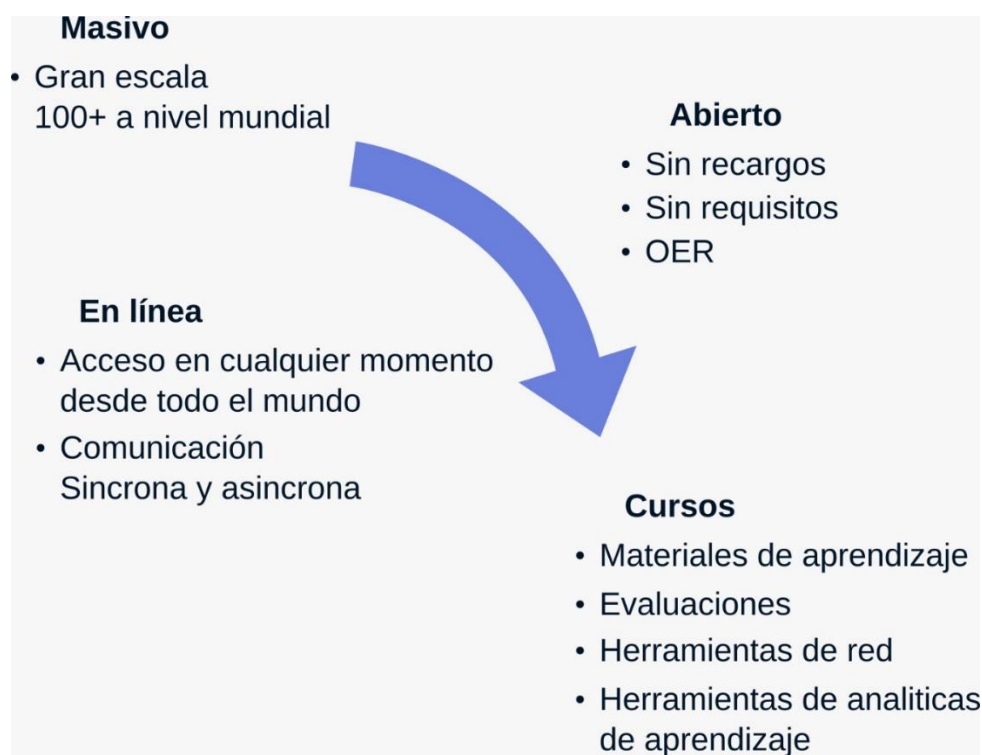


Figura 2: Características de los MOOCs (Yousef, Chatti, Schroeder, Wosnitza, & Jakobs, 2014)

Masivo: Según Terry Anderson (Anderson, 2013), el término “Masivo”, no es igual a determinar directamente el número de estudiantes inscritos o registrados en el curso, como se lo hace en un centro de educación convencional o a distancia, donde los registros que poseen nos da la cifra exacta de alumnos. El término en cuestión hace referencia a la gran escalabilidad que tiene para agregar participantes sin perder rendimiento o afectar su estudio por la cantidad de alumnos (Anderson & McGreal, 2012), en realidad esto será imperceptible para el usuario, que también difiere de los métodos tradicionales ya que un salón de clases tiene su capacidad limitada y establecida. Cabe resaltar que mientras



algunos MOOCs tienen unos cientos de alumnos existen otros con más de 150.000 usuarios registrados (Allen & Seaman, 2013).

Abiertos: Según Anderson y McGreal (2012), el término “Abierto”, se refiere a proporcionar una experiencia de aprendizaje a un gran número de participantes de todo el mundo y de calidad independientemente de: (1) Ubicación Geográfica, (2) Ideología y Política, (3) Contenido de libre acceso a los recursos, (4) Libre acceso a ser parte del curso, (5) Factores económicos como la gratuidad (Anderson & McGreal, 2012).

En línea: El término se refiere a la accesibilidad de estos recursos desde cada punto del mundo a través de Internet, en la cual se proporciona una interacción síncrona y asíncrona con los contenidos del curso y entre los participantes del MOOC (Stephen, 2013). En concordancia con la definición anterior Anderson (2013) indica en su definición que “En línea” es el aspecto menos polémico y discutido ya que los recursos (audio, video, podcasts, libros, presentaciones, actividades, etc.) y el modo de acceso a los mismos es en línea a través de un dispositivo con conexión a internet (PCs, tablets, Smartphones).

Curso: En términos de educación superior un curso es una unidad de enseñanza (Yousef et al., 2014), por lo que para realizar una mejor aproximación, según Youell (Youell, 2011) definió a la palabra “Curso” en un MOOC como un "compromiso académico coherente con un conjunto definido de resultados de aprendizaje".

2.5. Clasificación de los cursos abiertos masivos y en línea

El éxito de los MOOCs, ha dado como resultado una expansión de esta estrategia educativa a nivel mundial, por lo que es necesario clasificar los diferentes tipos de MOOC que se pueden encontrar (Atiaja & Guerrero, 2016). Además, los distintos modelos y enfoques de aprendizaje de los usuarios, obliga a los creadores de MOOCs a adaptarse empleando diferentes métodos de enseñanza (Kennedy, 2014). Cabe resaltar que, mediante la revisión bibliográfica, existen algunas clasificaciones de MOOCs, pero todas nacen a partir de dos tipos de MOOC principales que se detallan a continuación:

2.5.1. xMOOC

Los xMOOC fueron denominados cursos extendidos por las universidades de Harvard y Stanford, es por ello que el prefijo “x” corresponde a la palabra en inglés “extended”. Por otro lado, existen otras definiciones que se basan en el diseño del curso MOOC y autores como Wedekind (2013) define a los xMOOC como cursos masivos en línea con una ruta de aprendizaje fuertemente predeterminada, es decir; siguen un diseño “instructivista”, que hace referencia a que los objetivos de aprendizaje están predefinidos por un instructor y las vías

de aprendizaje estructuradas por el entorno, donde los estudiantes tienen interacciones limitadas con otros estudiantes, además en la **Figura 3** se puede ver como en este tipo de MOOC se conecta y consume estos recursos guiados (Littlejohn, 2013).

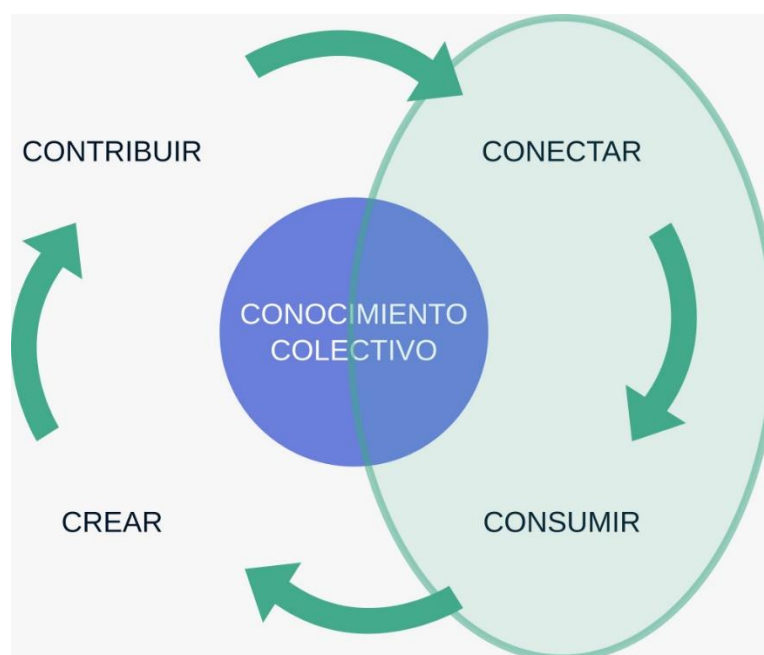


Figura 3: xMOOC según Littlejohn (2013)

Entre sus características principales, no deja de ser notorio su similitud con el modelo convencional conductista de un curso tradicional presencial, ya que la principal herramienta para la distribución de información de un xMOOC es la secuencia de videos, los cuales siguen un modelo de clases presenciales convencionales. Las otras actividades complementarias además de los videos, son lecturas, actividades y evaluaciones basadas en cuestionarios y respuestas de opción múltiple (Langer & Thillosen, 2013). Además, se plantean duraciones semanales de cada módulo, por lo general, entre 6 y 12 semanas (Sharples et al., 2015), donde las evaluaciones se realizan al terminar cada módulo y en su gran mayoría son auto calificadas o de opción múltiple, por otro lado al buscar más interacción entre estudiantes se plantean foros de discusión y evaluaciones por pares (Kerres & Preußler, 2013). Finalmente, las certificaciones que son entregadas a los estudiantes que culminan y aprueban el curso, lo que motiva a los mismos a obtener mejores calificaciones ya que por lo general este tipo de certificaciones tiene un costo (Nesterko, Seaton, Reich, & Chuang, 2013).

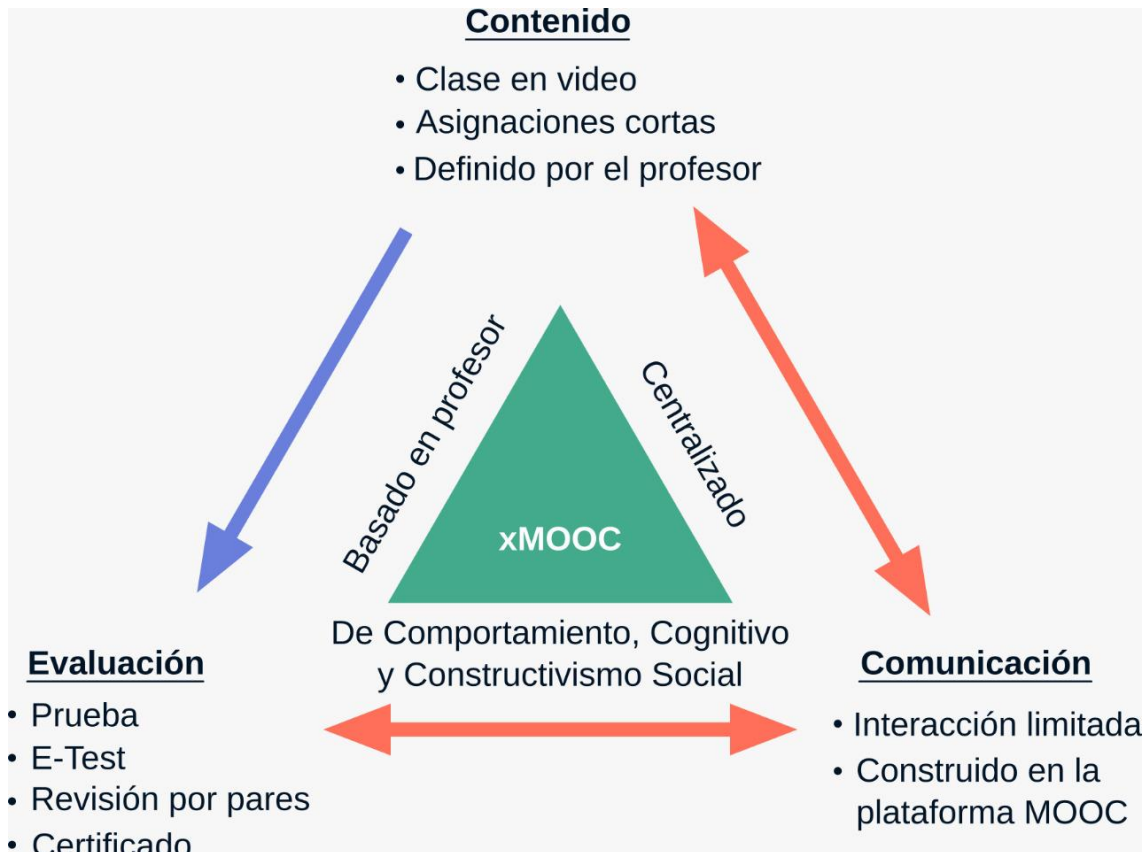


Figura 4: Características del xMOOC (Yousef et al., 2014)

El ejemplo más famoso de este tipo o clasificación de MOOC, es el curso de “Artificial Intelligence”, ofrecido por Sebastian Thrun profesor de Ciencias de la Computación de la Universidad de Stanford, donde se pudo notar el uso de las herramientas antes descritas (videos, actividades, lecturas, evaluaciones) y el modo que fueron presentadas a los estudiantes cuyo diseño era guiado o instructivista, además de la motivación añadida por la certificación de una Universidad importante al finalizar el curso con éxito (Littlejohn, 2013). Cabe resaltar que este tipo de MOOC es el más comercial y se distribuye en la mayoría de plataformas proveedoras de cursos MOOC como por ejemplo Coursera, EdX o Udemy (Yuan & Powell, 2013).

Es posible citar algunas ventajas y desventajas de los xMOOC:

Ventajas:

- Existe una alta disponibilidad en proveedores (Yuan & Powell, 2013), entre ellos los más importantes son Coursera, Edx, Udemy entre otras.
- Son los cursos que más se siguen al empezar con la educación en línea, ya que se aprovecha la similitud con el modelo convencional (Littlejohn, 2013).
- En estos cursos, se puede obtener una certificación al culminar con éxito un curso específico.



- Es accesible desde cualquier zona geográfica, donde solo es necesario un dispositivo que se pueda conectar a Internet como computadora, tableta o celular.
- Se puede acceder a este tipo de cursos sin poseer un título universitario, es decir; depende de la responsabilidad del estudiante en acatar los conocimientos previos que solicitan los cursos e inscribirse responsablemente, ya que la inscripción es libre.

Desventajas:

- La interacción entre miembros del curso es escasa y se limitan a foros y revisiones por pares.
- No son del todo abiertos, ya que muchos cursos de este tipo ofertan las evaluaciones bajo un modelo de pago, donde la certificación no es gratuita.

En cuanto a la estructura convencional que sigue un curso xMOOC, Pardos y Schneider (2013), recomiendan que cada curso debe caracterizarse en términos de su tema, audiencia y uso. Finalmente, en el **Anexo 1** se muestra una tabla (**Tabla 8**) en la que se indica algunos ejemplos de plataformas que distribuyen xMOOC.

2.5.2.cMOOC

Los cMOOC, se definen en base a su principio de diseño, el cual se lo conoce como “conectivismo”, donde Siemens (2005), define como una premisa básica del conectivismo, en que los estudiantes se conectan a través de redes digitales para aprender. En base a dicho principio de diseño Yuan y Powell (2013), indican que los cursos conectivistas se enfocan en el aprendizaje conectado y colaborativo. Finalmente una Downes (2012), define la esencia del conectivismo como la tesis de que el conocimiento se distribuye a través de una red de conexiones, y por lo tanto, el aprendizaje consiste en la capacidad de atravesar esas redes, por ende el conocimiento no se adquiere como si fuera una cosa, tampoco se transmite como un tipo de comunicación, más bien se busca tratar el aprendizaje como la formación de conexiones.

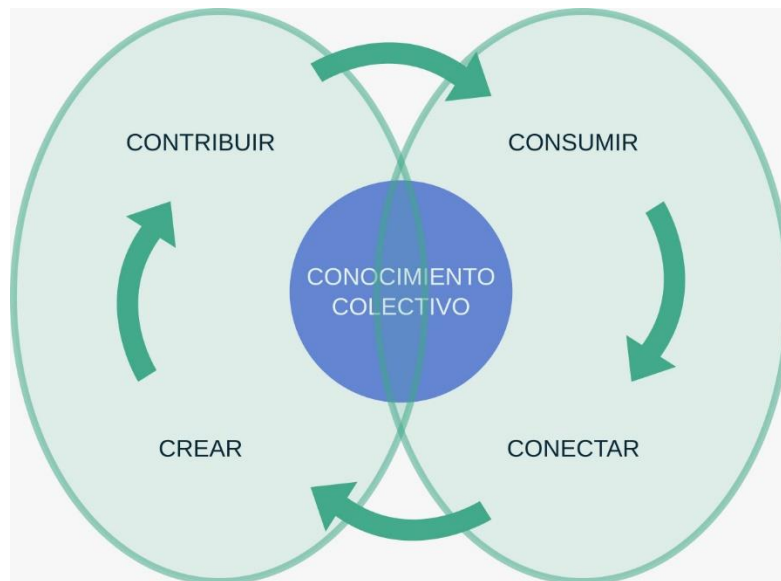


Figura 5: cMOOC según Littlejohn (2013)

En base a la definición antes mencionada, la característica principal de los cMOOC, es que los estudiantes pueden definir sus propios objetivos de aprendizaje, además de colaborar con la comunidad (curso) compartiendo conocimientos. Este modo de interactuar de forma colaborativa puede extenderse en diferentes redes, como blogs, wikis, grupos de redes sociales, etc., esto se lo realiza sin restricción de ningún maestro o tutor (Kruiderink, 2013). Para poder llevar a cabo estas definiciones y características, existen cuatro actividades clave en un cMOOC: (1) Agregación: en la que mediante una lista inicial de recursos en el sitio web, son agregadas y enviada a los participante mediante un boletín; (2) Remezcla: el proceso en donde se hacen todas las conexiones, y se documentan a través de blogs, marcadores sociales, tuiteos, etc.; (3) Reutilización: también conocido como constructivismo, hace referencia a que los participantes crean su propia conexión interna, y; (4) Retroalimentar: es decir, compartir nuevas conexiones con otros. Finalmente es importante indicar que los cMOOCs, se distribuyen en entornos de aprendizaje en red donde los estudiantes están en el centro del proceso de aprendizaje, por ello incluso las evaluaciones para una certificación en este tipo de MOOCs es secundaria, ya que prima el aprendizaje obtenido a una certificación (Bartolomé & Steffens, 2015).

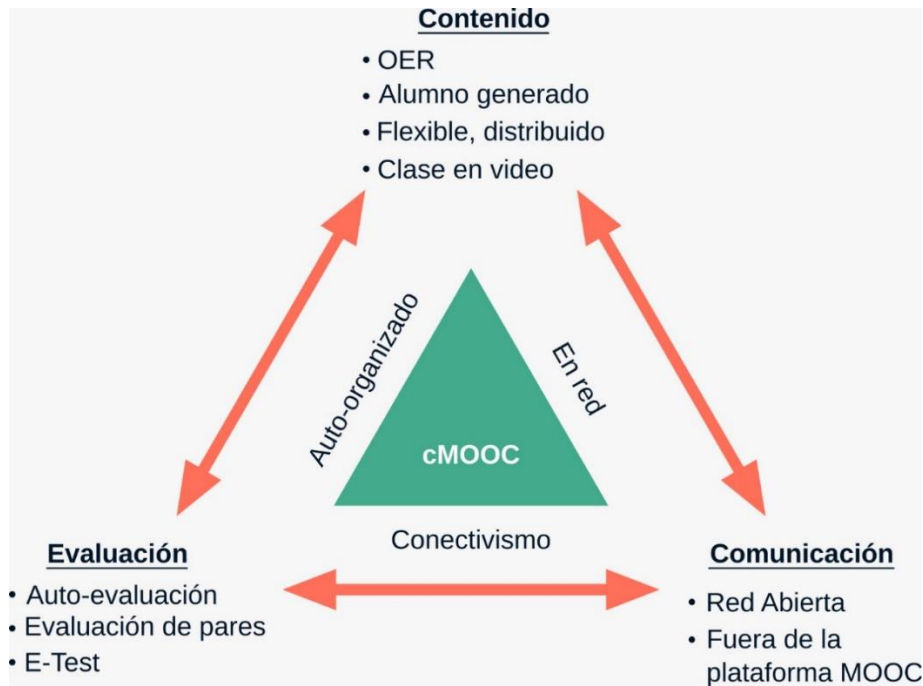


Figura 6: Características del cMOOC (Yousef et al., 2014)

Un ejemplo de este tipo de MOOCs, es coincidentemente el primer MOOC que se conoce según la literatura, el cual se denominó “Connectivism and Connective Knowledge”, que fue organizado y desplegado por George Siemens y Stephen Downes en la Universidad de Manitoba Canadá. Este curso tuvo una duración de 12 semanas y tuvo una masiva participación de aproximadamente 2.300 estudiantes de distintos países.

Ventajas:

- El conocimiento es lo principal, por lo que los estudiantes que ingresan a este tipo de cursos, dejan de lado el egoísmo y las evaluaciones.
- Existe mayor interacción entre los miembros del curso.
- La interacción de los miembros del curso, puede extenderse fuera de él por otra red social o grupos, donde comparten conocimiento sin restricciones (Kruiderink, 2013).
- Son completamente abiertos, como la premisa de su acrónimo lo indica.

Desventajas:

- Contienen contenido de menor calidad que los xMOOC (Yousef et al., 2014), ya que en este tipo de cursos se busca generar el contenido y compartir conocimientos.
- No se ofertan certificados avalados.
- Organización y diseño complejo para novatos, que desean empezar el aprendizaje en línea.



2.6. Recapitulación

En este capítulo se presenta una introducción a los cursos masivos abiertos y en línea, explicando su acrónimo en inglés conocido como MOOC (Massive Open Online Course), y también en español, donde existe una referencia por su acrónimo COMA (Cursos Online Masivos Abiertos) (Atiaja & Guerrero, 2016).

En cuanto a la definición, esta ha ido cambiando a lo largo del tiempo. Cabe resaltar que la mayoría hacen referencia a su acrónimo, por ello en este apartado como concepto general, McAuley (McAuley et al., 2010) define a los MOOC, como una integración de tutores expertos en el campo de estudio, a numerosos estudiantes que desean participar de manera activa en los recursos abiertos y de libre acceso que se ofertan en el curso.

Las características principales de los MOOC radican en su acrónimo: (1) Curso: compromiso académico coherente con un conjunto definido de resultados de aprendizaje (Youell, 2011), (2) Online: recursos disponibles en línea, y que permiten una interacción síncrona y asíncrona entre los participantes (Stephen, 2013), (3) Masivos: refiere a la gran escalabilidad para agregar alumnos sin afectar el rendimiento (Terry Anderson & McGreal, 2012), y (4) Abierto: se refiere a la libertad de acceder a un curso sin restricciones de tipo: geográficas, ideológicas, políticas y factores económicos (Terry Anderson & McGreal, 2012).

Los MOOCs se clasifican en: (1) xMOOC: que se basa en el diseño “instructivista”, donde los objetivos de aprendizaje son definidos por un instructor y las vías de aprendizaje estructuradas por el entorno, además las interacciones entre los estudiantes es limitada (Littlejohn, 2013). Por otro lado (2) cMOOC: sigue un diseño “conectivista”, donde los objetivos de aprendizaje tienden a ser definidos por el aprendiz, las vías de aprendizaje están abiertas y no definidas (Littlejohn, 2013), además existe mayor interacción entre los usuarios (estudiantes y tutores).



Capítulo 3: Plataforma Coursera

3.1. Introducción

En este capítulo, el lector encontrará una introducción a la plataforma proveedora de MOOCs “Coursera”. Se presentará una revisión sus características principales así también la arquitectura general de datos de la plataforma, lo que permitirá que el lector identifique las tablas de datos que se utilizarán más adelante en el desarrollo del caso de estudio.

Finalmente, en este capítulo concluye con el desarrollo de un modelo de datos diseñado para el desarrollo del caso de estudio, y una recapitulación con las ideas principales tratadas en este capítulo de la tesis.

3.2. Principales características de los cursos MOOC en Coursera

Coursera es la plataforma proveedora de MOOCs más grande del mundo y se puede acceder a ella mediante su sitio web oficial (<https://www.coursera.org/>), cuenta con más de 3.100 cursos activos, alrededor de 37 millones de usuarios, con más de 160 Universidades asociadas de más de 29 países (Coursera, 2019), Fue lanzada oficialmente en enero de 2012 como una iniciativa de dos profesores de la Universidad de Stanford, Andrew Ng y Daphne Koller. En 2018 Coursera se sitúa como la plataforma con más usuarios registrados (37 millones) y alcanzó una cifra record de ingresos superando los 140 millones de dólares por concepto de pago de certificados en línea, lanzando cerca de 675 cursos y 90 especializaciones (Shah, 2018a).

En cuanto a su diseño, Coursera se basa a cuatro principios:

- 1) El aprendizaje efectivo, según el Departamento de Educación de Estados Unidos, se obtienen mejores resultados en la educación en línea complementando el aprendizaje presencial;
- 2) Dominio del aprendizaje, un enfoque desarrollado por un centro de enseñanza por el psicólogo Benjamín Bloom, ayuda a los alumnos a entender un tema por completo antes de pasar al siguiente;
- 3) Evaluaciones de los compañeros, en revisiones no automatizadas, se propone revisión por pares en algunos casos, de tal manera que se proporciona una retroalimentación en el proceso de aprendizaje, y;
- 4) Aprendizaje Mezclado, con las universidades asociadas que refuerzan su conocimiento con MOOCs, mezclando lo presencial con lo virtual. (Shafiq, Wani, Mahajan, & Qadri, 2017).

Por otro lado, en un MOOC de Coursera, las video-lecturas y las evaluaciones se agrupan en secciones (normalmente por medio de semanas o lecciones). Cada lección es visible para los estudiantes en un momento predeterminado (el tiempo abierto), con el fin de estructurar el curso. Dentro de una lección, las video-lecturas y las evaluaciones pueden tener una fecha límite o no, para guiar más a los estudiantes a seguir un ritmo de estudio particular. Por último, las evaluaciones también tienen fechas límite (la hora de cierre), y el alumno puede intentar los cuestionarios varias veces, hasta cierto límite de envío (Mukala et al., 2015).

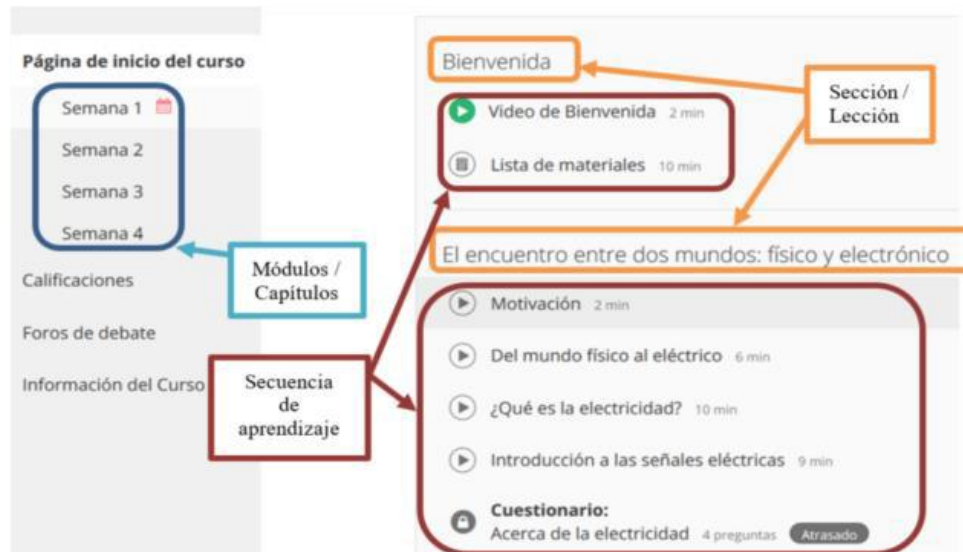


Figura 7: Interfaz principal y organización de los cursos en Coursera (Pérez-Álvarez, Maldonado-Mahauad, & Pérez-Sanagustin, 2018).

3.3. Arquitectura de los datos en Coursera

Coursera subdivide los datos en tres categorías principales: datos generales, datos de foros y datos de identificación personal. En total, el modelo estándar comprende más de 86 tablas (Coursera, 2016) que almacenan información sobre los privilegios de los usuarios, anuncios sobre el curso, todos los foros, evaluaciones, datos de evaluación, calificaciones del curso, detalles de las presentaciones, entre otras, los mismos que van organizando la arquitectura y el modelo de datos que maneja Coursera.

Debido a la gran cantidad de tablas que existen en el modelo de datos, Coursera plantea una arquitectura de datos como muestra la **Figura 8**, y estos grupos son:

- Course information
- Course content
- Learners
- Demographics
- Course progress

- Assesments
- Course grades
- Discussions
- Feedback

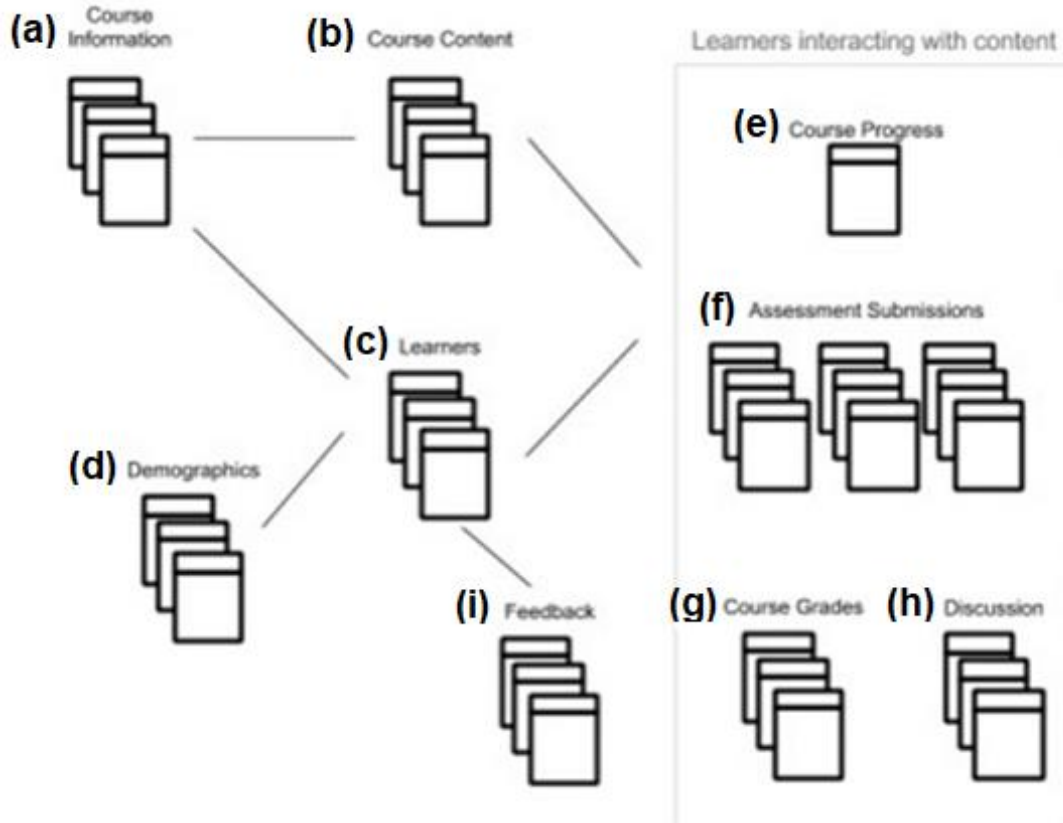


Figura 8: Arquitectura de Datos Coursera (Coursera, 2016)

El contenido de los grupos de tablas antes mencionadas, se alimentan desde que el instructor o tutor crea el curso y los contenidos, hasta que los estudiantes van realizando las actividades, por ello, según la guía de datos de Coursera para desarrolladores e investigadores (Coursera, 2016), se indica la secuencia de la creación de las tablas:

En primera instancia el instructor crea el curso con los datos de información, luego crea las diferentes actividades de las que consta el curso, como evaluaciones, video lecturas, cuestionarios, etc., con ello los grupos Course Information (**Figura 8 (a)**) y Course Content (**Figura 8 (b)**) van almacenando esa información en las diferentes tablas de dichos grupos.

Seguido de esto los alumnos empiezan a inscribirse en el curso y empiezan a alimentar a las tablas de Learners (**Figura 8 (c)**), y opcionalmente los que desean llenan cuestionarios de información más específica del usuario, y se alimenta las tablas del grupo Demographic (**Figura 8 (d)**), cabe resaltar que Coursera no muestra información sensible o sin autorización del usuario. En cuanto a la tabla de la base de datos llamada Demographic, los estudiantes tienen la posibilidad



de colaborar y alimentar de mejor manera los datos para la plataforma, con información en base a preguntas como:

- ¿En qué país vives actualmente?
- ¿Cuál es su género?
- ¿Cuál es el nivel más alto de la escuela que ha completado o el grado más alto que ha recibido?
- ¿Cuál de las siguientes opciones describe mejor su situación laboral actual?
- Además del inglés, ¿qué otros idiomas hablas?

Cuando el usuario empieza el curso, todas sus actividades empiezan a alimentar las tablas de Course Progress (**Figura 8 (e)**). Es importante indicar que existen 16 tipos de actividades en Coursera, que se indican en la **Tabla 1**.

id	Descripción	Categoría
1	Lecture	Lecture
7	Others	-----
10	Graded Programming	Programming
8	Graded Peer	Peer
9	Closed Peer	Peer
14	Ungraded lti	Lti
12	Phased peer	Peer
11	Ungraded programming	Programming
15	Slideshow	Slideshow
4	Peer	Peer
2	Assess open single page	Quiz
16	Staff graded	staffGraded

5	Quiz	Quiz
6	Exam	Quiz
3	Supplement	Supplement
13	Graded lti	Lti

Tabla 1: Tipos de actividades de un MOOC en Coursera (Pérez-Álvarez et al., 2018)

Para acciones más detalladas respecto a las evaluaciones, las tablas del grupo Assessments (**Figura 8 (f)**) empiezan a almacenar esta información, aquí se almacenan las evaluaciones, cuestionarios calificados, los intentos del usuario para resolver una evaluación, entre otros datos. Estas evaluaciones empiezan a generar calificaciones que son agregadas a la tabla Course Grades (**Figura 8 (g)**).

A medida que los usuarios empiezan a interactuar con los foros que provee la plataforma las tablas de Discussions (**Figura 8 (h)**) empiezan a almacenar estos datos y cuando los usuarios proporcionan comentarios hacia el curso sobre su contenido alimentan las tablas de Feedback (**Figura 8 (i)**).

Finalmente es importante indicar el proceso que emplea Coursera para almacenar los eventos o secuencia de actividades de los usuarios de la plataforma, y según la guía para desarrolladores de Coursera (Coursera, 2016), indica que todos los datos producto de las interacciones son enviadas a un servidor Amazon S3 en tiempo real, ya sea por el uso de Coursera en su sitio web o su aplicación móvil, conservando los datos para su posterior exportación. Además la confiabilidad de los datos vienen dada de los grandes esfuerzos de la comunidad de Coursera para la investigación, haciendo el mayor esfuerzo para obtener los datos de forma más organizada y controlada, pero también depende de la señal de internet que se tenga en ese momento, ya que con señal intermitente es complicado almacenar un log de eventos correcto, pero a su vez los esfuerzos de Coursera implementaron la denominada “lógica de reintento”, para intentar mejorar los datos en señales intermitentes de internet, reintentando capturar los datos cuando la solicitud inicial no tuvo éxito.

Los eventos se agrupan de lado del cliente antes de enviarlos al servidor, al realizarlo por lotes las marcas de tiempo tienen una variabilidad de segundos, cabe resaltar que se utiliza la marca de tiempo generada de lado del cliente. Además, Coursera está consciente de la posibilidad de perder datos en la transición de cliente a servidor, por lo que hace una medición, adjuntando un número de secuencia que aumenta monótonamente a cada evento construido en el cliente y una vez que los eventos llegan al servidor, observamos las brechas en los números de secuencia para comprender la pérdida. Se sabe que esta pérdida en promedio no supera el 6%, que es importante pero no

significativo para poder realizar una investigación o experimento. Cabe resaltar que en infraestructuras de internet deficientes como es el caso de India, esta pérdida puede suponer hasta el 15%.

3.4. Principales tablas de la base de datos

En esta sección se muestra las tablas de datos utilizadas para la generación del LOG de eventos, los mismos que nos permitirán analizar y el caso de estudio mediante MP. En la **Figura 9**, se puede observar el modelo de datos simplificado para el desarrollo del caso de estudio.

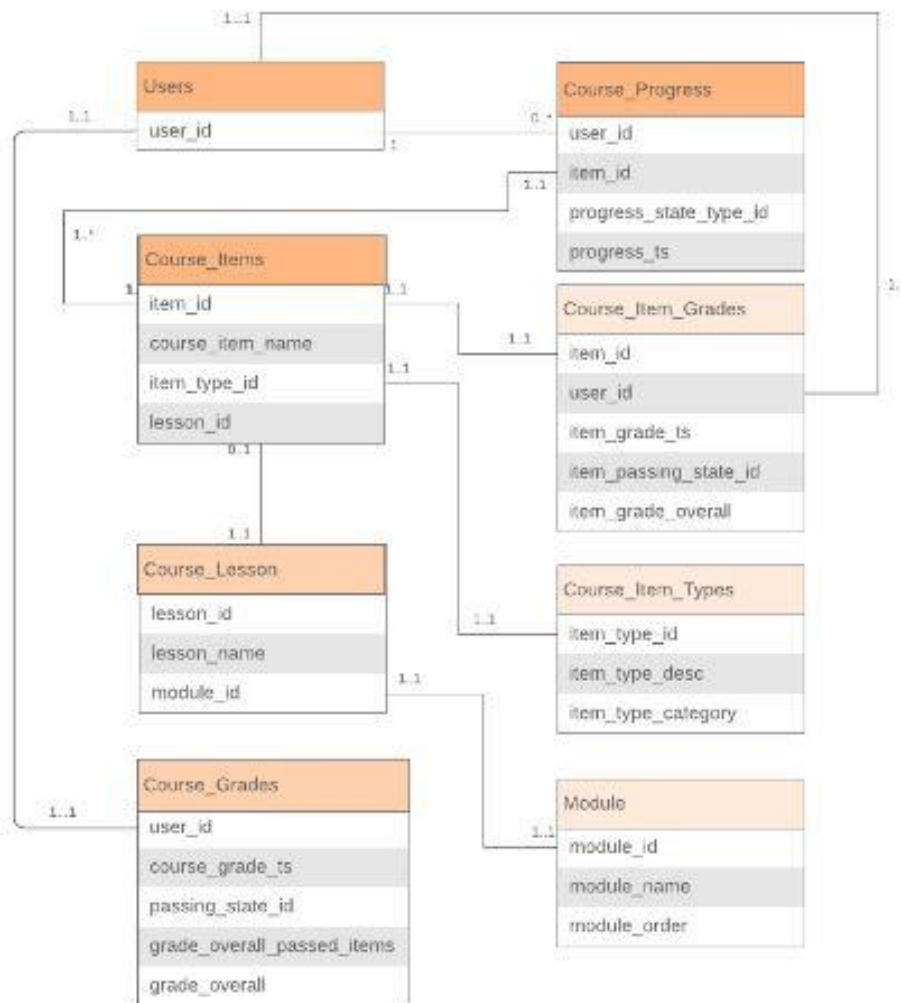


Figura 9: Modelo de datos simplificado para el desarrollo del caso de estudio

Users: Esta tabla almacena la información de los usuarios de la plataforma, independientemente del tipo de rol que éste desempeñe en la plataforma, por ejemplo: Tutor, estudiante, etc., para poder obtener solo los usuarios que son estudiantes del curso es necesario pre-procesar los datos.



Course Grades: Esta tabla almacena, las calificaciones generales, las mismas que se asocian al usuario por el campo *user_id*. Además, todos los registros tienen asociada una marca de tiempo y poseen un campo que indica el número de actividades que completaron en el curso los usuarios.

Course Progress: Esta tabla almacena el progreso del usuario, cabe resaltar que esta tabla incluye tanto actividades calificadas, como no calificadas, además incluye un estado de progreso, los mismos que son: (1) Iniciado y (2) Completado, seguido de la marca de tiempo para cada registro.

Course Items: Esta tabla almacena el nombre de la actividad, el id de la actividad, el id del tipo de actividad y el id de la lección, su importancia radica en poder asociar mediante los ids, otras tablas que enriquecen el contexto del modelo que se diseñó para el caso de estudio.

Course Item Grades: Esta tabla almacena, las calificaciones de las evaluaciones o actividades calificadas del curso, las mismas que se asocian al usuario por el campo *user_id*. Además, todos los registros tienen asociada una marca de tiempo.

Course Item Types: Son las actividades que mostraron anteriormente en la **Tabla 2**, las mismas que se logran integrar al modelo por el id del tipo de actividad que se encuentra en la tabla "Course Items".

Course Lesson: En esta tabla se almacena la información de cada lección y el orden que la lección tiene dentro de cada módulo, el mismo que está asociado por el id del módulo.

Module: En esta tabla se almacena el nombre de todos los módulos y el orden que tienen en el curso, es decir el orden de las semanas para participar en cada módulo.



3.5. Recapitulación.

Coursera fue uno de los primeros proveedores de MOOCs en el mundo y gracias a la visión de sus fundadores Andrew Ng y Daphne Koller de la Universidad de Stanford, llegó a ser la plataforma en el top cinco con la mayor cantidad de usuarios registrados y cursos ofertados.

Los datos que se generan en la plataforma Coursera son extensos. Es por esto que se ha subdividido en “categorías” o grupos grandes para mostrar una representación o visión general de los datos. Cuando se requiere datos específicos estos grupos pueden guiar, para ubicar las tablas que se pueden visualizar para obtener la información necesaria. Los grandes grupos son: Course Information, Course Content, Course Progress, Assesments, Course Grades, Discussions, Feedback, Learners, Demographics, para un mejor entendimiento se recomienda ver la **Figura 8**.

Finalmente, se realizó una aproximación mediante el modelo de datos simplificado, donde se visualizó las tablas más importantes que se consideraron para usarla en el caso de estudio planteado en esta tesina y como se relacionan entre ellas (**Figura 9**).

Capítulo 4: Minería de Procesos

4.1. Introducción

En este capítulo, el lector encontrará una revisión bibliográfica acerca de Minería de Procesos (MP). Aquí se tratará los principales conceptos de la minería de procesos, además de sus tipos y técnicas que se aplican para descubrir modelos. También, se presenta la revisión de las herramientas de software que se utilizan para dar soporte al descubrimiento de modelos de procesos. Finalmente, este capítulo presenta una recapitulación con las ideas principales tratadas.

4.2. Minería de Procesos

La minería de procesos es una disciplina de investigación relativamente joven que se ubica entre la inteligencia computacional y la minería de datos **Figura 10**, por una parte, y la modelación y análisis de procesos, por otra (Wil Van Der Aalst et al., 2011). Al colocar a la MP entre estas dos ramas de la ingeniería, se encuentra que la necesidad de los datos es crucial para poder elaborar el registro de eventos y aprovecharlos de forma significativa. Rudnitckaia (2017), se refiere a la MP como una profesión del futuro, donde las organizaciones que no sean capaces de utilizar “Big Data”, no solo como almacenamiento sino como procesamiento, de una manera inteligente no sobrevivirán. Además, se indica la necesidad de que el software utilizado genere dicha información automáticamente como muestra la **Figura 11**, ya que el conocimiento oculto en los registros de eventos, no se pueden convertir en información procesable, es decir; que los datos son omnipresentes, pero si no se maneja adecuadamente los mismos no sirve para poder realizar el análisis necesario para la generación de un LOG de eventos (Rudnitckaia, 2017).

La aplicación de la minería de procesos en el ámbito educativo es una línea de investigación emergente que centra su atención en modelar, generar y evaluar los procesos que los estudiantes y profesores llevan a cabo durante el desarrollo de una unidad o modulo. En este contexto la granularidad de lo que se entiende por proceso abarca desde el comportamiento de los estudiantes cuando realizan una actividad educativa hasta el flujo de trabajo que coordina la ejecución de todas las actividades de una unidad o módulo (Orellana García, A., 2018). Estas actividades son asociados a una marca de tiempo que ordenará cronológicamente la secuencia de las mismas.

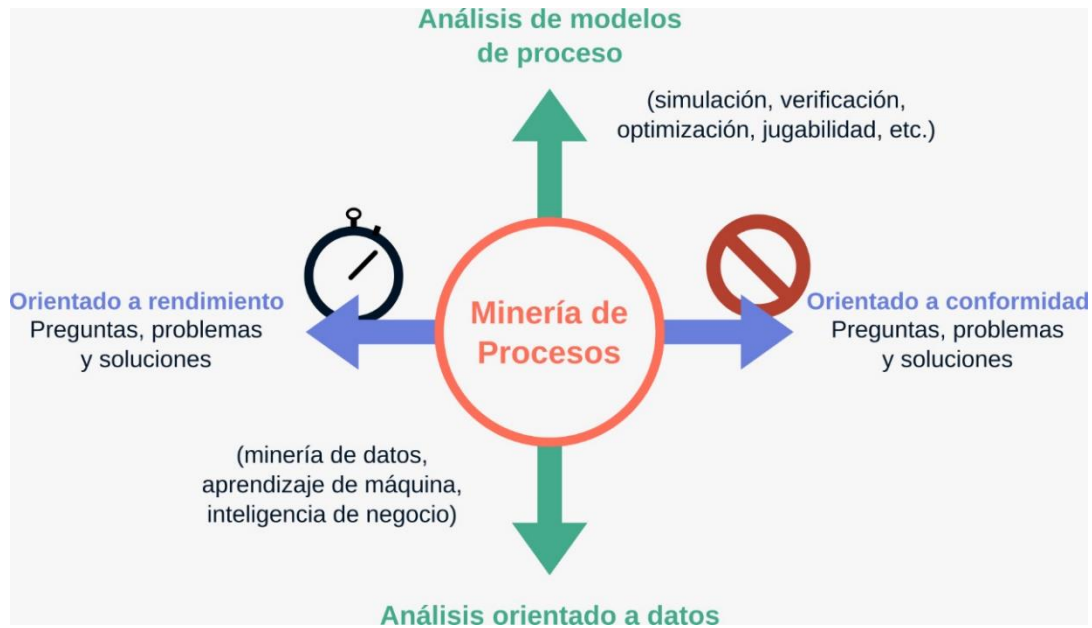


Figura 10: Posicionamiento de PM (Rudnitckaia, 2017)

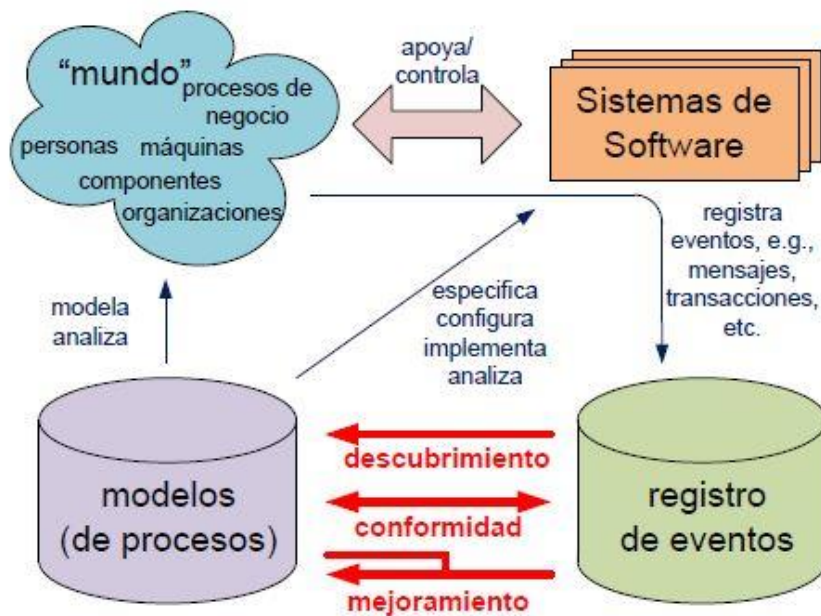


Figura 11: Secuencia para la aplicación del MP

4.3. Tipos de Minería de Proceso

Los principales tipos de MP son: (1) Descubrimiento, (2) Verificación de conformidad y, (3) Mejoramiento, los cuales se pueden observar en la **Figura 12**. A continuación, se realiza una descripción de cada uno de los tipos de MP.

4.3.1. Descubrimiento

Es el primer tipo de MP, es una técnica que a partir de un registro de eventos es capaz de producir un modelo, el mismo que se forma sin usar ninguna información adicional, es decir una técnica con datos a-priori. El modelo descubierto es generalmente un modelo de procesos (red de Petri, Diagrama de Actividad UML, etc.) (Wil Van Der Aalst et al., 2011), **Figura 12 (a)**.

4.3.2. Verificación de Conformidad

En este tipo de técnica de MP, que consiste en comparar un modelo de procesos existente con un registro de eventos del mismo proceso. La salida consiste en información de diagnóstico mostrando las diferencias y similitudes entre el modelo y el registro de datos. Esto se puede utilizar para constatar que los datos generan el mismo resultado y viceversa (Wil Van Der Aalst et al., 2011), **Figura 12 (b)**.

4.3.3. Mejoramiento

Con esta técnica de MP, se busca extender o mejorar un modelo de procesos. Por ejemplo, al usar marcas de tiempo en el registro de eventos, uno puede extender el modelo para mostrar cuellos de botella, niveles de servicio, tiempos de procesamiento, y frecuencias. Esta técnica también necesita como entrada un modelo y un registro de eventos, para obtener un modelo extendido como salida (Wil Van Der Aalst et al., 2011) **Figura 12 (c)**.

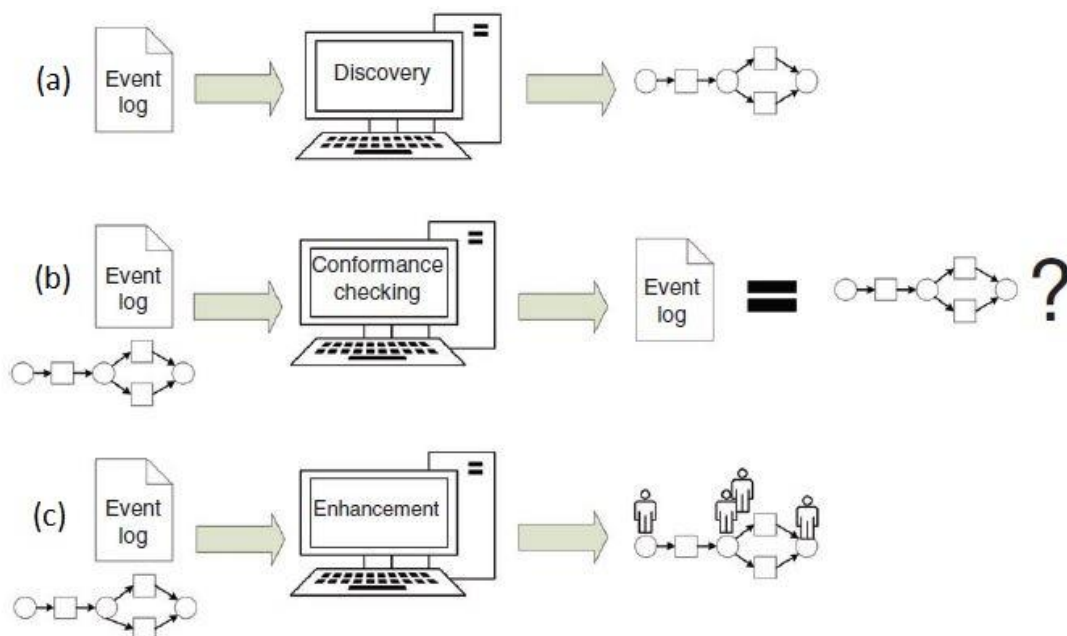


Figura 12: Tipos de Minería de Procesos en términos de entrada y salida (Bogarín, Cerezo, & Romero, 2018)

4.4. Algoritmos de Descubrimiento en MP

En esta sección se hace una introducción a los principales algoritmos de descubrimiento de procesos. Según Van der Aalst (2014), los principales algoritmos de descubrimiento que se encuentran disponibles en las herramientas para análisis son los siguientes:

- Heuristic Miner
- Fuzzy Miner
- Multi Phase Miner

4.4.1. Heuristic Miner

Fue desarrollado por el Dr. Ton Weijters, quien utilizó un enfoque heurístico para abordar muchos problemas existentes con el algoritmo alfa, haciendo que este algoritmo sea mucho más adecuado en la práctica. En la práctica se puede considerar lo siguiente para su uso adecuado (Fluxicon, 2010):

- **Salida:** Se obtiene como salida una red heurística.
- **Cuándo usarlo:** Se puede usar cuando tiene datos de la vida real con no demasiados eventos diferentes, o cuando necesita un modelo de red de Petri para un análisis posterior en alguna herramienta de minería de procesos.

El algoritmo Heuristic Miner, explota la perspectiva del flujo de control de un modelo de proceso. Para hacerlo, solo considera el orden de los eventos dentro de un caso. En otras palabras, el orden de los eventos entre los casos no es importante (Weijters, van der Aalst, & de Medeiros;, 2006).

Heuristic Miner deriva los conectores XOR y AND de las relaciones de dependencia. Para encontrar un modelo de proceso sobre la base de un registro de eventos, el registro debe analizarse en busca de dependencias causales, por ejemplo, si una actividad siempre es seguida por otra actividad, es probable que exista una relación de dependencia entre ambas actividades (Weijters et al., 2006). Puede abstraerse de comportamientos y ruidos excepcionales (dejando fuera los bordes) y, por lo tanto, también es adecuado para muchos registros de la vida real (Fluxicon, 2010).

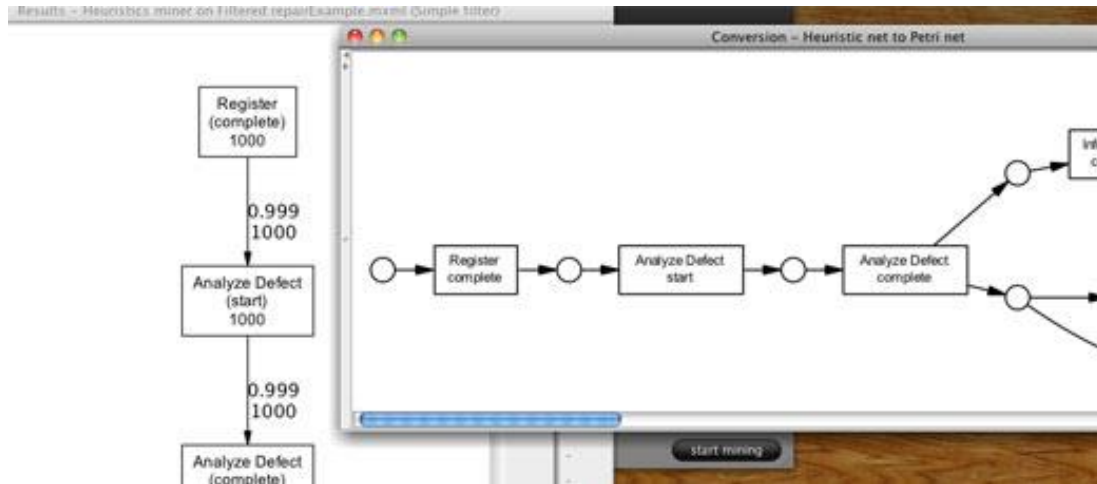


Figura 13: Resultado de un análisis con un Algoritmo de Minería Heurística

4.4.2. Fuzzy Miner

El algoritmo de Fuzzy Miner o Minería Difusa en español, es uno de los algoritmos de descubrimiento de procesos más jóvenes, y fue desarrollado por el cofundador de Fluxicon, Christian W. Günther, en 2007 (Fluxicon, 2010).

- **Salida:** Se un obtiene un modelo difuso.
- **Cuándo usarlo:** cuando tiene datos de registro complejos y no estructurados, o cuando desea simplificar el modelo de manera interactiva.

La característica principal radica en ser el primer algoritmo que aborda directamente los problemas de un gran número de actividades y un comportamiento altamente desestructurado (Fluxicon, 2010), es por ello, que se usa para modelar comportamientos de usuarios en ambientes MOOCs, y la aplicación de este algoritmo en el caso de uso abordado en el siguiente capítulo.

La minería difusa utiliza métricas de correlación para simplificar interactivamente el modelo de proceso al nivel deseado de abstracción. En comparación con el minero heurístico, también puede omitir actividades menos importantes (u ocultarlas en grupos) si tiene cientos de ellas (Fluxicon, 2010).

El algoritmo de Minería Difusa, no se puede convertir a otros tipos de lenguajes de modelado de procesos, pero puede usarlo para animar el registro de eventos en la parte superior del modelo creado para tener una idea del comportamiento dinámico del proceso (Fluxicon, 2010).

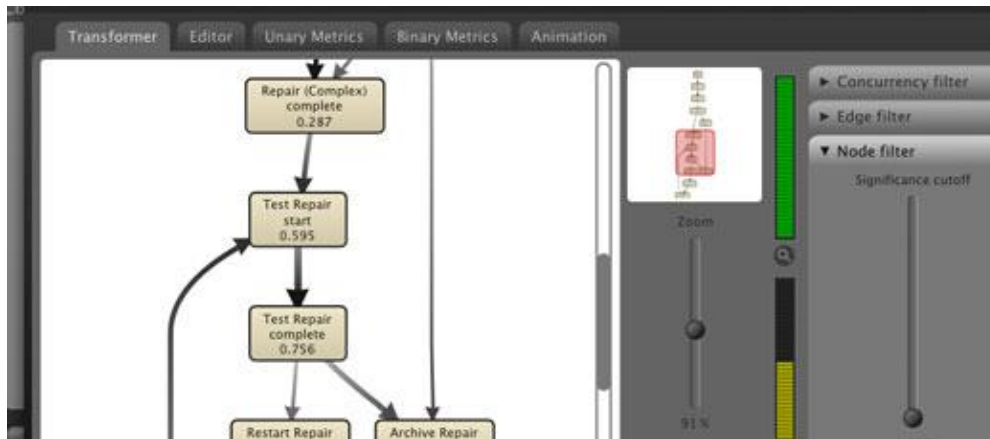


Figura 14: Resultado de un análisis con un Algoritmo de Minería Difusa

4.4.3. Multi Phase Miner

El algoritmo Multi Phase Miner o Minería Multi Fase, se desarrolló por el Dr. Boudewijn van Dongen, un veterano de la minería de procesos y desarrollador líder de la herramienta de minería de procesos ProM, además es el primer algoritmo que utilizó explícitamente la semántica de la compuerta lógica OR, lo que le permite expresar un comportamiento complejo en modelos relativamente bien estructurados. Por ello, se usa cuando tiene datos de registro simples y estructurados, su salida corresponde a una cadena de proceso controlada por eventos (EPC) (Fluxicon, 2010).

- **Salida:** Cadena de proceso controlada por eventos (EPC).
- **Cuándo usarlo:** cuando se tiene datos de registro simples y estructurados.

Este algoritmo pliega los conectores XOR, AND y OR de las ejecuciones y muestra el modelo resultante como un EPC. Una de las ventajas, es que construye un modelo que siempre "se ajusta" al registro de eventos completo. Sin embargo, rara vez es útil para procesos más complejos porque el modelo se vuelve ilegible.

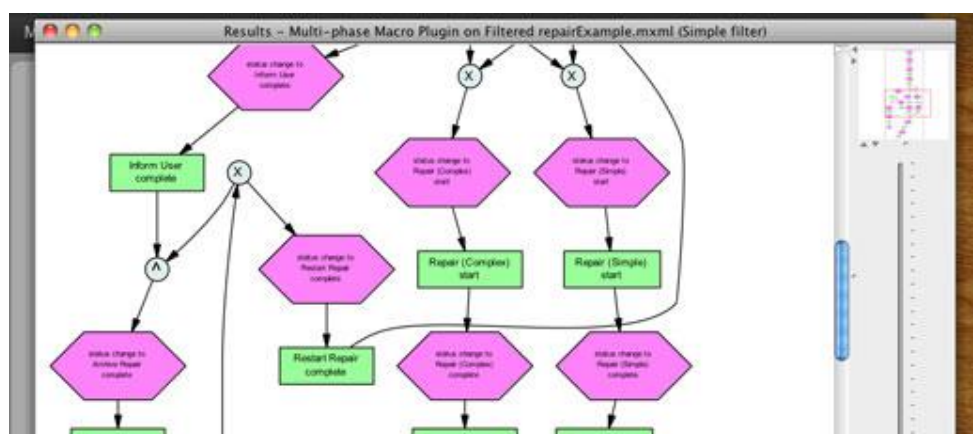


Figura 15: Resultado de un análisis con un Algoritmo de Minería Multi Fase

4.5. Principales herramientas para Minería de Procesos

Debido al impacto significativo de la minería de procesos, los esfuerzos por desarrollar herramientas que permitan su uso automatizado y rápido han crecido, es por ello que existe gran variedad de las mismas en el mercado, a continuación, se mostrarán dos herramientas altamente usadas.

4.5.1. ProM

(ProM, 2019) ProM (Process Mining Framework) es un marco extensible que admite una amplia variedad de técnicas de minería de procesos en forma de complementos. Es independiente de la plataforma, ya que se implementa en Java y se puede descargar de forma gratuita.

ProM 6 se distribuye en partes, lo que ofrece la máxima flexibilidad. Primero, el núcleo de ProM 6 se distribuye como un paquete descargable utilizando la licencia de código abierto GNU Public License (GPL). Esto significa que puede descargar e instalar ProM 6 sin restricciones, pero que cualquier software que use el núcleo también debe distribuirse utilizando la licencia GPL. En segundo lugar, los complementos ProM 6 se distribuyen como paquetes separados, que generalmente usan la licencia de fuente abierta de la Licencia pública menor de GNU (L- GPL). Tenga en cuenta que el uso de software de terceros puede requerir la emisión de un complemento con una licencia diferente a la L-GPL, pero en principio los complementos ProM 6.3 utilizarán la licencia L- GPL. Esto significa que puede descargar e instalar los complementos sin restricciones, y que es libre de distribuir el software que utiliza estos complementos con su propia licencia. Sin embargo, si distribuye una versión modificada de dicho complemento, también debe distribuir este complemento modificado bajo la licencia L- GPL (ProM, 2019).

La misión de ProM, es convertirse en la plataforma de minería de procesos estándar de facto en el mundo académico al establecer una comunidad activa y reconocida de contribuyentes y usuarios, y crear conciencia sobre el poder de la tecnología de minería de procesos mediante la promoción de aplicaciones y la absorción industrial (Process Mining Group, 2016).

Además la visión de ProM, es avanzar activamente en el estado de la tecnología de minería de procesos mediante el desarrollo de métodos que realmente funcionen, creando una comunidad abierta y proporcionando una plataforma estable y fácilmente extensible, que soporte óptimamente la minería de procesos (Process Mining Group, 2016).

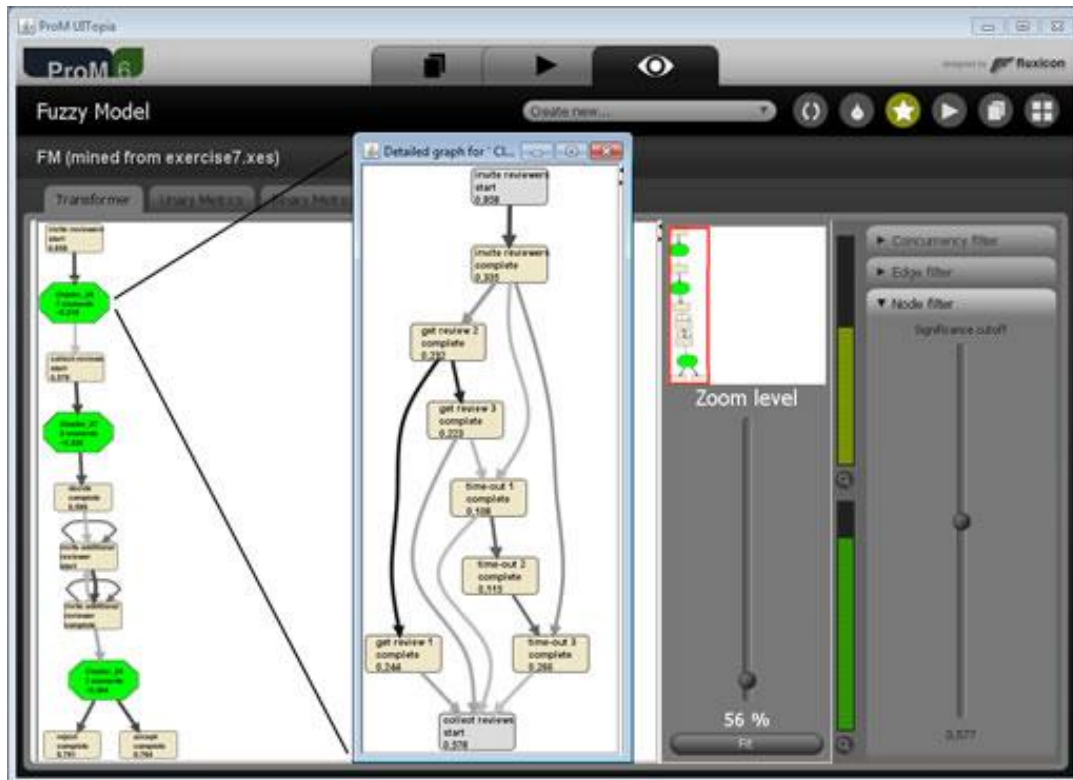


Figura 16: ProM, interfaz gráfica

4.5.2. DISCO

Disco fue construido por ex académicos líderes con más de ocho años de experiencia en minería de procesos. Con Disco, se ha utilizado en experiencia propia de los creadores, innumerables proyectos de minería de procesos y se ha diseñado una herramienta que se adapta perfectamente al flujo de trabajo de los profesionales. Disco contiene los algoritmos de minería de procesos más rápidos (Fuzzy Miner) y el marco de filtrado y gestión de registros más eficiente (Fuxicon, 2019). Entre las principales características de este software se presenta lo siguiente:

Descubrimiento automatizado de procesos: La revolucionaria tecnología de minería de procesos en Disco puede crear mapas de procesos, directamente desde los datos en bruto, de forma automática. Además, se puede elegir el nivel de abstracción deseado, entre seis visualizaciones métricas de proceso proyectadas directamente en el mapa y complementario a esto se puede crear filtros directamente a partir de actividades o rutas (Fuxicon, 2019).

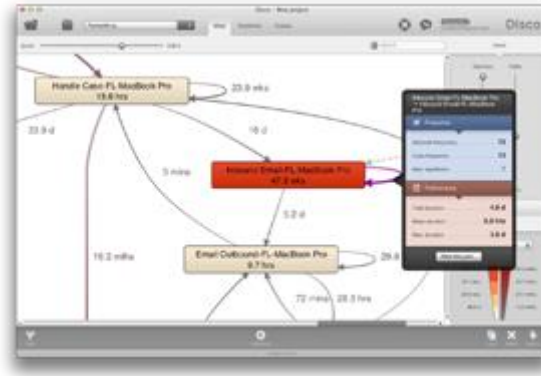


Figura 17: Descubrimiento Automatizado de Procesos en DISCO

Animación del mapa de procesos: Con Disco se puede crear animaciones visualmente atractivas, donde se muestra el proceso tal como sucedió directamente en el mapa de proceso. Es importante recalcar, que la animación puede ayudar a detectar cuellos de botella instantáneamente donde se acumula el trabajo (Fuxicon, 2019).

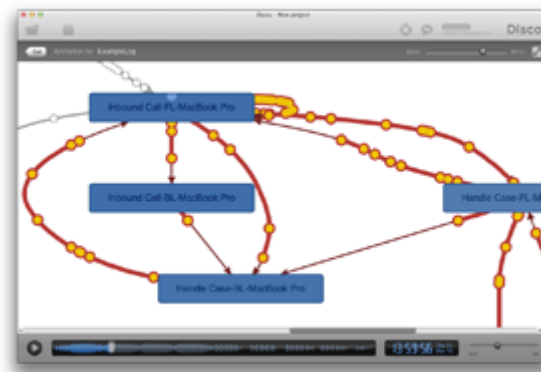


Figura 18: Animación del mapa de procesos en DISCO

Estadísticas detalladas: La vista de estadísticas en Disco puede responder a muchas preguntas, tales como: ¿Con qué frecuencia se ha realizado esta actividad?, ¿En realidad te tardas x cantidad de tiempo en tus actividades?, entre otras. Con esta sección se puede obtener una descripción general de los datos de gráficos interactivos y profundizar en información detallada sobre cada actividad, recurso y valor de atributo (Fuxicon, 2019).



Figura 19: Estadísticas detalladas en DISCO

Casos: En la vista de casos, puede inspeccionar rápidamente el historial completo de los casos relevantes y encontrarlos con la función de búsqueda en vivo de Disco. Y con el explorador de variaciones que posee Disco, puede ver qué casos siguen la ruta del proceso del día normal de trabajo y cuáles son excepciones (Fuxicon, 2019).



Figura 20: Casos en Disco

Filtros: Dispone de un conjunto de filtros de registro fáciles de usar y muy rápidos, Disco hace que sea fácil limpiar los datos de su proceso y enfocar el análisis. En Disco, el filtrado de registros es un ciudadano de primera clase, es decir; que considera los datos como el eslabón más importante. Ya sea que desee desglosar el rendimiento del caso, el marco temporal, la variación, los atributos, las relaciones de eventos o los puntos finales, todo lo anteriormente mencionado lo puede realizar con Disco (Fuxicon, 2019).

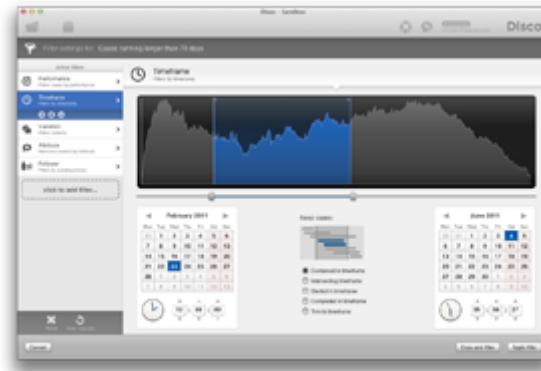
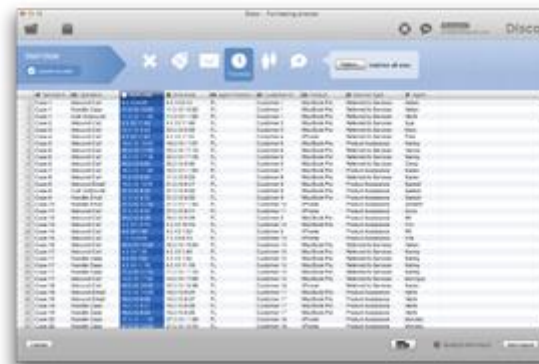


Figura 21: Filtros en Disco

Importar y exportar: Se puede importar archivos CSV y MS Excel complejos que con Disco. Su característica de mapeo inteligente, hace que la configuración sea eficiente, y ordenar los datos en magnitudes grandes es más rápido que con cualquier otra solución. Pero Disco también leerá y escribirá datos pre-configurados en el formato MXML de ProM y en XES, el formato oficial aprobado por IEEE para el intercambio de registros de eventos (Fuxicon, 2019).



Evento	Fecha	Actividad	Estado	Actividad	Estado	Actividad	Estado
Evento 1	2019-01-01	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 2	2019-01-02	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 3	2019-01-03	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 4	2019-01-04	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 5	2019-01-05	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 6	2019-01-06	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 7	2019-01-07	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 8	2019-01-08	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 9	2019-01-09	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3
Evento 10	2019-01-10	Actividad 1	Estado 1	Actividad 2	Estado 2	Actividad 3	Estado 3

Figura 22: Importar y Exportar datos en Disco



4.6. Recapitulación.

En este capítulo, se presentó una introducción a la Minería de Procesos (MP), ubicándola entre inteligencia computacional y la minería de datos por una parte, y la modelación y análisis de procesos, por otra (Wil Van Der Aalst et al., 2011).

Además, se presentó los principales tipos de MP: (1) Descubrimiento: que se basa en tener como entrada un registro de evento y obtener a-priori un modelo que se ajusta al registro de eventos dado. (2) Verificación de la Conformidad, que toma como entrada un modelo y el registro de eventos y obtiene una retroalimentación acerca de similitudes y diferencias entre el modelo inicial y el registro de datos, y finalmente (3) Mejoramiento, que busca extender el modelo de entrada en basa al registro de eventos para mejorar el modelo (p.e: eliminar cuellos de botella).

En cuanto a los algoritmos para descubrimiento de procesos, los principales y más usados son: (1) Heuristic Miner: Se puede usar cuando tiene datos de la vida real con no demasiados eventos diferentes, o cuando necesita un modelo de red de Petri para un análisis posterior en alguna herramienta de minería de procesos (Fluxicon, 2010), (2) Fuzzy Miner: se usa principalmente cuando se posee datos de registros complejos y no estructurados y/o cuando se desea simplificar el modelo de manera interactiva (Fluxicon, 2010), es por ello, que se usa para modelar comportamientos de usuarios en ambientes MOOCs, y la aplicación de este algoritmo en el caso de uso abordado en el siguiente capítulo y finalmente (3) Multi Phase Miner: construye un modelo que siempre "se ajusta" al registro de eventos completo. Sin embargo, rara vez es útil para procesos más complejos porque el modelo se vuelve ilegible (Fluxicon, 2010).

Finalmente se realizó una revisión de las herramientas para análisis de MP, las cuales son: (1) ProM: es un marco extensible que admite una amplia variedad de técnicas de minería de procesos en forma de complementos. y (2) Disco: que posee una tecnología de minería de procesos en Disco puede crear mapas de procesos, directamente desde los datos en bruto, de forma automática.



5. Caso de Estudio

5.1. Introducción

En este capítulo, se desarrollará un caso de estudio y se presentarán los resultados que se obtuvieron. A continuación, se describe el enfoque metodológico del caso de estudio, la estructura del curso, las características de los participantes en el estudio. Posteriormente se presenta el desarrollo de los procedimientos para el análisis de los datos y el uso de las técnicas de MP, con el propósito de dar respuesta a las preguntas de investigación planteadas. Finalmente, en este capítulo se presentarán los principales resultados.

5.2. Tipo y enfoque del caso de estudio

Dado que el objetivo de este estudio es investigar el comportamiento de los estudiantes de un MOOC, el núcleo del caso de estudio es de tipo exploratorio. Para ello, mediante el uso de minería de procesos se busca identificar el comportamiento de dichos estudiantes y caracterizar por medio de las estrategias de aprendizaje empleadas sus sesiones de estudio. Este estudio tiene un enfoque de metodología cuantitativa.

5.3. Contexto: MOOC y participantes

Este estudio se lleva a cabo en el contexto de un MOOC, el cual fue desplegado en la plataforma Coursera y ofertado por la Universidad Católica de Chile en idioma español. El tema del curso trata sobre “La Web Semántica”, y fue ofertado el 07 de diciembre de 2015, finalizando el 25 de febrero de 2016 su primera cohorte. El curso se ofreció de forma abierta para las personas tanto profesionales como no profesionales, que deseen adquirir la experiencia de la Web Semántica.

El curso consistía en siete módulos, donde cada módulo posee una serie de actividades, que comprenden video-lecturas y evaluaciones tanto sumativas como formativas. En términos generales el curso consta de 37 video-lecturas, tres actividades suplementarias que contienen cuestionarios no calificados, 7 evaluaciones, una por módulo y finalmente 2 evaluaciones por pares (es decir; que se evaluaron entre estudiantes).

La información descrita anteriormente indica que solo nueve actividades fueron calificadas y tomadas en consideración para el promedio general del curso. Para aprobar el curso se debía aprobar todas las actividades requeridas y cumplir con el umbral de aprobación del mismo con calificaciones acumuladas, lo cual determina si se aprueba o no el curso, estas actividades son: (1) Evaluaciones

modulares y (2) Evaluaciones por pares. La distribución porcentual de las actividades es la siguiente: video-lecturas 75.51%, cuestionarios suplementarios 6.13%, evaluaciones modulares 14.28% y evaluaciones por pares 4.08%, dando un total de 18.36% de actividades consideradas para aprobar o no el curso (**Figura 23**).

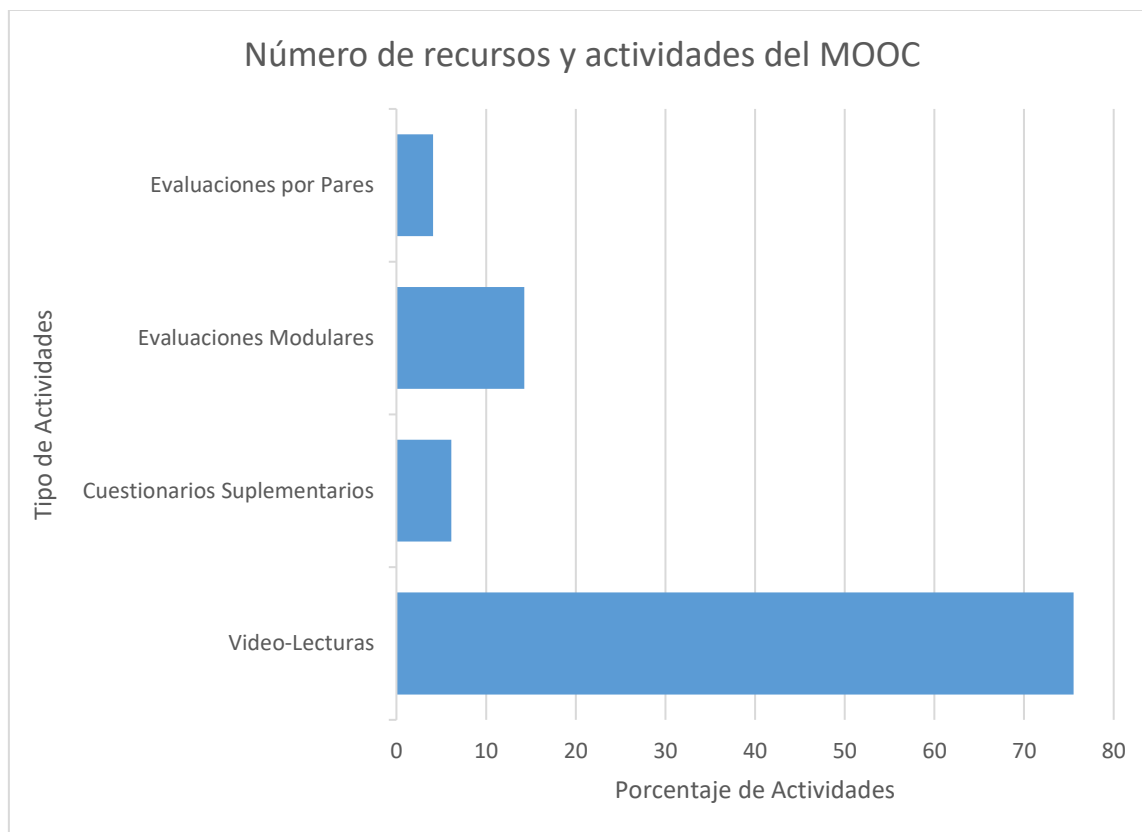


Figura 23: Distribución porcentual de actividades en el MOOC del caso de estudio

En cuanto a los participantes del curso, se registraron 1.550 (N=1.550) estudiantes, de los cuales se realizó un filtro para considerar en este estudio los que al menos terminaron una actividad, dando una reducción de 131 estudiantes, por lo tanto, se consideró a 1.419 estudiantes registrados en el curso y que al menos completaron una actividad. **Las Tablas 2, 3, 4** presentan la información demográfica de los estudiantes.

Género	Porcentaje
Masculino	77.38%
Femenino	22.62%

Tabla 2: Distribución de Género

Rango de Edad	Porcentaje
Menos de 25	1.19%
25-35	30.95%
36-45	38.09%
46-55	19.04%
Más de 55	10.71%

Tabla 3: Distribución de Edades

Nivel Académico	Porcentaje
Educación Secundaria	17.85%
Universidad	54.76%
Postgrado	27.38%

Tabla 4: Distribución de Nivel Académico

5.4. Metodología PM2 para Minería de Procesos

Para responder las preguntas de investigación, se usó la adaptación de la metodología de MP denominada PM2 (Van Eck, Lu, Leemans, & Van Der Aalst, 2015) realizada por Jorge Maldonado y sus colegas en (Maldonado et al., 2016). PM2 es una simplificación y flexibilización de otras metodologías de MP como L*Life-cycle model (W. M. P. van der Aalst, 2014) y que se ha utilizado en distintas disciplinas como en el área de la salud y negocios (Chaves & Córdoba, 2014; Rojas, Muñoz-Gama, Sepúlveda, & Capurro, 2016). Las etapas resultantes de esta adaptación son las siguientes: Objetivos, Extracción de datos, Generación del Log de eventos y Descubrimiento y análisis del modelo. La **Figura 24** presenta de forma gráfica estas etapas y se describen en los siguientes apartados.

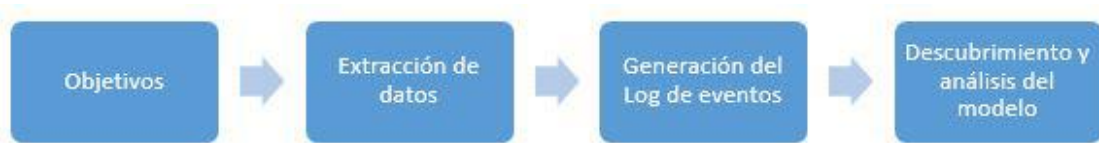


Figura 24: Etapas para generar el modelo de procesos, adaptado de PM2 (Van Eck et al., 2015)

5.4.1. Objetivos

En esta etapa se definen los objetivos. Estos se plantean en base a las preguntas de investigación, las mismas que son:

- P.I.1 ¿Qué tipo de sesiones realizan los estudiantes en un MOOC?
- P.I.2 ¿Qué secuencias de aprendizaje realizan los estudiantes en las diferentes sesiones?
- P.I.3 ¿Qué diferencia existe entre las sesiones de los estudiantes que completan el curso y de los que no?

5.4.2. Extracción de datos

Para estudiar el comportamiento de los estudiantes dentro del curso, la plataforma Coursera genera un conjunto de ficheros de extensión CSV. Estos ficheros contienen todas las tablas de la plataforma Coursera, y que almacena la interacción de los estudiantes con los recursos del curso.

Mediante un proceso de modelado que se muestra en la **Figura 9**, se seleccionaron las siguientes tablas que fueron creadas e importadas a la base de datos PostgreSQL, estas tablas son:

- Users.
- Course grades.
- Course progress.
- Course items.
- Course item grades.
- Course item types.
- Course lesson.
- Course module.

5.4.2.1. Fuentes de datos

Para poder modelar el comportamiento de los estudiantes, Coursera provee un “**export**” que contiene una carpeta con archivos con extensión CSV y HTML. Cada una de las tablas que almacenan la interacción de los estudiantes con los contenidos se almacenan en el fichero de extensión CSV (**Figura 25**) y los de extensión HTML (**Figura 26**) contienen la descripción de cada uno de los campos de las tablas.

A	B	C	D	E	F	G	H	I
course_id	ucchile_user_id	course_grade_ts	course_passing_state_id	course_grade_overall	passed_items	course_grad		
eDkkVBq9EeWjrA6seF25aw,	7ca52036f1633d2aa59ad5b9c697477e33ef41bb,	2015-11-21 08:27:33.665,	0,0,0,0					
eDkkVBq9EeWjrA6seF25aw,	e00be642d20b20ed80e63b85f4bcb61b914054a9,	2015-12-24 17:05:18.362,	0,5,0.575,	0,0				
eDkkVBq9EeWjrA6seF25aw,	8497cc5da7e98762afd4732b0f71bead507e57f9,	2015-10-12 02:03:29.814,	0,0,0,0					
eDkkVBq9EeWjrA6seF25aw,	48271c412bd8f0ab0124a502d55e7f63d00289d9,	2015-12-23 21:26:45.962,	0,0,0,0					
eDkkVBq9EeWjrA6seF25aw,	fdbe2c23abef4ecd645e95a07d020b563aeec1c9,	2015-10-10 23:13:56.636,	0,0,0,0					
eDkkVBq9EeWjrA6seF25aw,	1229c718106c85baa06a31d16b3ca91bdc69be70,	2015-10-17 06:19:21.761,	0,0,0.025,	0,0				
eDkkVBq9EeWjrA6seF25aw,	00a9e8a0e09d4cefa3559a0db8656d8a3ab814b2,	2015-10-08 03:07:40.553,	0,0,0,0					

Figura 25: Fragmento de un archivo CSV de la tabla course_grades

course_grades

Description

[No table description available]

Columns

Name	Description
course_id	22 character unique string identifying an individual course
ucchile_user_id	Encrypted Coursera user id for ucchile data.
course_grade_ts	The timestamp for when a learner's course grade has changed. This occurs everytime a grading event takes place.
course_passing_state_id	A learner can be in one of four "states" with respect to passing a course: 0 means they have started the course but not passed, 1 means they have completed the course without identity verification, 2 means they've completed the course with identity verification, 3 means the course is not passable.
course_grade_overall_passed_items	The number of items in the course that the learner has passed.
course_grade_overall	The grade a learner has in the course, according to the instructor-set grading policy, based on passed items.
course_grade_verified_passed_items	The number of items in the course that the learner has passed and verified their identity.
course_grade_verified	The grade a learner has in the course, according to the instructor-set grading policy, based on items that have been passed and for which the learner's identity has been verified.

SQL create statement

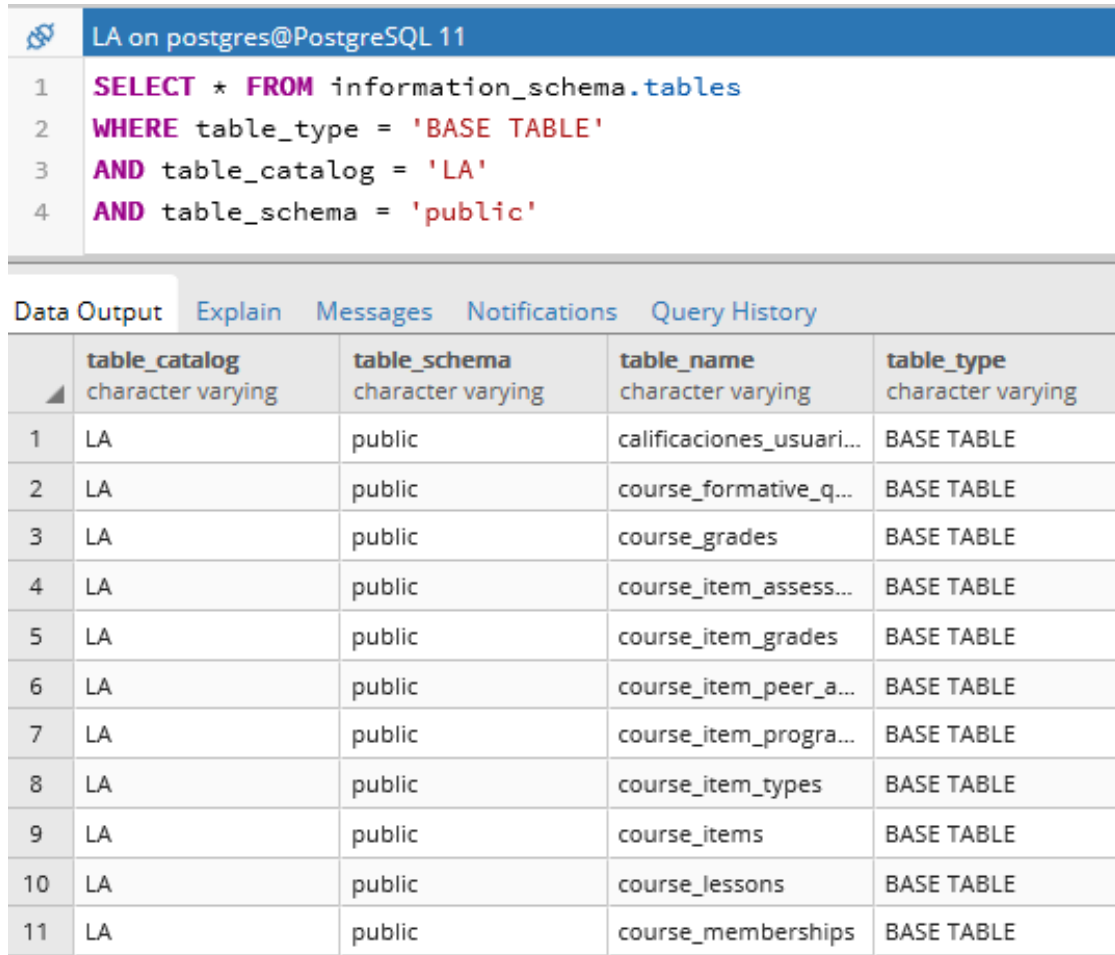
```
CREATE TABLE course_grades (
  course_id varchar
  ,ucchile_user_id VARCHAR(50) NOT NULL
  ,course_grade_ts timestamp
  ,course_passing_state_id int4
  ,course_grade_overall_passed_items int4
  ,course_grade_overall float4
  ,course_grade_verified_passed_items int4
  ,course_grade_verified float4
);
```

Figura 26: Ejemplo de un fichero HTML de la tabla course_grades

Para poder utilizar los datos y realizar consultas se utilizó la base de datos relacional PostgreSQL, cabe resaltar que, en cada fichero HTML, incluye el código o script para crear las tablas, para posterior a eso poder importar los datos a la base de datos desde los ficheros de extensión CSV, es importante

mencionar que no es necesario cargar todas las tablas a la base de datos, sino solo las necesarias, las mismas que fueron definidas en el modelo de datos del capítulo 3, en la sección 3.5, específicamente en la **Figura 9**.

Una vez que se importó las tablas necesarias a PostgreSQL acorde al modelo de datos, se realizó algunas consultas y se crearon tablas adicionales con la información que más adelante será utilizada (**Figura 27**).



```
LA on postgres@PostgreSQL 11
1 SELECT * FROM information_schema.tables
2 WHERE table_type = 'BASE TABLE'
3 AND table_catalog = 'LA'
4 AND table_schema = 'public'
```

	table_catalog character varying	table_schema character varying	table_name character varying	table_type character varying
1	LA	public	calificaciones_usuari...	BASE TABLE
2	LA	public	course_formative_q...	BASE TABLE
3	LA	public	course_grades	BASE TABLE
4	LA	public	course_item_assess...	BASE TABLE
5	LA	public	course_item_grades	BASE TABLE
6	LA	public	course_item_peer_a...	BASE TABLE
7	LA	public	course_item_progra...	BASE TABLE
8	LA	public	course_item_types	BASE TABLE
9	LA	public	course_items	BASE TABLE
10	LA	public	course_lessons	BASE TABLE
11	LA	public	course_memberships	BASE TABLE

Figura 27: Fragmento de tablas importadas del modelo de datos Coursera y auxiliares del caso de estudio

5.4.2.2. Limpieza de datos

Posterior a esto se utilizó un software para depurar datos y cambiar términos de registros con ciertas condiciones, de modo que sea más fácil la comprensión de los resultados, esta herramienta es de minería de datos y su nombre es Pentaho Data Integration, el cual permite a los usuarios que usan la herramienta ingerir, mezclar, limpiar y preparar datos diversos de cualquier fuente, con ayuda de herramientas visuales para eliminar la codificación y la complejidad, Pentaho pone los datos de mejor calidad al alcance de TI y el negocio (Hitachi Vantara, 2019). Alguna de las transformaciones que se realizaron fueron las siguientes:

1. En las calificaciones de las evaluaciones, el valor “null” era recurrente, ya que si el estudiante no rendía la evaluación no se colocaba la calificación de cero (0), sino el valor que tomaba era de “null”, para ello con la herramienta Pentaho Data Integration, mediante la transformación llamada “if field value is null”, se configuro que cambie los valores de null por el número 0.

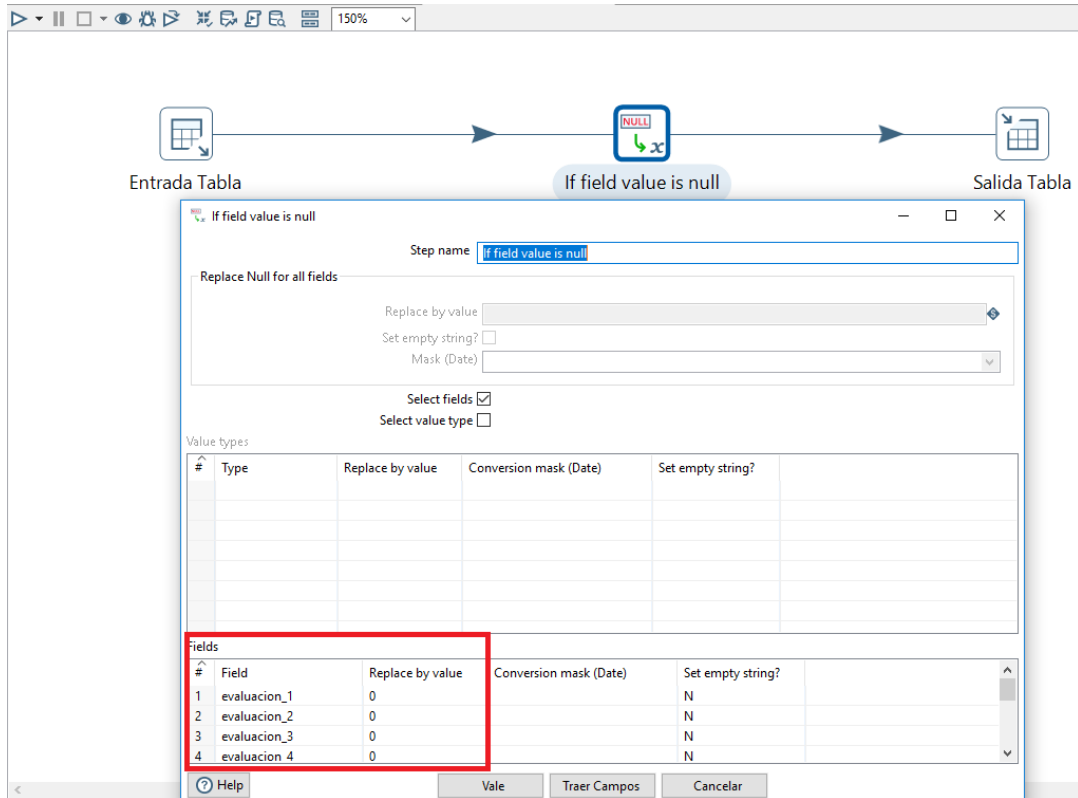


Figura 28: Pentaho Data Integration - Transformación "if field value is null"

2. Existen actividades que fueron necesarias reemplazar, por ejemplo; para una mejor comprensión del log de actividades que se seguía preparando, fue necesario utilizar una herramienta de transformación llamada “mapeo de valores”, que consiste en cambiar el valor de un campo por otro, en este caso corresponde a un cambio en las etiquetas para dar sentido al análisis y estos sean más descriptivos, ya que en muchos casos estos valores eran el resultado de una concatenación de valores en diferentes idiomas.

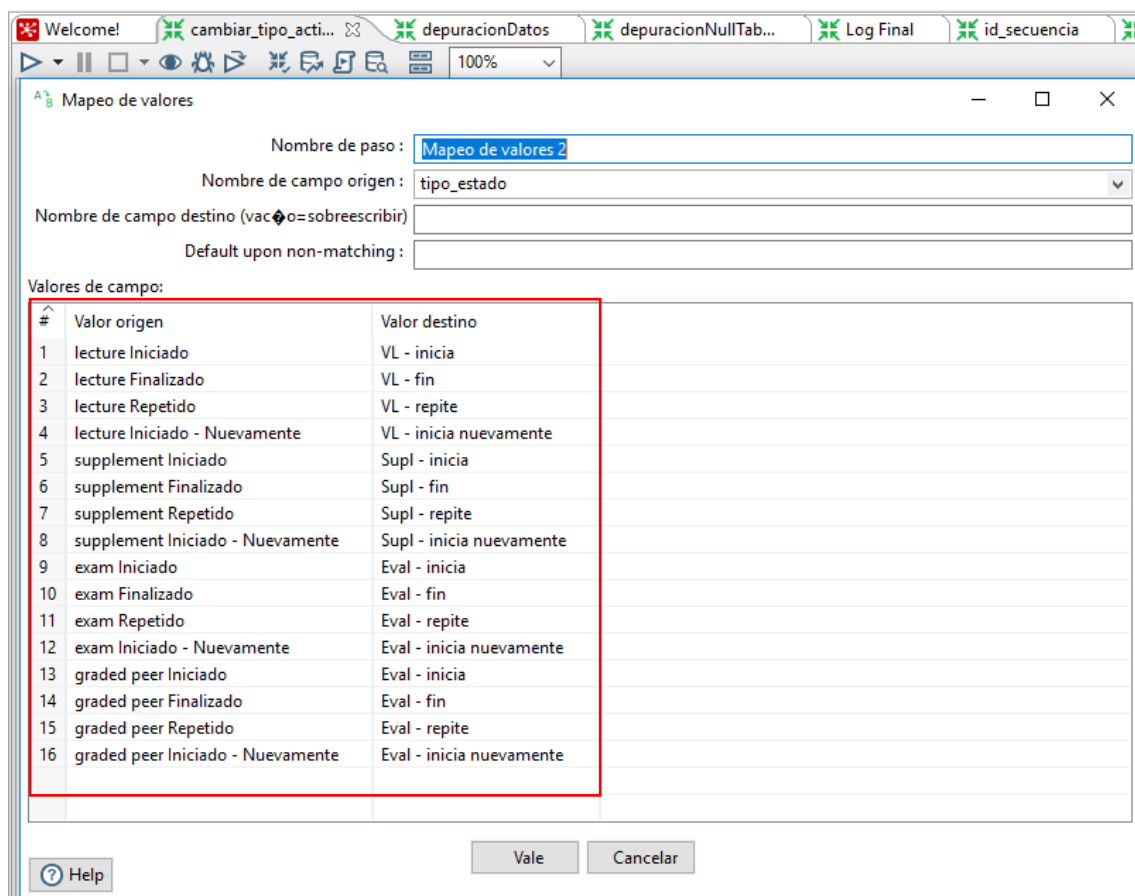


Figura 29: Pentaho Data Integration - Transformación "Mapeo de Valores"

Como siguiente paso, mediante scripts que se realizaron en el lenguaje de programación Python, se obtuvo información que fue necesario calcularla, por ejemplo; se obtuvo el número de sesiones que realizó el estudiante en el MOOC. Para esto se desarrolló un algoritmo, que se basó en calcular los tiempos de inactividad entre actividades, y que estos no superen los 40 minutos, ajustándose al rango de inactividad que plantea Kovanović (2015) en una de sus investigaciones. Para considerar estos tiempos de inactividad, se agruparon todas las actividades de cada estudiante y mediante la marca de tiempo (un campo generado en la base de datos cuando el estudiante realiza alguna actividad), se calculó dicha marca con la anterior, si ésta estaba dentro de los 40 minutos se seguía considerando parte de una misma sesión, en caso contrario, continuaba hacia la siguiente sesión. Dentro de este mismo algoritmo se calculó la duración de cada sesión, que consiste en una resta entre la marca de tiempo de la última actividad dentro del rango de inactividad planteada, con la primera marca de tiempo, estos valores se calcularon en minutos. Finalmente, al haberlo agrupado por estudiantes, fue posible agregar una lista de las actividades que se realizaban por cada sesión, con esto mejora la capacidad de interpretación de los datos ya que se puede saber qué actividad realizó en determinada sesión de estudio (**Anexo 2**).

Además, se agregó un campo muy importante para el desarrollo del presente estudio de caso, ya que se agregó a cada actividad de cada usuario un estado,

los cuales se basaron en la etiqueta que el modelo de datos le otorgaba, este consistía en “1” como estado “Iniciado” y “2” como estado “Finalizado”, pero al momento de realizar el análisis se pudo observar que en realidad no eran suficientes dichos estados, ya que los estudiantes iniciaban nuevamente las video-lecturas cuando no las terminaban en primera instancia, y cuando las terminaban en muchos casos las volvían a repetir, pero dichos estados se mantenían siempre en “1” y “2”, y fue lo mismo para el resto de actividades principalmente para las evaluaciones. Para esto, se consideró implementar un algoritmo, que al igual que el algoritmo anterior (**Anexo 2**), agrupo a los estudiantes con todas sus actividades y mediante este identificador propuesto en el modelo original de datos de Coursera (“1” y “2”), los estados fueron analizados y cuando una misma actividad de un estudiante aparecía por primera vez en estado “1”, el estado agregado en el nuevo campo era (1) Iniciado, mientras que nuevamente aparecía dicha actividad con el mismo estado, es decir con “1”, quería decir que el estudiante nunca completó dicha actividad, por ende se utilizó una nueva etiqueta para ese estado en nuestro nuevo campo que se denominó como (2) Inicia-nuevamente, por otro lado cuando una actividad de un estudiante aparecía por primera vez con la etiqueta “2”, esta actividad fue terminada y se agregó el estado (3) Finalizado, finalmente cuando la misma actividad aparecía más de una vez con la etiqueta “2”, indicaba que el estudiante que finalizó una actividad la repitió nuevamente (**Anexo 3**). Luego finalmente, estos datos se concatenaron con el tipo de actividad realizada y fueron cambiados por textos más descriptivos como muestra la **Figura 29**, y a continuación, se indica una tabla (**Tabla 5**) resumen con la información relacionada:

Interacción	Definición
VL-inicia	Inicia una video-lectura, es un estado que se mantiene apenas un estudiante abre una video lectura.
VL-inicia nuevamente	Una video-lectura que fue iniciada y previamente no se completó.
VL-fin	Cuando una video-lectura se terminó de ver.
VL-repite	Cuando una video lectura que previamente se terminó de ver es abierta nuevamente.
Eval-inicia	Inicia una evaluación, es un estado que se mantiene apenas un estudiante intenta resolver la evaluación.
Eval-inicia nuevamente	Se trata de una evaluación que anteriormente fue abierta, pero que no se completó o no se aprobó.

Eval-fin	Cuando una evaluación fue aprobada en base al umbral de calificaciones requerido.
Eval-repite	Cuando una evaluación previamente aprobada, es abierta nuevamente.

Tabla 5: Definición de las interacciones con los tipos de actividades del curso

Finalmente, se realizó un script adicional para obtener información para saber el número de interacciones del usuario por actividad e interacciones totales, para ello se obtuvo el número de interacciones con las video-lecturas y con los exámenes que son las actividades más recurrentes. Este algoritmo se basó en un contador aplicado a cada actividad en cada sesión, cuando se cambiaba de sesión este contador vuelve a tomar el valor de cero y empieza nuevamente a contar las interacciones, sucesivamente para cada sesión de estudio (**Anexo 4**).

```
1 import psycopg2
2 #Conectar PostgreSQL con Python
3 try:
4     connection = psycopg2.connect(user = "adminPython",
5                                   password = "Bernal 2206 1993",
6                                   host = "127.0.0.1",
7                                   port = "5432",
8                                   database = "LA")
9
10    cursor = connection.cursor()
11
12    #Seleccionar el log para el cluster
13    sql= "select * from user_session_califications_id_sequence;"
14    #ejecutar comando sql
15    cursor.execute(sql)
16    usuarios = cursor.fetchall();
17    print("Conectado a la Base de Datos de Learning Analytics")
18
19 except (Exception, psycopg2.Error) as error :
20     print ("Error while connecting to PostgreSQL", error)
```

Figura 30: Fragmento de código de los scripts mostrados en los Anexos 1 y 2

5.4.3. Generación del log de eventos

Posterior a la etapa de extracción, se generan los logs de eventos. Esto a partir de la integración de los datos que se obtuvieron y depuraron, que constan de la información detallada de como los estudiantes fueron desarrollando el curso a lo largo del tiempo dentro de la plataforma Coursera. La información fue recopilada en una tabla del modelo de datos que nos servirá para responder las preguntas de investigación planteadas.

A continuación, se detalla cómo un solo log de eventos se divide según el análisis que sea necesario, en base a las preguntas de investigación:

P.1.1 ¿Qué tipo de sesiones realizan los estudiantes en un MOOC? Para dar respuesta a esta pregunta de investigación se generó una tabla en la base de datos con la información necesaria, para poder develar el tipo de sesiones que los estudiantes realizan, como muestra la **Tabla 6**. Cabe resaltar, que las columnas o variables que se utilizaron para caracterizar las sesiones son las siguientes:



- `duration_per_session`: contiene el tiempo en minutos de la duración de cada sesión de los estudiantes.
- `num_int_vl`: contiene el número de interacciones con actividades de video lectura, que el estudiante realiza por sesión.
- `num_int_exam`: contiene el número de interacciones con actividades de evaluación, que el estudiante realiza por sesión.
- `num_int_total`: contiene el número de interacciones con todas las actividades del curso.

Finalmente, se procedió a utilizar técnicas de aprendizaje máquina para agrupar a los estudiantes según sus características. En este caso se hizo uso del algoritmo k-means. El número de clusters resultantes fueron comprobados de forma cualitativa y cuantitativa mediante la técnica del codo en la curva y el score de silueta. Más adelante se muestra el desarrollo completo de este proceso. Este campo adicional puede ser observado en las **Tablas 7 y 8** en la última columna de ambas tablas.

<code>case_id</code>	<code>nro_session</code>	<code>duration_per_session</code>	<code>num_int_vl</code>	<code>num_int_exam</code>	<code>num_int_total</code>
003a30 18747ce 400e80 a6689b 758498 861d7ca dc	1	"10.46"	7	0	7
00f066f 678584 77a97a 61a52fa 8ccdb0e 2c4215b	2	"55.58"	7	3	10
011ff41 dfa7cc2 cf9bb89 a73fd9a c1ac74e ef4d3	3	"38.51"	8	0	8
014f7e3 a7a9fbb 9d7445 560628f bcefdcd c81970	4	"66.63"	2	8	10
014f7e3 a7a9fbb 9d7445	5	"65.33"	13	13	26

560628f bcefdcd c81970					
.....					

Tabla 6: Fragmento de los datos usados para generar el clúster

P.I.2 ¿Qué secuencias de aprendizaje realizan los estudiantes en las diferentes sesiones? Para dar respuesta a esta pregunta de investigación se generó un log de eventos, que permite estudiar las secuencias de los estudiantes desde dos enfoques distintos pero complementarios: a) a nivel micro; b) a nivel macro. El primero considerado como nivel micro que se indica en la **Tabla 7**, que consisten en la interacción con los recursos del curso, como por ejemplo, video-lecturas, evaluaciones, cuestionarios, entre otros, y nivel macro que se indica en la **Tabla 8**, al conjunto de acciones por tópico y por semanas del MOOC (Maldonado et al., 2016).

1) A nivel de interacción con los objetos del curso (micro): a este nivel se definen los eventos como el resultado de las interacciones del estudiante con los objetos del MOOC (videos-lecturas, evaluaciones, cuestionarios, evaluaciones por pares) en sesiones de aprendizaje. Estas actividades también se organizan según la recurrencia que se ha tenido sobre la misma, es decir; los estados pueden ser: (1) Iniciado, (2) Repetido, (3) Finalizado y (4) Inicia nuevamente (este estado se repite frecuentemente en las evaluaciones, ya que mientras no se apruebe la misma no cambia el estado hacia Finalizado o Repetido) (**Tabla 7**).

Case_id	Time stamp	Actividad	Sesión	Estado	Aprueba o Reprueba	Clúster
"003a30187 47ce400e80 a6689b7584 98861d7cad c"	"2015-10- 13 17:03:42.9 58"	"Video-Lectura"	1	"VL - inicia"	0	0
"39797a087 1c0ca36462 134baec8ae d963968c98 b"	"2015-12- 12 09:04:30.7 47"	"Evaluación"	10	"Eval-inicia- nuevamente"	1	0
"25e0fdd4ce 4f08c7921e1 c9fa227b0b6 6ed7923a"	"2015-10- 08 17:29:26.2 57"	"Cuestionario"	3	"Supl-fin"	0	1

.....

Tabla 7: Fragmento del log de eventos utilizado para el análisis a nivel micro

2) A nivel de interacción con las lecciones del curso (macro): a este nivel se definen los eventos como resultado de las interacciones del estudiante a lo largo de las 7 semanas con las lecciones del curso (**Tabla 8**).

Case_id	Time stamp	Lección	Sesión	Aprueba o Reprueba	Clúster
"003a3018747ce400e80a6689b758498861d7cadc"	"2015-10-13 17:03:42.958"	"Vocabularios RDF"	3	0	1
"39797a0871c0ca36462134baec8aed963968c98b"	"2015-12-12 09:04:30.747"	"Conceptos de RDF Schema"	6	1	1
"25e0fdd4ce4f08c7921e1c9fa227b0b66ed7923a"	"2015-10-08 17:29:26.257"	"Carta de bienvenida"	1	0	0

.....

Tabla 8: Fragmento del log de eventos utilizado para el análisis a nivel macro

P.I.3 ¿Qué diferencia existe entre las sesiones de los estudiantes que completan el curso y de los que no? Para dar respuesta a esta pregunta de investigación, en base al log de eventos que tiene información acerca de si el estudiante aprueba o no el curso, en la cual existe el identificador "0" para quienes no aprueben el curso, es decir; no lograron superar el umbral de aprobación del mismo que corresponde al 80% de las evaluaciones en promedio. Por otra parte, el identificador "1" es asignado a los estudiantes que sí aprueben el curso, por consecuencia obtuvieron un promedio general de 80% o más de las evaluaciones.

El log de eventos que se generó, según sea necesario como se mostró en las **Tablas 7 y 8**, en cada caso particular se consideran solo ciertas columnas con datos. En el primer caso, servirá para analizar la interacción con los objetos del curso y que se compone de: (a) Case ID: que identifica de forma única al conjunto de eventos, (b) Time stamp: que indica el inicio de un evento, (c) Actividad: que indica la interacción con un objeto del curso; (d) Sesión: que indica el número de

la sesión, e) Estado: que indica el estado de una actividad (iniciado, repetido, finalizado, inicia nuevamente), f) Aprueba o Reprueba; que indica si el estudiante culminó con éxito o no el curso y g) Clúster: que indica el grupo en el que se asignó dicha sesión según el algoritmo k-means. Un fragmento de este tipo se muestra en la **Tabla 7**. En el segundo caso, el análisis del log de eventos servirá para analizar la interacción del estudiante con las actividades del curso, donde se repiten algunos campos, pero varía “Actividad”, en este caso va el nombre de la actividad y esta puede ser de cualquier tipo (video-lectura, evaluación, cuestionario, revisión por pares) (**Tabla 8**).

5.4.4. Descubrimiento y análisis de los modelos

Posterior a la generación del log de eventos, se procede a realizar el descubrimiento de los modelos para su posterior análisis.

P.1.1 ¿Qué tipo de sesiones realizan los estudiantes en un MOOC? Para responder esta pregunta de investigación, se realizó un enfoque en base a las sesiones, las cuales se armaron con variables que recopilan información que no está dentro del modelo de datos de Coursera de forma explícita, es decir; tuvieron que calcularse en base a las interacciones de los estudiantes con la plataforma (**Anexo 2**) y (**Anexo 4**). Para cada uno de los estudiantes se calcularon las siguientes variables: (1) duración de cada sesión (**Figura 31 (a)**), (2) número de interacciones con las video-lecturas (**Figura 31 (b)**), (3) número de interacciones con las evaluaciones (**Figura 31 (c)**), (4) número de interacciones totales por sesión (**Figura 31 (d)**), un fragmento de cómo son los datos reales que se agruparon por cada estudiante y sesión se indica en la **Tabla 6**. Esto se realizó con el objetivo de caracterizar grupos de estudiantes y analizar el tipo de sesiones de estos grupos.

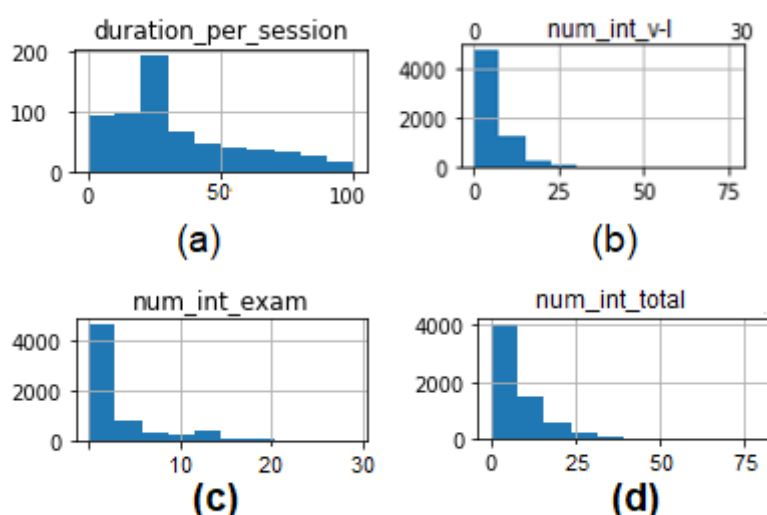


Figura 31: Distribución de las variables para caracterizar los grupos del clúster

Posterior a la identificación de variables para el agrupamiento de los estudiantes y sus sesiones (**Tabla 6**), se procedió a utilizar técnicas de aprendizaje máquina para agrupar a los estudiantes según sus características. En este caso se hizo uso del algoritmo para agrupamiento de estudiantes k-means (**Figura 32**). El número de clústers resultantes fueron comprobados de forma cualitativa y cuantitativa. La primera fue la técnica cualitativa del “Codo en la Curva” (**Figura 33**), que se basa en la apreciación visual de la gráfica, dado que en el eje “x” se indica el número de clústers y mediante una función en el eje “y” forma un “codo” en alguna parte del plano (enmarcado en color rojo de la **Figura 33**). Por otro lado, la técnica cuantitativa de la “Silueta” (**Figura 34**), que se usó como índice para medir la calidad de una agrupación final (dos clusters), donde el promedio de las distancias que más se aproxime a “1” es el valor idóneo de clusters (Lletí, Ortiz, Sarabia, & Sánchez, 2004).

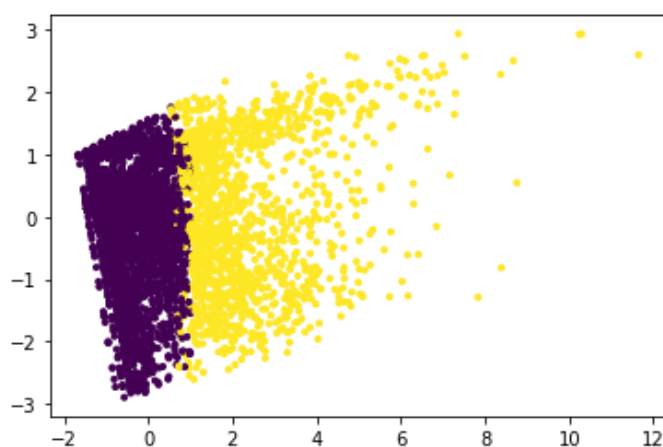


Figura 32: Clúster 0 (morado) y clúster 1 (amarillo)

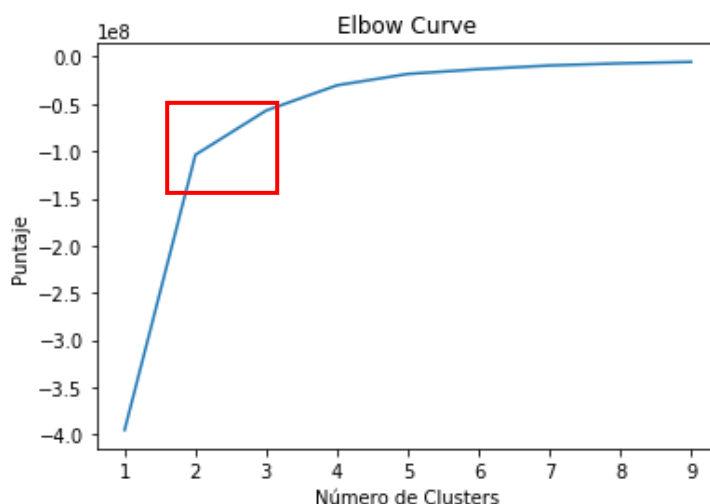


Figura 33: Elbow Curve, para determinar el número ideal de grupos

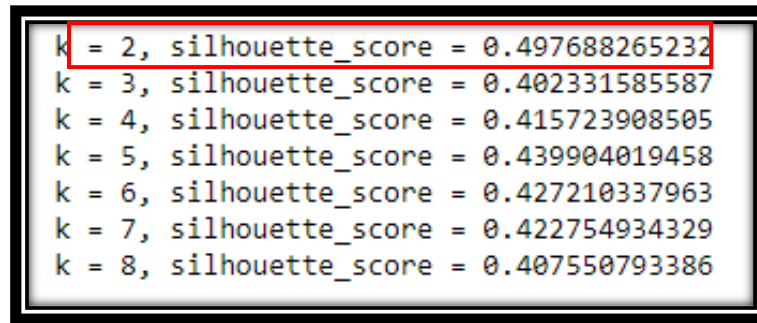


Figura 34: Silhouette Score para determinar el número ideal de grupos

Los grupos que se obtuvieron son: (1) Clúster 0 y (2) Clúster 1, los cuales se describen a continuación:

Clúster 0: formado por 1.292 estudiantes. Estos estudiantes trabajan en promedio por cada sesión 29.74 minutos dentro del MOOC. Interactúan en promedio con 3.53 video-lecturas y 0.7 evaluaciones. Dado que el grupo es determinado por las características de las sesiones, un mismo estudiante puede estar en el Clúster 0 o el Clúster 1, según su comportamiento por sesión, de esta manera en este grupo forman parte los siete estudiantes que aprobaron el curso y 1.285 estudiantes no aprobaron.

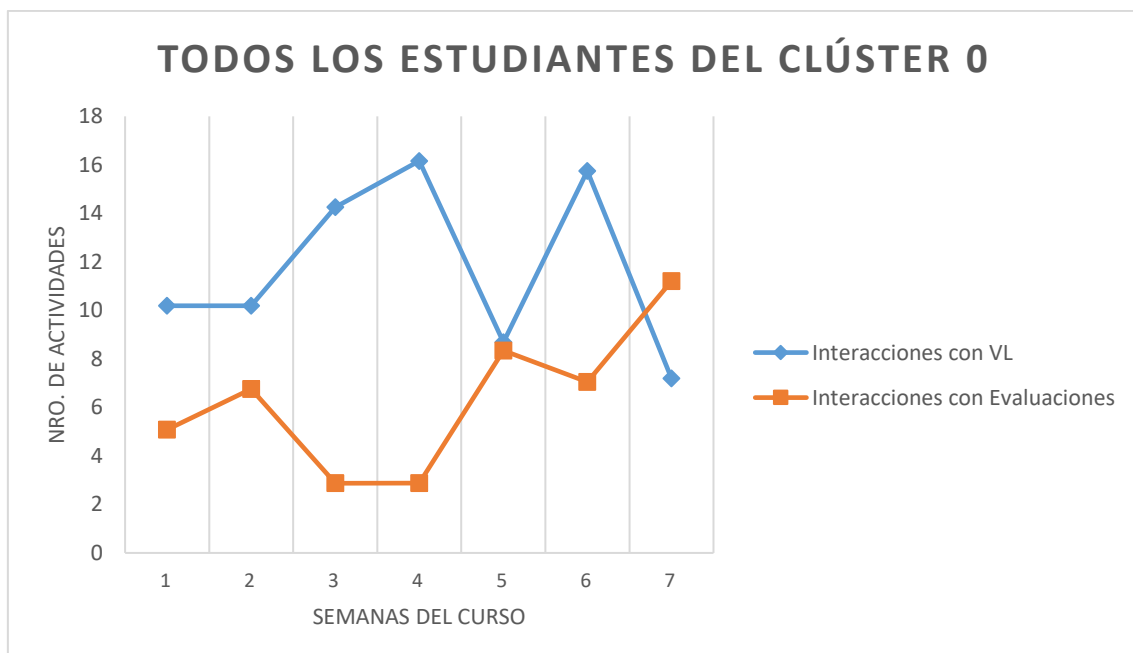


Figura 35: Comportamiento promedio por semana Clúster 0

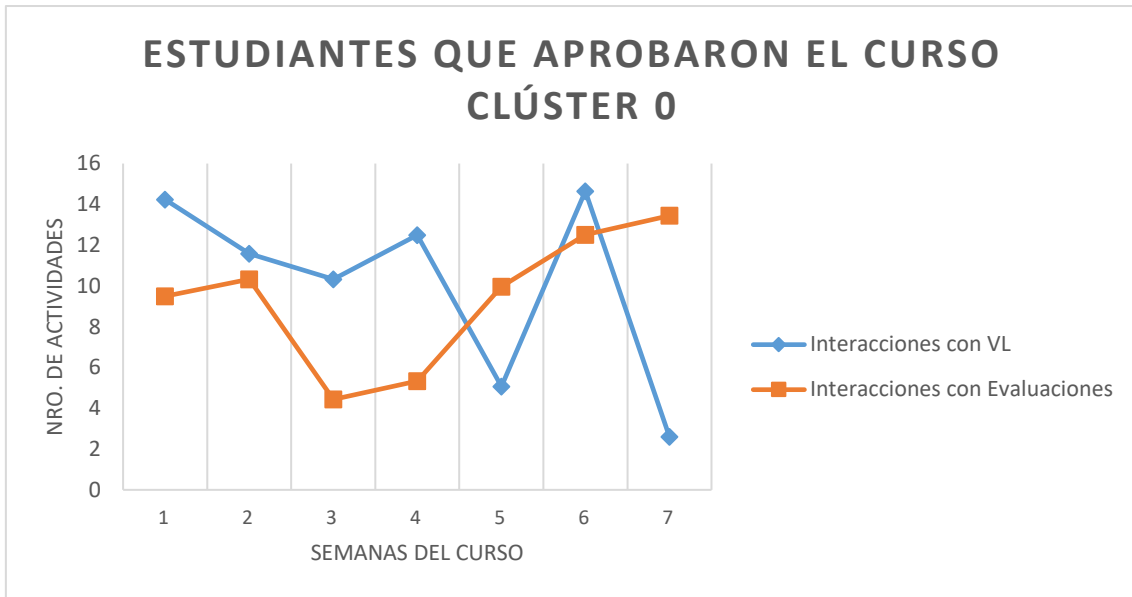


Figura 36: Comportamiento promedio por semana de los estudiantes que aprobaron el curso del Clúster 0

Clúster 1: formado por 653 estudiantes, trabajan en promedio por cada sesión 45.25 minutos dentro del MOOC. Interactúan en promedio con 12.31 videolecturas y 8.30 evaluaciones. Dado que el grupo es determinado por las sesiones como se explicó anteriormente, un mismo estudiante puede estar en el Clúster 0 y el Clúster 1, según su comportamiento por sesión, de esta manera en este grupo forman parte los siete estudiantes que aprobaron el curso y 646 estudiantes no aprobaron.

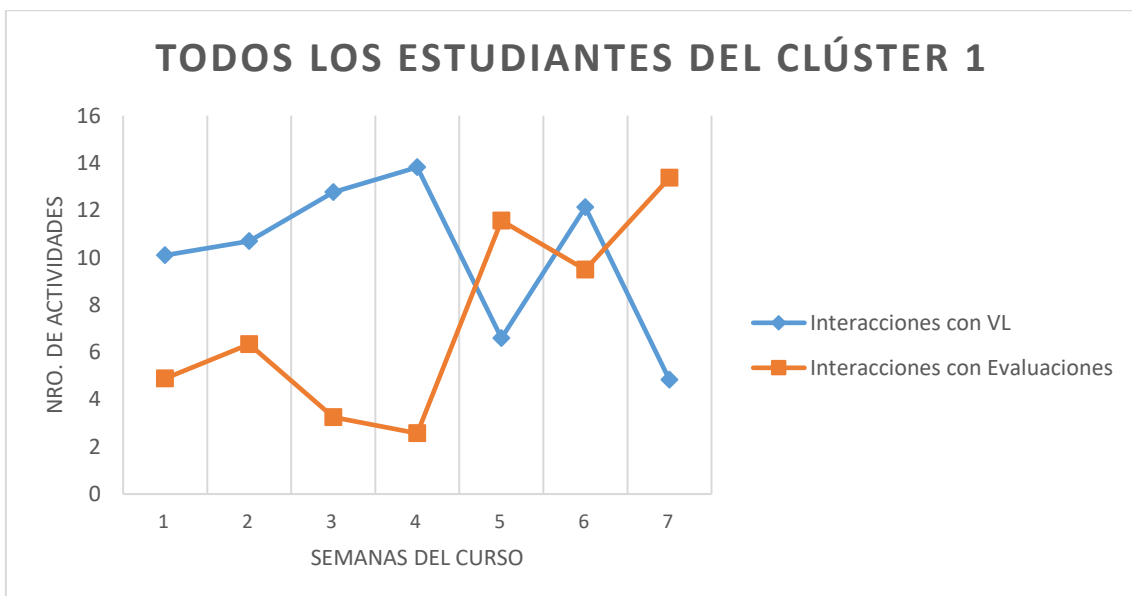


Figura 37: Comportamiento promedio por semana Clúster 1

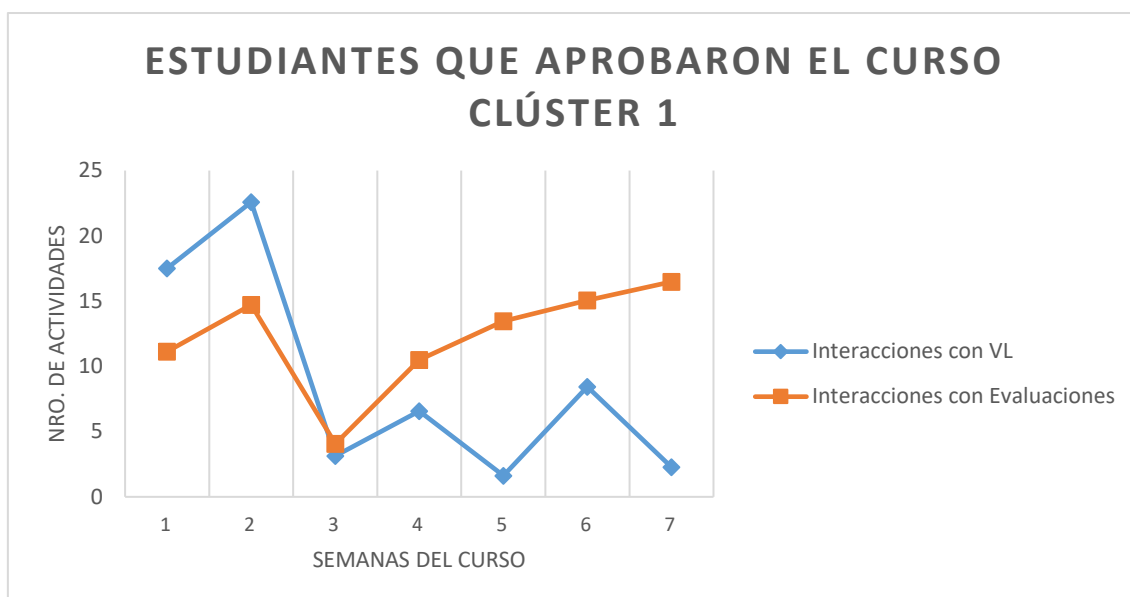


Figura 38: Comportamiento promedio por semana de los estudiantes que aprobaron el curso del Clúster 1

R1. Se generaron dos grupos denominados Clúster 0 y Clúster 1. En la **Tabla 9**, se indica los datos del Clúster 0 y el Clúster 1, los cuales se calcularon para aplicar la prueba estadística “t de Student”, en la cual, se pudo observar diferencias significativas entre cada una de las variables consideradas para el agrupamiento, es decir; que el Clúster 0 y el Clúster 1 son diferentes, y cada uno de ellos es homogéneo.

	Clúster 0	Clúster 1	t	p
Duración por sesión	26.06	36.69	-21.81281432957502	0.001***
Número de interacciones con Video-lecturas por sesión	3.39	10.58	-58.03606809079478	0.001***
Número de interacciones con Evaluaciones por sesión	1.308	14.07	-79.64195423937103	0.001***
Número de interacciones totales por sesión	3.83	14.52	-110.2091638551896	0.001***
Nota: ***p<0.001				

Tabla 9: Prueba t de Student aplicada a los datos del Clúster 0 y el Clúster 1

Luego de revisar las diferencias antes descritas de los clústers mediante el algoritmo de MP de descubrimiento, con la ayuda de la herramienta DISCO, se extrajeron modelos de procesos para cada clúster y contestar la P.I.2. y P.I.3.

La **Figura 39**, presenta una secuencia básica del modelo de procesos que a continuación se explica para su entendimiento. Este modelo es un extracto del modelo general, que se lee de la siguiente manera: (1) Existen 1.185 sesiones que iniciaron, (2) la actividad de Video-lectura inició 9.631 veces, de las cuales mediante la dirección de la flecha (\rightarrow) indica hacia donde se dirige después de dicha actividad e interpretar como transiciones de una actividad hacia otra, (3) Se puede observar en la gráfica, que después de empezar la actividad de Video-lectura los estudiantes tomaron dos opciones; la primera es que 6.953 veces llegaron "Video-lectura fin", y la segunda que 621 veces abandonaron la sesión después de iniciar una Video-lectura. (4) Después de terminar una Video-lectura, los estudiantes tomaron nuevamente dos caminos, el primero 6.029 veces regresó a la Video-Lectura y en segundo lugar 703 veces inició una evaluación (Eval-inicio), en la cual se realizaron 930 veces esta actividad. Finalmente (5) 346 veces la dirección de la flecha (\rightarrow) marca el curso hace la actividad "Eval-fin", la cual se realizó 485 veces y 385 veces regresaron a Iniciar una Video-lectura. En la gráfica además se puede observar que mientras más veces se tome un camino o dirección la flecha (\rightarrow) esta es más gruesa, y que los colores en las cajas que indican la actividad, son más azules a mayor número de repeticiones y más tenue a menor número.

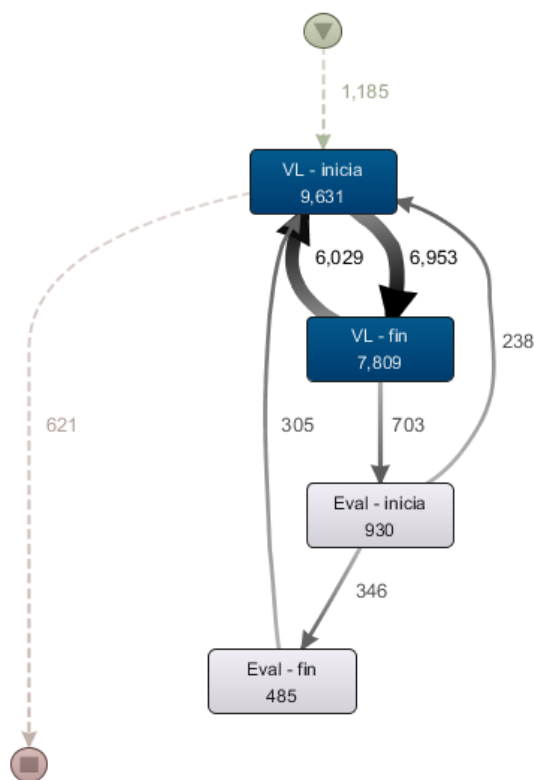


Figura 39: Ejemplo de interpretación de un modelo de procesos



P.1.2 ¿Qué secuencias de aprendizaje realizan los estudiantes en las diferentes sesiones? Para dar respuesta a esta pregunta, el enfoque analítico utilizado fue: (a) analizar el modelo de procesos de los estudiantes del Clúster 0 y Clúster 1 a nivel micro; (b) analizar el modelo de procesos de los estudiantes del Clúster 0 y Clúster 1 a nivel macro.

Nivel Micro

R.2. Los estudiantes del Clúster 0 y del Clúster 1, tienden a seguir la misma secuencia de actividades. Se identifica un comportamiento similar entre los dos grupos tanto el Clúster 0 como el Clúster 1, en la que las actividades siguen un orden secuencial e idéntico como se puede observar en las **Figura 40** y **Figura 41**. Los itinerarios que más se repiten de ambos grupos son (en orden de mayor a menor frecuencia):

1. VL-inicia → VL-fin → VL-inicia
2. VL-inicia → VL-fin → Eval-inicia → Eval-inicia nuevamente → Eval-fin

R.3. Los estudiantes del Clúster 0 y del Clúster 1, en las sesiones de trabajo, tienden a iniciar una evaluación, repetirla hasta lograr aprobarla, sin volver a realizar video-lecturas intermedias. Esto se puede apreciar en el itinerario número 2 antes descrito en R.1. y en la **Figura 40** y **Figura 41** enmarcado en color rojo, además, se puede observar que las sesiones que realizaron los estudiantes tanto del Clúster 0 como del Clúster 1, repitieron un gran número de veces “Eval-inicia nuevamente” con 10.971 y 2997 veces respectivamente cada clúster.

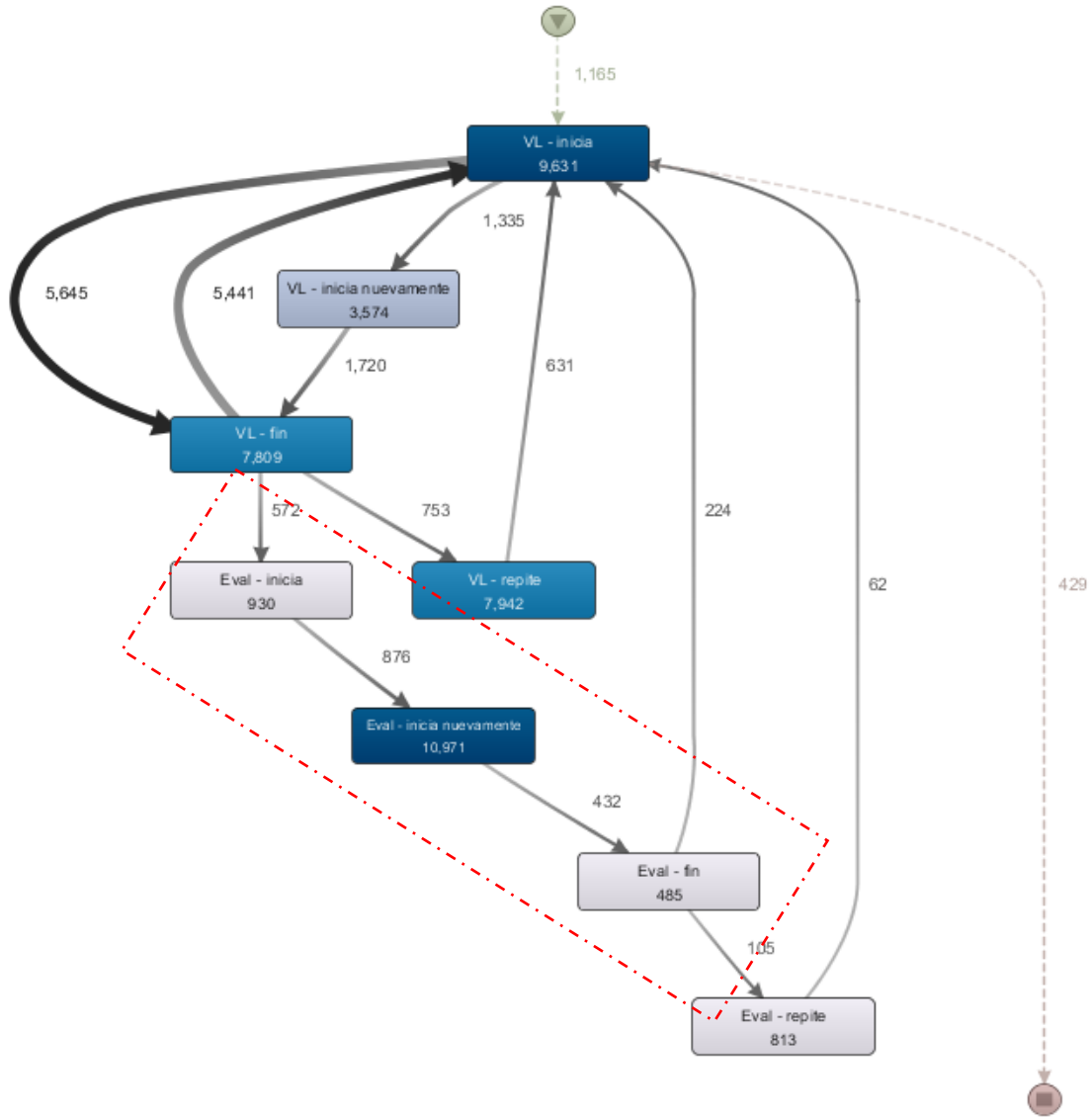


Figura 40: Clúster_0, modelo de procesos a nivel micro

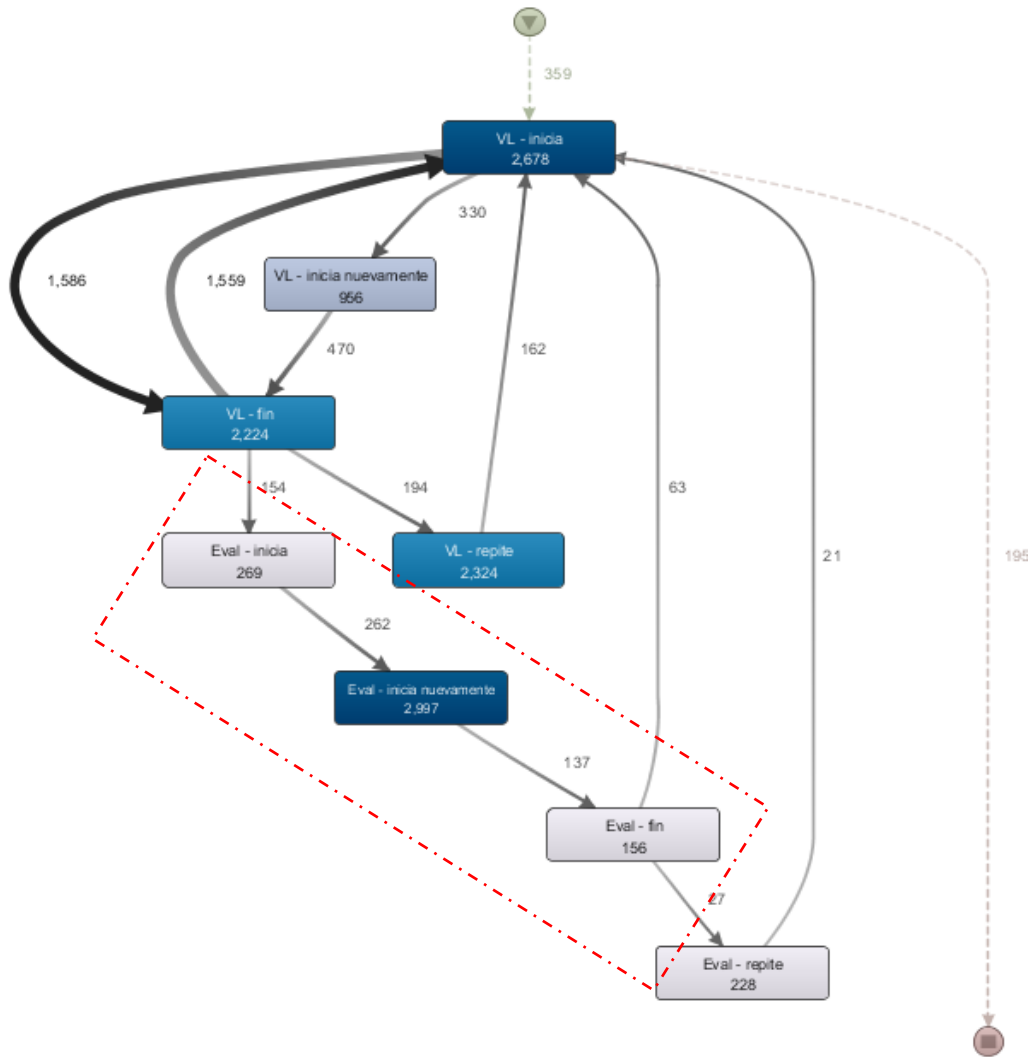


Figura 41: Clúster_1, modelo de procesos a nivel micro

Nivel Macro

R.4. Las sesiones de los estudiantes del Clúster 0 y del Clúster 1, durante las dos primeras semanas en el MOOC, siguen el curso de forma secuencial, pero además repiten varias veces la primera video-lectura de los respectivos módulos y regresan a la primera actividad (video-lectura) después de realizar la evaluación del módulo. Este comportamiento se observó en las figuras de los clústers: Clúster 0 y Clúster 1 de la primera semana (**Figura 42 (a)**) y (**Figura 42 (b)**), y en la segunda semana (**Figura 43 (a)**) y (**Figura 43 (b)**), que corresponden al clúster 0 y 1 respectivamente, donde se enmarcó en color rojo las iteraciones de la primera actividad (video-lectura) y en color azul cuando posterior a las actividades de evaluación regresan a la primera video lectura de cada módulo.

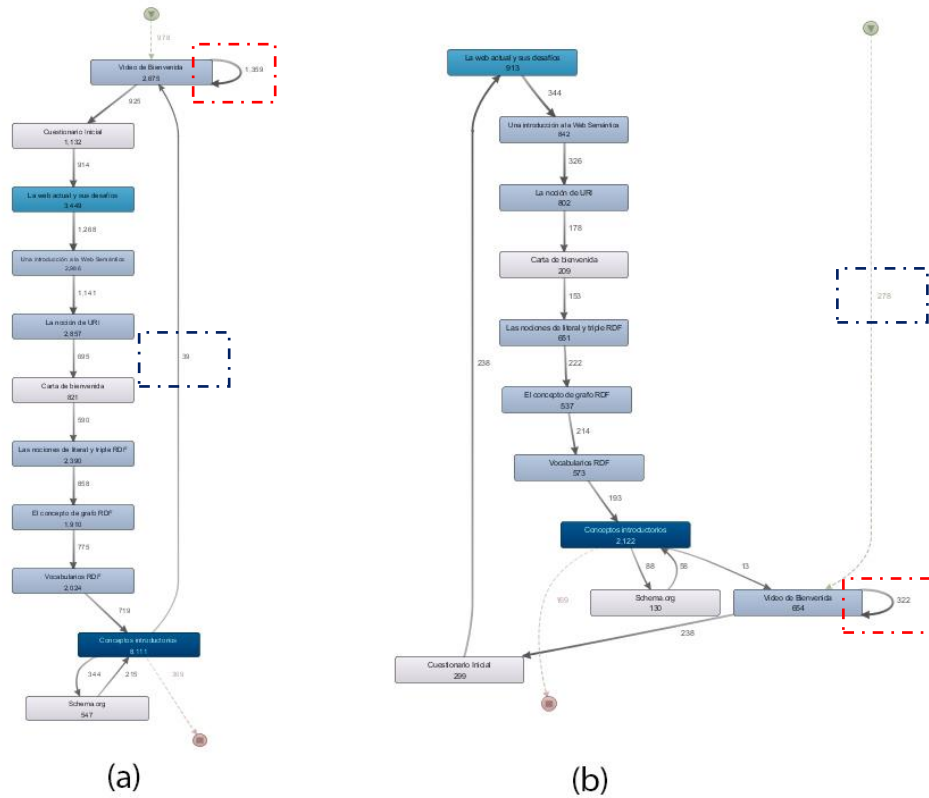


Figura 42: Modelo de procesos en la primera semana del MOOC

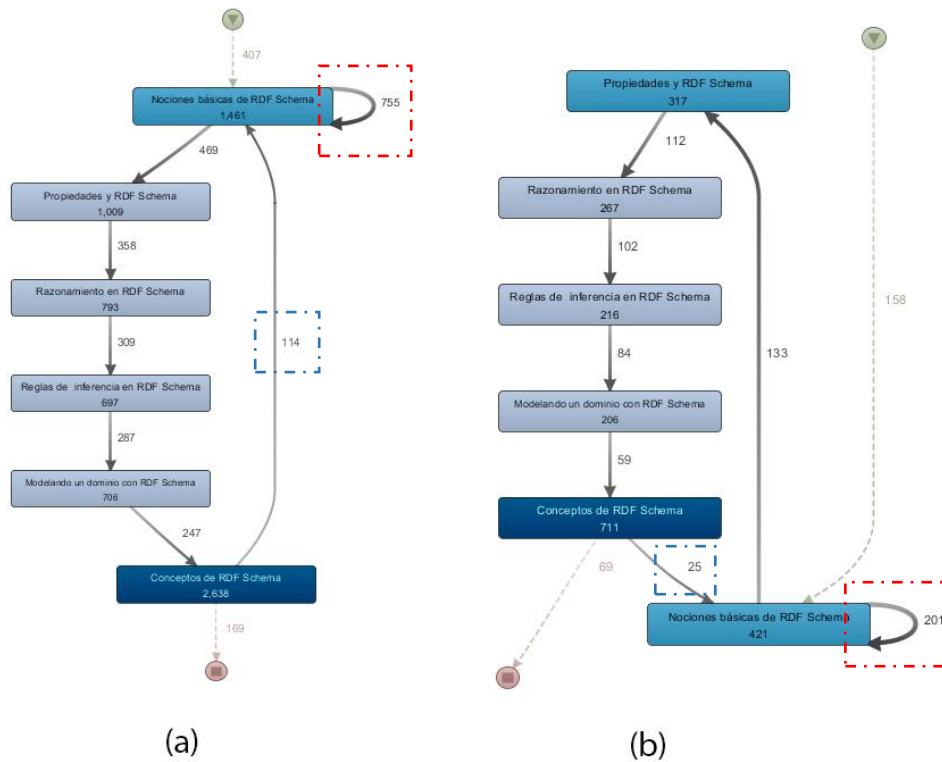


Figura 43: Modelo de procesos, segunda semana del MOOC

R.5. Las sesiones de los estudiantes del Clúster 0 y del Clúster 1 durante la tercera semana, tienden seguir en secuencia lineal el curso, pero además regresan al inicio de cada módulo después de completar la evaluación del módulo y la evaluación por pares. Esto se puede observar en la *Figura 44 (a)* y *Figura 44 (b)*, enmarcado en color verde.

R.6. Las sesiones de los estudiantes en el Clúster 1 en la tercera semana del MOOC, repite de forma iterativa en la primera actividad del módulo (video-lectura) en la tercera semana del curso. Esto se puede observar en la *Figura 44 (b)* enmarcado en color rojo, donde además se indica que el Clúster 0 (*Figura 44 (a)*), no tiene esa secuencia repetitiva.

R.7. Algunas sesiones (29 sesiones) del Clúster 1 de la tercera semana del MOOC abandonaron la sesión de estudio en la tercera semana después de la primera actividad. Esto se puede observar en la *Figura 44 (b)*, enmarcado en color azul.

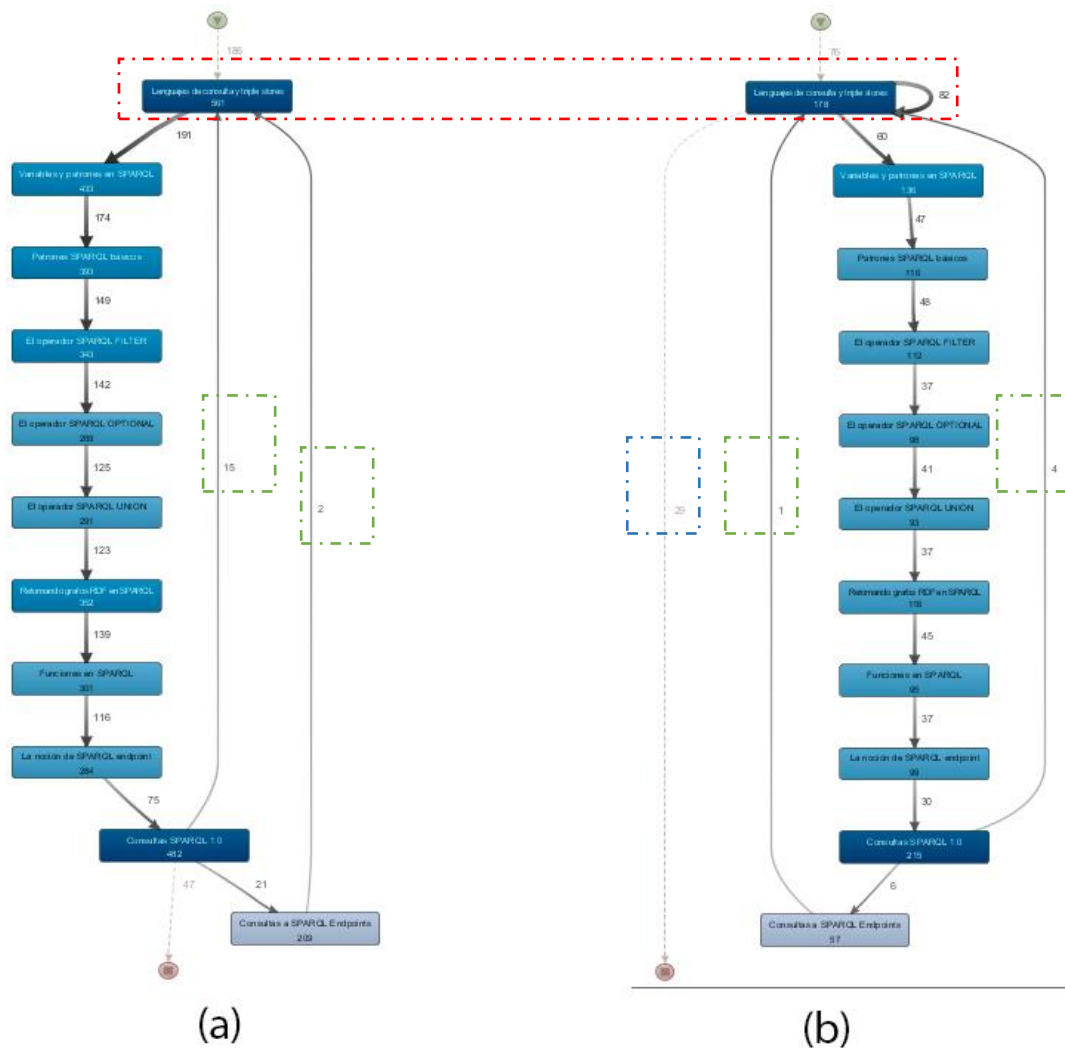


Figura 44: Modelo de procesos, tercera semana del MOOC

R.8. Las sesiones de los estudiantes del Clúster 0 y Clúster 1 en la cuarta semana, siguen un orden de secuencia lineal planteado por el curso, sin embargo, después de la evaluación modular regresan a una actividad de video-lectura. En cuanto al Clúster 0 (**Figura 45 (a)**) los estudiantes regresan a la primera actividad del módulo que es una video-lectura (enmarcado en rojo), mientras que el Clúster 1 (**Figura 45 (b)**), regresa a una actividad intermedia, pero que también es una video lectura (enmarcado en rojo). Además, el Clúster 1 (**Figura 45 (b)**), repite iterativamente la primera actividad del módulo (enmarcado en azul).

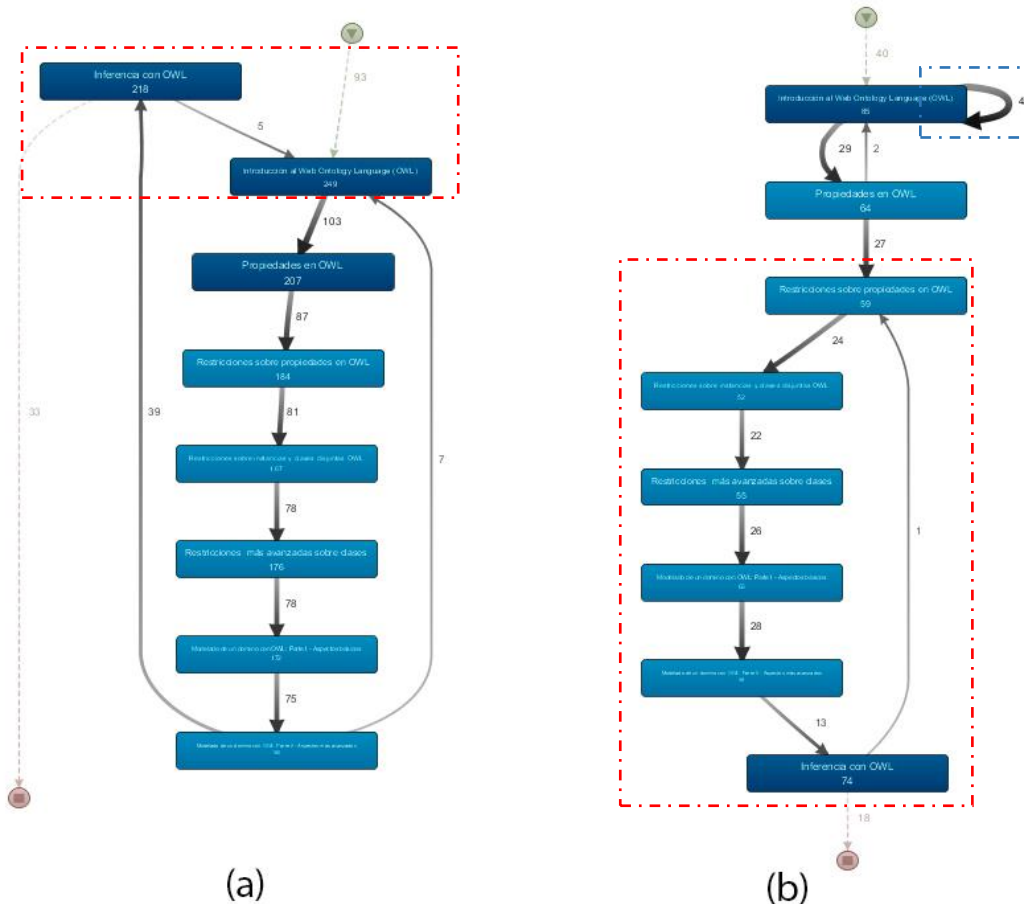


Figura 45: Modelo de procesos, cuarta semana del MOOC

R.9. Las sesiones de los estudiantes en la quinta semana del MOOC, tienden a ser de mayor navegación hacia actividades anteriores tanto el Clúster 0 como el Clúster 1. Esto se puede observar en la **Figura 46 (a y b)**, donde la secuencia de aprendizaje si bien es cierto sigue el orden del curso, también regresan hacia actividades previas.

R.10. Las sesiones de los estudiantes en la quinta semana del MOOC, para el Clúster 0, tiende a ser iterativo en actividades de video-lectura, mientras que las sesiones de los estudiantes en el Clúster 1, tienden a ser iterativas

en la evaluación modular. Esto se puede observar en la **Figura 46 (a)** del Clúster 0, que se encuentra enmarcado en azul, y repiten actividades de tipo video-lecturas. Por otra parte, el Clúster uno realiza ese comportamiento, pero en la evaluación modular que se puede observar en la **Figura 46 (b)**, enmarcado en color azul.

R.11. Las sesiones de los estudiantes en la quinta semana del MOOC, para el Clúster 0, tienden a terminar la sesión de trabajo después de una video lectura, mientras que las sesiones de los estudiantes del Clúster 1, tienden a terminar la sesión de trabajo del módulo después de la evaluación. Se puede observar este comportamiento enmarcado en color rojo, **Figura 46 (a)** para el Clúster 0 y **Figura 46 (b)** para el Clúster 1.

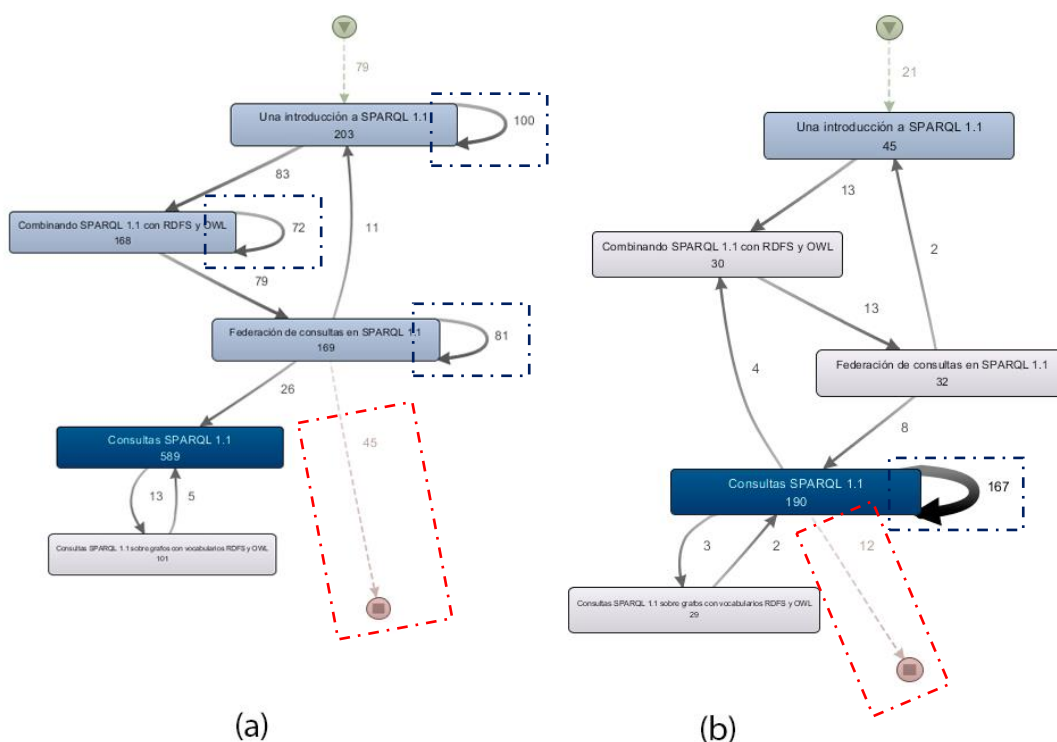


Figura 46: Modelo de procesos, quinta semana del MOOC

R.12. Las sesiones de los estudiantes en la sexta y séptima semana del MOOC, para el Clúster 0 y el Clúster 1, tienden a seguir las mismas secuencias, ya que a este nivel deben llegar las personas que culminaron el curso. Esto se puede observar en las **Figuras 47 y 48**, donde existen ligeras diferencias, como por ejemplo en el Clúster 0 de la sexta semana **Figuras 487 (a)** los estudiantes terminan su sesión de trabajo después de una actividad de video-lectura (enmarcado en color azul), mientras que, en la misma semana, pero el Clúster 1 **Figura 47 (b)**, los estudiantes terminan su sesión de trabajo en el módulo después de una actividad de evaluación (enmarcado en color azul).

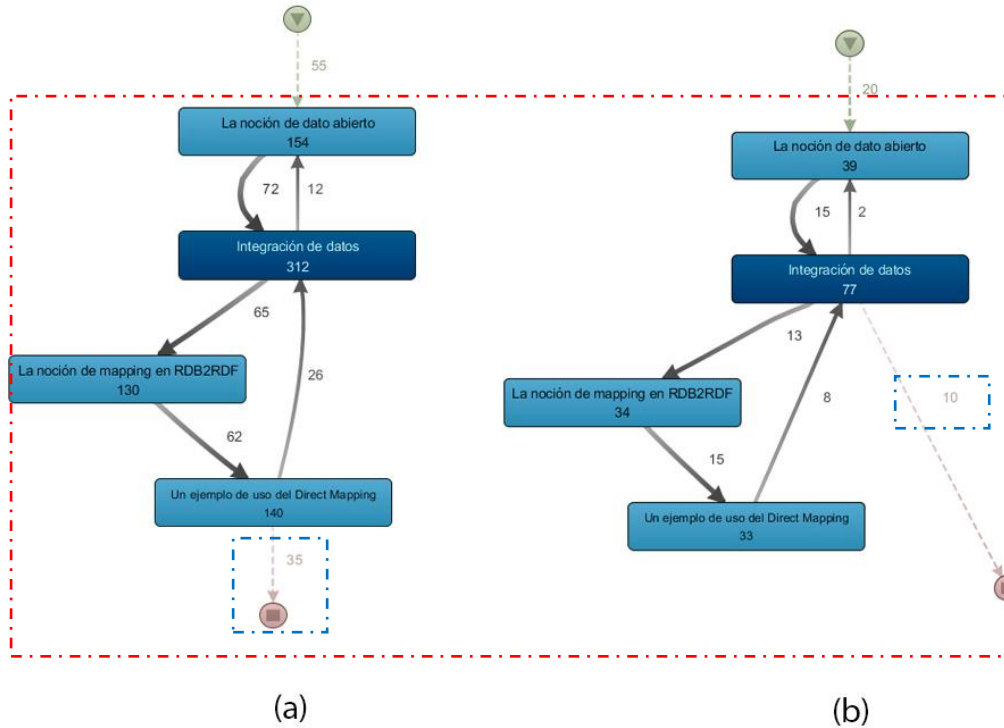


Figura 47: Modelo de procesos, sexta semana del MOOC

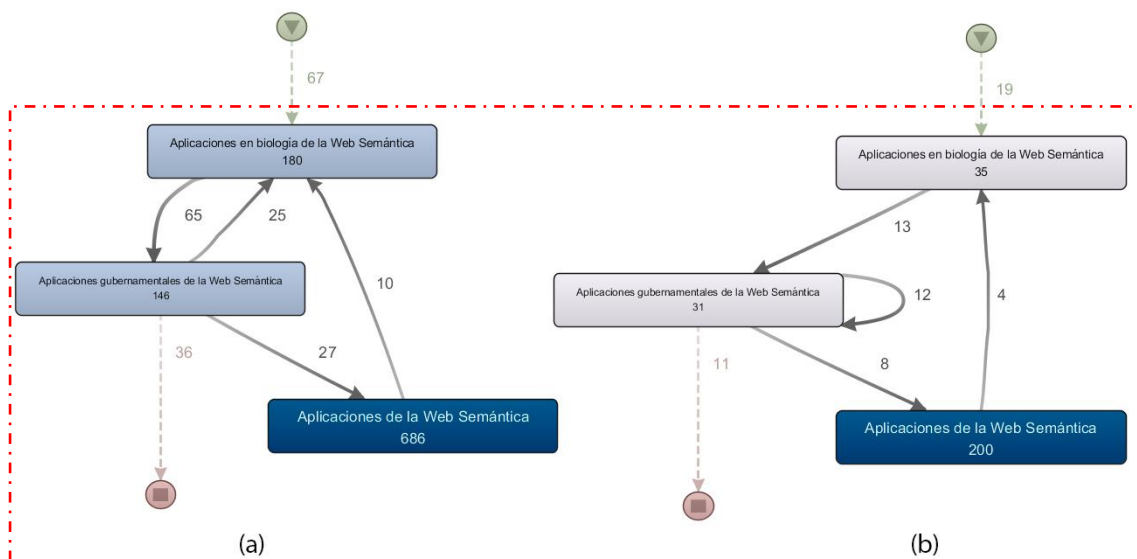


Figura 48: Modelo de procesos, séptima semana del MOOC

P.I.3 ¿Qué diferencia existe entre las sesiones de los estudiantes que completan el curso y de los que no? Para responder esta pregunta, los estudiantes fueron clasificados por la plataforma Coursera con los valores (0) No Aprueba y (1) Aprueba. En base a estas pautas, se obtuvieron los siguientes resultados:



Nivel Micro

R.13. Los estudiantes que aprobaron el curso, a nivel micro tienden a intentar finalizar una evaluación (Eval-inicia nuevamente) de forma consecutiva, mientras que los estudiantes que no aprueban el curso siguen una secuencia lineal sin repetir de forma consecutiva el intento de finalizar. Se puede observar en las **Figuras 49** y **50**, la similitud a nivel micro de los estudiantes que aprueban y los que no aprueban el curso, en el que después de realizar las actividades de Video-Lectura los estudiantes proceden a realizar los exámenes y lo repiten hasta lograr conseguirlo (enmarcado en color rojo).

R.14. Los estudiantes que no aprueban el curso, tienden a abandonarlo después de realizar una actividad de video-lectura. En este caso, se observa que los estudiantes que no aprobaron el curso, tienden más a la deserción (enmarcado en color azul), este comportamiento se puede apreciar en la **Figura 50**, que luego de terminar una actividad de video-lectura culmina el curso, mientras que en la **Figura 49**, los estudiantes que aprueban el curso, terminan sus sesiones de trabajo después de finalizar actividades de evaluación.

Nivel Macro

R.15. Los estudiantes que aprueban el curso, tienden a ser más ordenados en la secuencia de actividades al momento de realizarlas. Por otro lado, los estudiantes que no aprueban el curso, empiezan con un comportamiento ordenado similar al de los estudiantes que aprueban el curso, pero en ciertas etapas del curso se desordenan. Además, se observa que abandonan el curso en etapas tempranas. En la **Figura 51**, se muestra el modelo de procesos de los estudiantes que aprobaron el curso (b) como de los que no aprobaron (a), el comportamiento similar está enmarcado en color rojo, y cuando empiezan a desordenarse del grupo que no aprobaron (a), se puede observar enmarcado en color azul.

Análisis de comportamiento de las sesiones por semana

R.16. Los estudiantes que aprueban el MOOC, realizan más sesiones en promedio respecto a los estudiantes que no aprueban el MOOC. En la **figura 52**, se puede apreciar, que los estudiantes que aprueban el curso realizan más sesiones respecto a los que no aprueban el curso, con 34.14 y 17.32 sesiones promedio respectivamente. Además, se observa que, en cuanto al promedio semanal de sesiones, son más estables y homogéneos, que fluctúan entre 31.43 y 36.71 sesiones por semana. Por otro lado, los estudiantes que no aprueban el curso, van de menos a más en cuanto al promedio de número de sesiones promedio por semana, estos fluctúan entre 8.34 y 22.31 sesiones por semana.

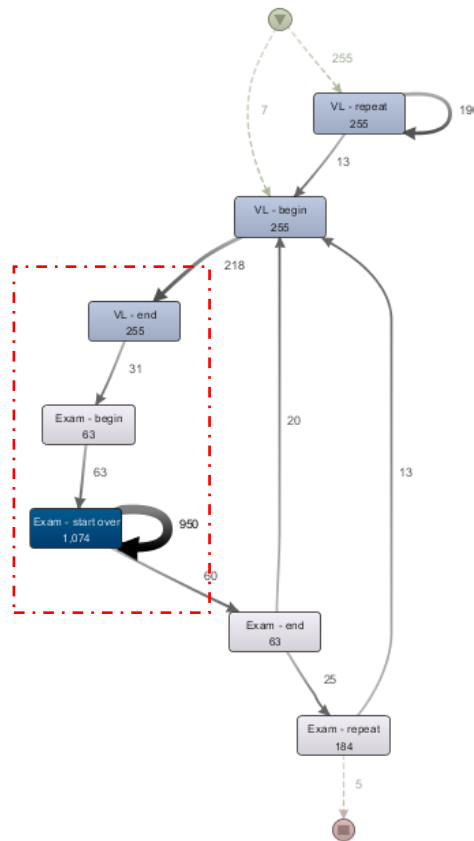


Figura 49: Modelo de proceso, nivel micro de los estudiantes que aprueban el curso

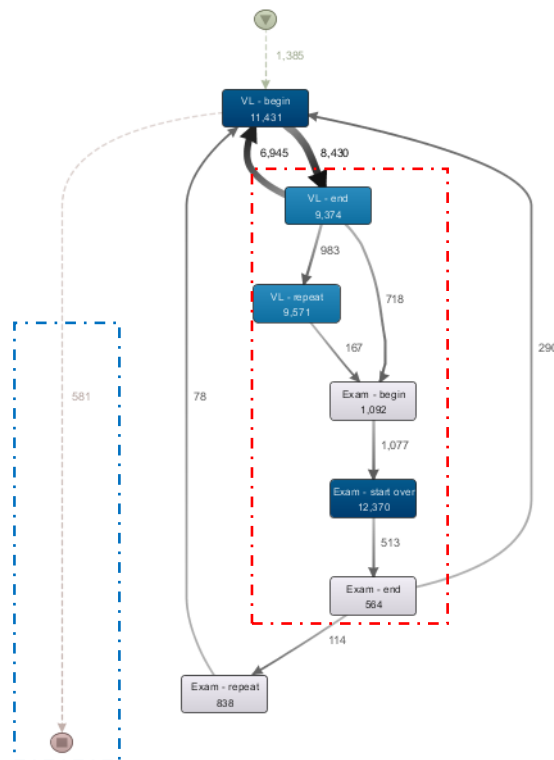


Figura 50: Modelo de proceso, nivel micro de los estudiantes que no aprueban el curso

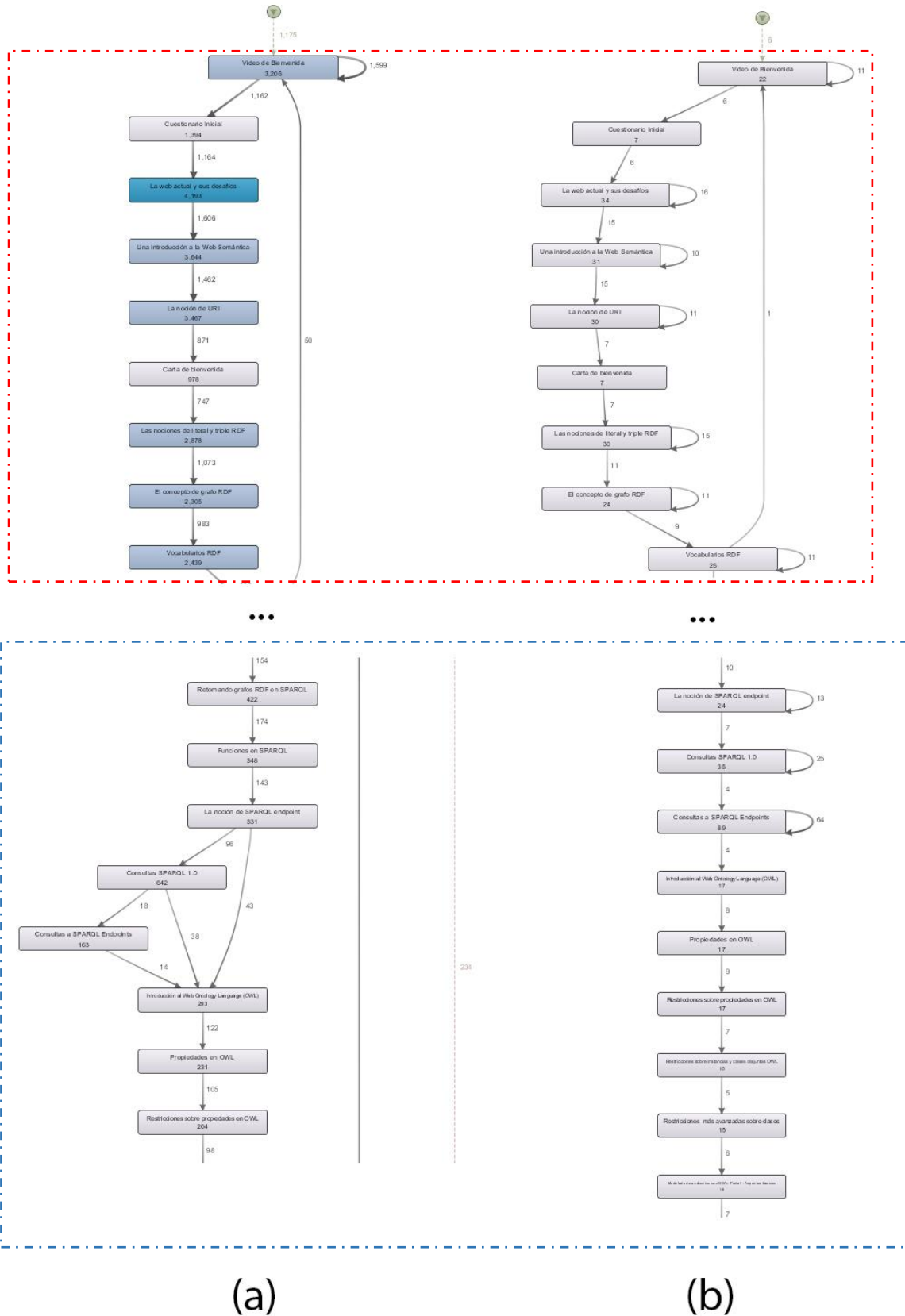


Figura 51: Comportamiento de los que aprueban vs no aprueban

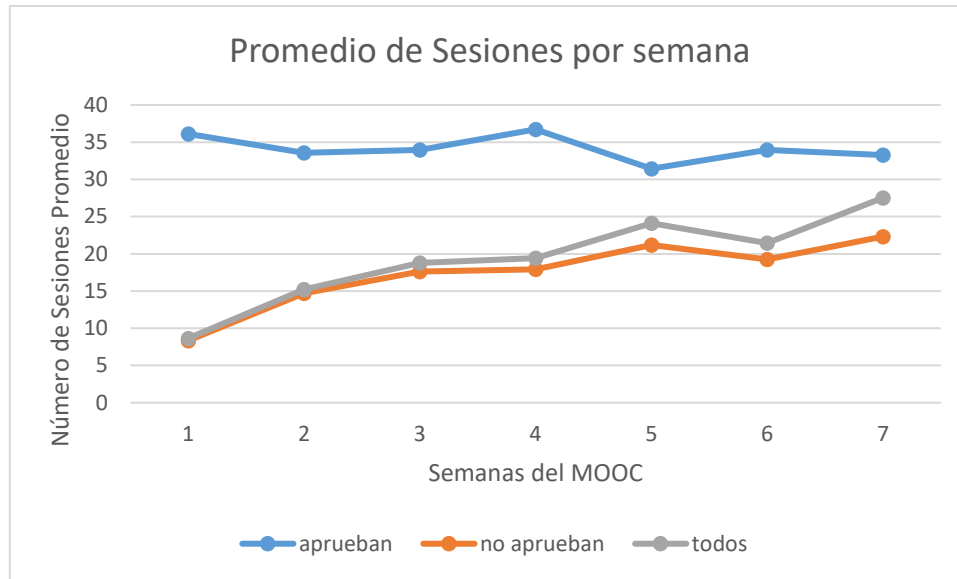


Figura 52: Promedio de sesiones realizadas por semana de los estudiantes del MOOC

5.5. Recapitulación

En este capítulo se presentó el desarrollo de un caso de estudio para dar respuesta a las tres preguntas de investigación planteadas. Para esto se siguió las fases de la metodología PM2, mediante la cual se definió los objetivos, se detalló la extracción y limpieza de los datos, como se generaron los logs de eventos y finalmente se presentó los modelos de procesos para su análisis.

Para el caso de la P.I.1. que consistía en ¿Qué tipo de sesiones realizan los estudiantes en un MOOC?, donde se obtuvieron dos grupos de estudiantes basados en sus sesiones de estudio, los cuales se formaron de la siguiente manera: (1) Clúster 0: formado por 1.292 estudiantes, trabajan en promedio por cada sesión 29.7 minutos en el MOOC. Además, interactúan en promedio con 3.53 video-lecturas y 0.7 evaluaciones y, (2) Clúster 1: formado por 653 estudiantes, trabajan en promedio por cada sesión 45.49 minutos en el MOOC. Además, interactúan en promedio con 12.31 video-lecturas y 8.30 evaluaciones.

Para el caso de la P.I.2. que consistía en ¿Qué secuencias de aprendizaje realizan los estudiantes en las diferentes sesiones?

A nivel macro, se pudo observar que el Clúster 0 realizan secuencias a nivel macro por cada semana, donde en la primera y segunda semana se realizaron las actividades en el orden secuencial del curso, donde la primera actividad (video-lectura) es repetida iterativamente, la diferencia entre la primera y la segunda semana, se da luego de la última actividad hubo sesiones de estudiantes en las que regresaron a revisar nuevamente la primera actividad del módulo, en la segunda semana. En la tercera y cuarta semana, se observa un comportamiento que sigue siendo secuencial, en relaciones a las lecciones del curso, pero empiezan los estudiantes a regresar a la primera actividad del



módulo respectivo, desde otras actividades. Finalmente, en las últimas tres semanas se empieza a observar iteración de actividades no sólo de la primera actividad del módulo sino de algunas actividades diferentes, además se empieza a evidenciar un cambio en la secuencia de las actividades, donde se empieza a navegar de forma más desordenada las actividades del curso, avanzando y retrocediendo entre actividades, por ejemplo; los estudiantes están regresando desde una actividad de evaluación hacia una video-lectura, o de una video lectura regresa a una video lectura anterior, correspondientes a lecciones del módulo o semana del curso.

A nivel micro, en las que el orden de secuencia es: video-lectura inicio a video-lectura fin, con una bifurcación de video-lectura inicia nuevamente, ya que en algunos casos no terminan de ver la video-lectura y entonces se genera este estado (VL- inicia nuevamente), diferente de Video-lectura repite, ya que ese estado se da solo cuando se terminó la video-lectura. Al finalizar las actividades de Video-lectura, empiezan las evaluaciones en las que inician y las repiten hasta aprobar. Se pudo observar que los estudiantes que empezaban las evaluaciones, no la aprobaron directamente en ningún caso, por lo que intentaron varias veces aprobar la misma, y fueron pocos los que lo hicieron y muchos menos los que repitieron la evaluación, probablemente para mejorar la calificación.

Para el Clúster 1 realizan secuencias a nivel macro por cada semana, donde en la primera y segunda semana se realizaron las actividades en el orden secuencial del curso, pero solo en la segunda semana hubo iteraciones en la primera actividad de dicho módulo, además en la segunda semana se regresó de la última actividad hacia la primera. En la tercera y cuarta semana, se observa un comportamiento que sigue siendo secuencial, con repeticiones iterativas en la primera actividad, donde ese sigue un orden establecido por el curso para completar las actividades, con algunos retornos hacia las primeras actividades del módulo desde las últimas actividades del mismo. Además, en la tercera semana existen sesiones (26) en las que solo se completó la primera actividad del módulo y ya no se realizaron más en el módulo. Finalmente, en las últimas tres semanas se empieza a observar un cambio en la secuencia de las actividades, donde se observa algunas iteraciones en actividades de orden intermedio de cada módulo, además de navegación más retroalimentada, es decir; al pasar cierta actividad se retorna a otra diferente, donde se intuye que la razón es encontrar una retroalimentación y fortalecimiento en ciertos aspectos. Por otro lado, a nivel micro, el comportamiento es completamente igual al del Clúster 0.

Para el caso de la P.I.3. que consistía en ¿Qué diferencia existe entre las sesiones de los estudiantes que completan el curso y de los que no?, se pudo observar el comportamiento a nivel micro similar, en cuanto a la secuencia, pero se pudo observar que para los estudiantes que aprobaron el MOOC, la actividad de “Eval-inicia nuevamente”, se repitió consecutivamente. Por otro lado, los estudiantes hacían un recorrido más lineal. Además, otra diferencia que se



observó, fue que los estudiantes que aprobaron el curso, terminaban su sesión de trabajo después de una evaluación, mientras que los estudiantes que no aprobaron lo hacían desde una actividad diferente (video-lectura). A nivel macro, se evidenció que ambos grupos, es decir; los que aprueban y los que no aprueban, en los primeros módulos o semanas del curso ambos son ordenados y avanzan de forma secuencial y los estudiantes que aprueban el curso mantienen ese comportamiento hasta el final, mientras que, los estudiantes que no aprueban empiezan a tener cierto desorden desde los módulos más avanzado y deserción, es decir abandonar el MOOC.

Por otro lado, para la P.I.3., además, se observó que los estudiantes que aprueban el curso realizan más sesiones promedio por semana (34.14 sesiones) respecto a los que no aprueban el curso (17.32 sesiones). Además, los estudiantes que aprueban el curso, presentan una regularidad en cuanto al número de sesiones por semana, es decir sin muchas fluctuaciones, mientras que los estudiantes que no aprueban el curso presentan un comportamiento que va de menos a más, es decir; empiezan con pocas sesiones por semana y este valor va aumentando según aumentan las semanas.

6. Conclusiones y trabajos futuros

En este capítulo se presentan las principales conclusiones de esta tesis y el trabajo futuro.

6.1. Conclusiones

La motivación primaria de esta tesis ha sido la de explorar nuevas oportunidades y retos que la analítica de aprendizaje ofrece para estudiar el comportamiento de los estudiantes en cursos MOOC.

Específicamente tres preguntas de investigación han sido resueltas con esta tesis:

- P.I.1 ¿Qué tipo de sesiones realizan los estudiantes en un MOOC?
- P.I.2 ¿Qué secuencias de aprendizaje realizan los estudiantes en las diferentes sesiones?
- P.I.3 ¿Qué diferencia existe entre las sesiones de los estudiantes que completan el curso y de los que no?

De los resultados obtenidos y en base a las preguntas de investigación planeadas en el capítulo uno, se puede concluir:

1.- Respecto a la P.I.1.:

- R.1. Se generaron dos grupos denominados Clúster 0 y Clúster 1.
 - Al considerar las variables de comportamiento por sesión, el aporte de esta tesis, fue demostrar que variables como la duración de sesión, número de interacciones con video-lecturas por sesión, número de interacciones con evaluaciones por sesión y número de interacciones totales por sesión, muestran diferencias estadísticas significativas, con las cuales es posible obtener grupos más homogéneos.

2.- Respecto a la P.I.2., se obtuvo los siguientes resultados:

- R.2. Los estudiantes del Clúster 0 y del Clúster 1, tienden a seguir la misma secuencia de actividades a nivel micro.
 - A nivel micro, es exactamente el mismo modelo para ambos grupos, tanto el Clúster 0 y el Clúster 1, se puede concluir que el modelo es similar ya que de los 1.500 estudiantes que iniciaron el curso, solo lo terminaron y aprobaron 7 estudiantes, por ende, el comportamiento de



la mayoría se ve reflejado en la similitud de los mismos. Además, cabe resaltar que un mismo estudiante según su sesión de estudio podía pertenecer al Clúster 0 o al Clúster 1.

- R.3. Los estudiantes del Clúster 0 y del Clúster 1, en las sesiones de trabajo, tienden a iniciar una evaluación, repetirla hasta lograr aprobarla, sin volver a realizar video-lecturas intermedias.
 - Este comportamiento, demuestra que los estudiantes no se retroalimentaban con sesiones de video-lecturas, lo que los llevaba a intentar sucesivamente aprobar una evaluación y de ese modo avanzar en el MOOC.
- R.4. Las sesiones de los estudiantes del Clúster 0 y del Clúster 1, durante las dos primeras semanas en el MOOC, siguen el curso de forma secuencial, pero además repiten varias veces la primera video-lectura de los respectivos módulos y regresan a la primera actividad (video-lectura) después de realizar la evaluación del módulo
 - En las dos primeras semanas se puede apreciar mayor flujo de trabajo, en cuanto al número de actividades e interacciones realizadas, además de seguir un orden dictado por el curso, además la similitud con los estudiantes con estilo de aprendizaje reflexivo del estudio de Maldonado y demás (2016), en el que los estudiantes regresan al inicio del módulo después de realizar las evaluaciones.
- R.5. Las sesiones de los estudiantes del Clúster 0 y del Clúster 1 durante la tercera semana, tienden seguir en secuencia lineal el curso, pero además regresan al inicio de cada módulo después de completar la evaluación del módulo y la evaluación por pares.
- R.6. Las sesiones de los estudiantes en el Clúster 1 en la tercera semana del MOOC, repite de forma iterativa en la primera actividad del módulo (video-lectura) en la tercera semana del curso.
 - En el Clúster 1, empiezan a existir actividades repetitivas que puede deberse a dos alternativas: (1) Refuerzo teórico o (2) Poco entendimiento de esa actividad por parte del alumno, debido al mal diseño del curso.
- R.7. Algunas sesiones (29 sesiones) del clúster 1 en la tercera semana del MOOC abandonaron la sesión de estudio en la tercera semana después de la primera actividad.



- Desde la semana 3 se empieza a observar que estudiantes del Clúster 1 no terminan las actividades del módulo, se puede deber a un abandono del curso o al poco interés en el módulo.
- R.8. Las sesiones de los estudiantes del Clúster 0 y Clúster 1 en la cuarta semana, siguen un orden de secuencia lineal planteado por el curso, sin embargo, después de la evaluación modular regresan a una actividad de video-lectura.
- R.9. Las sesiones de los estudiantes en la quinta semana del MOOC, tienden a ser de mayor navegación hacia actividades anteriores tanto el Clúster 0 como el Clúster 1.
 - Los estudiantes empiezan a tener una retroalimentación más evidente, como se mencionó anteriormente se puede deber a un refuerzo teórico o a un poco entendimiento del mismo, en el que hace falta revisar varias veces el contenido por su nivel de dificultad.
- R.10. Las sesiones de los estudiantes en la quinta semana del MOOC, para el Clúster 0, tiende a ser iterativo en actividades de video-lectura, mientras que las sesiones de los estudiantes en el Clúster 1, tienden a ser iterativas en la evaluación modular.
- R.11. Las sesiones de los estudiantes en la quinta semana del MOOC, para el Clúster 0, tienden a terminar la sesión de trabajo después de una video lectura, mientras que las sesiones de los estudiantes del Clúster 1, tienden a terminar la sesión de trabajo del módulo después de la evaluación.
 - En la quinta semana, según la R.10 y R.11, se intuye que la perseverancia de los estudiantes del Clúster 1, fue compensada, ya que en dicho grupo después de finalizar la evaluación, mientras que los estudiantes del Clúster 0, repiten las actividades de video-lectura, y terminan la sesión de trabajo después de la misma.
- R.12. Las sesiones de los estudiantes en la sexta y séptima semana del MOOC, para el Clúster 0 y el Clúster 1, tienden a seguir las mismas secuencias, ya que a este nivel deben llegar las personas que culminaron el curso.
 - La similitud en estos modelos, radica en el que seguramente solo lo realizaron los estudiantes que culminaron el curso, además se puede observar por el color de las cajas de actividades (tenues) que fueron pocas las sesiones que se trabajaron en esta semana.



- Existe por ende un abandono en etapas previas a las dos últimas semanas considerable, para finalmente concluir el curso solo siete estudiantes.
- Estos resultados eran esperados, ya que el índice de aprobación de un MOOC bordea el 5% según Kizilcec (2017), en este caso particular el índice de aprobación fue de 0.46%, por ello los últimos módulos tuvieron pocas interacciones con las actividades del curso.

3.- Respecto a la P.I.3., se obtuvo los siguientes resultados:

- R.13. Los estudiantes que aprobaron el curso, a nivel micro tienden a intentar finalizar una evaluación (Eval-inicia nuevamente) de forma consecutiva, mientras que los estudiantes que no aprueban el curso siguen una secuencia lineal sin repetir de forma consecutiva el intento de finalizar.
- R.14. Los estudiantes que no aprueban el curso, tienden a abandonarlo después de realizar una actividad de video-lectura.
 - Un aporte acerca de los resultados en R.13 y R.14, es que los grupos identificados como los estudiantes que aprueban el curso y los que no lo hacen, se observa un comportamiento similar a nivel de actividades, en donde varía en la deserción que a diferencia de los estudiantes que aprueban el curso se observa que después de las evaluaciones o actividades de examen se finaliza el curso, mientras que de los estudiantes que no aprueban el curso se observa que se terminan después de una video-lectura, es decir; abandonaron el curso en una actividad y no la retomaron.
- R.15. Los estudiantes que aprueban el curso, tienden a ser más ordenados en la secuencia de actividades al momento de realizarlas. Por otro lado, los estudiantes que no aprueban el curso, empiezan con un comportamiento ordenado similar al de los estudiantes que aprueban el curso, pero en ciertas etapas del curso se desordenan. Además, se observa que abandonan el curso en etapas tempranas.
 - Este estudio tiene un sesgo por la cantidad de estudiantes que reprobaron el MOOC, con una tasa de no aprobación de 99.64%. Cabe resaltar que las secuencias de aprendizaje fueron similares en gran parte del curso, la respuesta a esto es esperada, ya que, según los datos demográficos, según la muestra de la distribución académica de los participantes **Tabla 4**, se observa que el 82.14% de estudiantes del curso, tenían una instrucción académica entre Universidad y Postgrado.



- En base a esto es posible concluir que los estudiantes que no aprobaron el curso, no lo es debido a la falta de conocimientos previos.
- R.16. Los estudiantes que aprueban el MOOC, realizan más sesiones en promedio respecto a los estudiantes que no aprueban el MOOC.
 - Se puede concluir que los estudiantes que aprobaron el curso trabajan más intensamente y de forma constante a lo largo de las siete semanas de duración del MOOC.
 - Otra conclusión es que los estudiantes que no aprobaron el curso y llegaron a las últimas semanas, aumentaron en promedio el número de sesiones realizadas, mientras que el primero módulo en promedio existía poco trabajo en cuanto al número de sesiones que realizaron este tipo de estudiantes.
 - En la **Figura 52**, se puede apreciar en la gráfica, que el comportamiento marcado en los modelos de procesos tan similar, se debe a la incidencia de los estudiantes que no aprobaron el curso, ya que el número de estos, es la mayoría con un 99.64%, lo que marca un sesgo en los modelos de procesos.

6.2. Trabajo Futuro

Esta tesis aparte de las contribuciones antes mencionadas, abre un camino para otras investigaciones a partir de cada una de las preguntas de investigación. Además, se recomienda garantizar, mediante nuevos experimentos la validez de este estudio, para ello se adjuntan los scripts, y se puede tomar este documento como guía para replicar el experimento (es posible que la arquitectura de datos de Coursera, haya variado levemente considerando que se usaron datos de un curso desplegado en el 2015).

En relación a la P.I.1. una línea de trabajo futuro es:

- Estudiar más a detalle las sesiones que realizan los estudiantes, con nuevos parámetros que logren determinar cuáles son las sesiones que mejor se aprovechan y por qué.

En relación a la P.I.2. la línea de trabajo futuro es la siguiente:

- Complementar este estudio como guía y aplicarlo a un curso más actual, donde se estudien las sesiones y las repercusiones con el trabajo colaborativo (foros y revisiones por pares), ya que en este estudio se limitó ya que era una cantidad no relevante las personas que participaron en dichas actividades colaborativas.
- Implementar un cuestionario adicional de estilos de aprendizaje para complementar al estudio actual que se basa en sesiones de los



estudiantes (duración por sesión, número de sesiones, interacción con actividades como video-lecturas, evaluaciones, totales).

- Implementar un cuestionario adicional de auto regulación de aprendizaje (SRL), para complementar al estudio actual que se basa en sesiones de los estudiantes (duración por sesión, número de sesiones, interacción con actividades como video-lecturas, evaluaciones, totales).
- Complementar este estudio tomando en cuenta las acciones que realiza el usuario dentro del navegador, mientras realiza su sesión de estudio.

En relación a la P.I.3. la línea de trabajo futuro es la siguiente:

- Complementar la clasificación de los estudiantes que aprueban y los que no aprueban, con variables de rendimiento académico como, por ejemplo: Promedio general del curso, calificaciones por actividades calificadas, entre otras.

Bibliografía

- Aalst, W. M. P. van der. (2014). *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer. <https://doi.org/10.1007/978-3-642-19345-3>
- Aalst, W Van Der, b, V. D. A. W. ., A.a, A., A.K.A.ax, D. M., F.z, A., ba, B. T. ., ... M.b, W. (2012). Process mining manifesto. *Lecture Notes in Business Information Processing*, 99 LNBIP(PART 1), 169–194. https://doi.org/10.1007/978-3-642-28108-2_19
- Aalst, Wil Van Der, Adriansyah, A., Karla, A., Medeiros, A. De, Arcieri, F., Blickle, T., ... Wynn, M. (2011). Manifiesto sobre Minería de Procesos.
- Allen, I. E., & Seaman, J. (2013). Changing course: ten years of tracking online education in the United States. *Nursing Standard (Royal College of Nursing (Great Britain) : 1987)*.
- Anderson, T. (2013). Promise and/or peril: MOOCs and open and distance learning. *Commonwealth of Learning*, (March), 1–9. Retrieved from http://scholar.google.co.uk/scholar?q=promise+and%2For+peril%3A+MOOCs&btnG=&hl=en&as_sdt=0%2C5#1
- Anderson, Terry, & McGreal, R. (2012). Disruptive pedagogies and technologies in universities. *Educational Technology and Society*.
- Atiaja, A. (2016). Los MOOCs : evolución , problemas y perspectivas hasta el 2015, 5(3), 105–123.
- Bartolomé, A., & Steffens, K. (2015). ¿Son los MOOC una alternativa de aprendizaje? *Are MOOCs Promising Learning Environments?*, 22(44), 91–99. Retrieved from 10.3916/C44-2015-10%5Cn<http://search.ebscohost.com/login.aspx?direct=true&db=a9h&AN=100244633&lang=es&site=ehost-live>
- Bernal-González, M. C. (2015). *Abandono de los estudiantes en los MOOC*.
- Bogarín, A., Cerezo, R., & Romero, C. (2018). A survey on educational process mining. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(1). <https://doi.org/10.1002/widm.1230>
- Chaves, M. A., & Córdoba, E. R. (2014). Deciphering event logs in SharePoint Server: A methodology based on process mining. In *Proceedings of the 2014 Latin American Computing Conference, CLEI 2014*. <https://doi.org/10.1109/CLEI.2014.6965174>
- Cole, M. (1971). *The Cultural Context of Learning and Thinking: An Exploration in Experimental Anthropology*.



- Coursera. (2016). Data Guide.
- Coursera. (2019). About Us. Retrieved from <https://blog.coursera.org/about-old/>
- Davis, D., Seaton, D., Hauff, C., & Houben, G.-J. (2018). Toward large-scale learning design. <https://doi.org/10.1145/3231644.3231663>
- Downes, S. (2012). *Connectivism and Connective Knowledge: essays on meaning and learning networks*. National Research Council Canada, <http://www.>
- Fincham, O. E., Gasevic, D. V., Jovanovic, J. M., & Pardo, A. (2018). From Study Tactics to Learning Strategies: An Analytical Method for Extracting Interpretable Representations. *IEEE Transactions on Learning Technologies*. <https://doi.org/10.1109/TLT.2018.2823317>
- Fluxicon. (2010). ProM Tips — Which Mining Algorithm Should You Use? Retrieved from <https://fluxicon.com/blog/2010/10/prom-tips-mining-algorithm/>
- Fuxicon. (2019). Disco. Retrieved from <https://fluxicon.com/disco/>
- Geigle, C., & Zhai, C. X. (2017). Modeling MOOC student behavior with two-layer hidden markov models. In *L@S 2017 - Proceedings of the 4th (2017) ACM Conference on Learning at Scale*. <https://doi.org/10.1145/3051457.3053986>
- Hitachi Vantara. (2019). PENTAHO DATA INTEGRATION. Retrieved from <https://www.hitachivantara.com/en-us/products/big-data-integration-analytics/pentaho-data-integration.html>
- Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *Internet and Higher Education*. <https://doi.org/10.1016/j.iheduc.2017.02.001>
- Juhaňák, L., Zounek, J., & Rohlíková, L. (2017). Using process mining to analyze students' quiz-taking behavior patterns in a learning management system. *Computers in Human Behavior*. <https://doi.org/10.1016/j.chb.2017.12.015>
- Kennedy, J. (2014). Characteristics of massive open online courses (MOOCs): A research review, 2009-2012. *Journal of Interactive Online Learning*.
- Kerres, M., & Preußler, A. (2013). Zum didaktischen Potenzial der Vorlesung: Auslaufmodell oder Zukunftsformat? In *Hochschuldidaktik im Zeichen von Heterogenität und Vielfalt*.
- Kizilcec, R. F., & Cohen, G. L. (2017). Eight-minute self-regulation intervention raises educational attainment at scale in individualist but not collectivist cultures. *Proceedings of the National Academy of Sciences*. <https://doi.org/10.1073/pnas.1611898114>



- Kovanović, V., Gašević, D., Dawson, S., Joksimović, S., Baker, R. S., & Hatala, M. (2015). Penetrating the black box of time-on-task estimation. In *ACM International Conference Proceeding Series*.
<https://doi.org/10.1145/2723576.2723623>
- Kruiderink, N. (2013). *Open buffet of higher education*.
- Langer, V., & Thilloßen, A. (2013). Freie Online-Angebote für Selbstlernende. Lebenslanges Lernen mit dem Internet. *Lehrbuch Für Lernen Und Lehren Mit Technologien*.
- Littlejohn, A. (2013). CEMCA EdTech Notes: Understanding Massive Open Online Courses. Retrieved from <http://www.cemca.org.in>
- Liyanagunawardena, T. R., Adams, A. A., & Williams, S. A. (2013). MOOCs : A Systematic Study of the Published, *14*(3).
- Lletí, R., Ortiz, M. C., Sarabia, L. A., & Sánchez, M. S. (2004). Selecting variables for k-means cluster analysis by using a genetic algorithm that optimises the silhouettes. *Analytica Chimica Acta*.
<https://doi.org/10.1016/j.aca.2003.12.020>
- Maldonado-Mahauad, J., Pérez-Sanagustín, M., Kizilcec, R. F., Morales, N., & Muñoz-Gama, J. (2018). Mining theory-based patterns from Big data: Identifying self-regulated learning strategies in Massive Open Online Courses. *Computers in Human Behavior*.
<https://doi.org/10.1016/j.chb.2017.11.011>
- Maldonado, J. J., Palta, R., Vázquez, J., Bermeo, J. L., Pérez-sanagustín, M., & Muñoz-gama, J. (2016). Exploring Differences in How Learners Navigate in MOOCs Based on Self-Regulated Learning and Learning Styles.
- Mcauley, A. A., Stewart, B., Siemens, G., & Cormier, D. (2010). *MOOC Model for Digital Practice*. Onlineaccessed 170411.
- Mukala, M. P., Buijs, J. C. A. M., Leemans, M., & van der Aalst, W. M. P. (2015). Learning analytics on coursera event data, (Simpda). Retrieved from <http://repository.tue.nl/5d173ca1-eb71-47e4-b560-715cdacb91be>
- Nesterko, S. O., Seaton, D., Reich, J., & Chuang, I. (2013). Evaluating Geographic Data in MOOCs. *NIPS Workshop on Data Driven Education (Pp. 1–7)*.
- Normandi Atiaja Atiaja, L., & Segundo Guerrero Proenza, R. (2016). MOOCs: Origin, characterization, principal problems and challenges in higher education. *Journal of E-Learning and Knowledge Society*.
- Orellana García, A., & R. L. (2018). Tecnologías Para El Análisis De Procesos Educativos Desde Big Data : Learning Analytics Y Process Mining Technologies for the Analysis of Educational Processes From Big Data :



Learning Analytics and Process Mining.

- Pardos, Z. A., & Schneider, E. (2013). AIED 2013 workshop, 1.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' Introduction to the Special Section on Replicability in Psychological Science: A Crisis of Confidence? *Perspectives on Psychological Science*. <https://doi.org/10.1177/1745691612465253>
- Pérez-Álvarez, R., Maldonado-Mahauad, J., & Pérez-Sanagustín, M. (2018). How to map learning activities through URLs? the case of Coursera platform. *CEUR Workshop Proceedings, 2224*, 25–34.
- Process Mining Group. (2016). ProM. Retrieved from <http://www.processmining.org/prom/start>
- ProM. (2019). ProM Tools. Retrieved from <http://www.promtools.org/doku.php>
- Rojas, E., Muñoz-Gama, J., Sepúlveda, M., & Capurro, D. (2016). Process mining in healthcare: A literature review. *Journal of Biomedical Informatics*. <https://doi.org/10.1016/j.jbi.2016.04.007>
- Romero, M., & Usart, M. (2014). The time factor in MOOCs: Time-on-task, interaction temporal patterns, and time perspectives in a MOOC. In *CSEDU 2014 - Proceedings of the 6th International Conference on Computer Supported Education*.
- Rudnitckaia, J. (2017). Process mining. Data science in action. *Process Mining: Data Science in Action*. <https://doi.org/10.1007/978-3-662-49851-4>
- Saint, J., Gašević, D., P. A. (2018). Detecting Learning Strategies Through Process Mining. *European Conference on Technology Enhanced Learning (2018)*, 385–398.
- Sapunar-Opazo, D., Pérez-Álvarez, R., Maldonado-Mahauad, J., Alario-Hoyos, C., & Pérez-Sanagustín, M. (2018). Analyzing learners' activity beyond the MOOC. In *CEUR Workshop Proceedings*.
- Shafiq, H., Wani, Z., Mahajan, I., & Qadri, U. (2017). Courses beyond borders: A case study of MOOC platform Coursera. *Library Philosophy & Practice*, (November), 1–15. Retrieved from <http://search.ebscohost.com/login.aspx?direct=true&db=llf&AN=127888639&site=ehost-live>
- Shah, D. (2018a). By The Numbers: MOOCs in 2018. Retrieved from <https://www.classcentral.com/report/mooc-stats-2018/>
- Shah, D. (2018b). Year of MOOC-based Degrees: A Review of MOOC Stats and Trends in 2018. Retrieved from <https://www.classcentral.com/report/moocs-stats-and-trends-2018/>



Sharples, M., Adams, A., Alozie, N., Ferguson, R., Fitzgerald, E., Gaved, M., ... Yarnall, L. (2015). *Innovating Pedagogy 2015. Open University Innovation Report 4.*

Siemens, G. (2005). Connectivism: A Learning Theory for the Digital Age. *International Journal of Instructional Technology and Distance Learning*. Retrieved from http://www.itdl.org/Journal/Jan_05/article01.htm

Stephen, B. (2013). Back to the Future with MOOCs. *Icicte*.

Torres, D. (2013). Reflexiones y primeros resultados de Moocs en Iberoamérica: Unedcoma y UNX. *Campus Virtuales*. Retrieved from <http://www.revistacampusvirtuales.es/campusvirtuales/2/5.pdf>

Trigwell, K., & Prosser, M. (1991). Improving the quality of student learning: the influence of learning context and student approaches to learning on learning outcomes. *Higher Education*. <https://doi.org/10.1007/BF00132290>

Van den Beemt, A., Buijs, J., & Van der Aalst, W. (2018). Analysing Structured Learning Behaviour in Massive Open Online Courses (MOOCs): An Approach Based on Process Mining and Clustering. *The International Review of Research in Open and Distributed Learning*, 19(5). <https://doi.org/10.19173/irrodl.v19i5.3748>

Van Der Aalst, Wil, Adriansyah, A., Alves De Medeiros, A. K., Arcieri, F., Baier, T., Blickle, T., ... Wynn, M. (2011). Manifiesto sobre Minería de Procesos. In *International Conference on Business Process Management*.

Van Eck, M. L., Lu, X., Leemans, S. J. J., & Van Der Aalst, W. M. P. (2015). PM²: A process mining project methodology. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. https://doi.org/10.1007/978-3-319-19069-3_19

Wang, Y., & Baker, R. S. J. D. (2015). Content or platform : Why do students complete MOOCs ? *Journal of Online Learning and Teaching*, 11(1), 17–30. <https://doi.org/10.1097/ICO.0b013e31812dfad1>

Wedekind, J. (2013). MOOCs – eine Herausforderung für die Hochschulen? In *Hochschuldidaktik im Zeichen von Heterogenität und Vielfalt*.

Weijters, A. J. M. M., van der Aalst, W. M. P., & de Medeiros, A. K. A. (2006). Process Mining with the HeuristicsMiner Algorithm. *Beta Working Papers*.

Youell, A. (2011). What Is a Course, 1–26. Retrieved from http://www.hesa.ac.uk/dox/publications/The_Course_Report.pdf

Yousef, A. M. F., Chatti, M. A., Schroeder, U., Wosnitza, M., & Jakobs, H. (2014). MOOCs a review of the state-of-the-art. In *CSEdu 2014 - Proceedings of the 6th International Conference on Computer Supported Education*.



Yuan, L., & Powell, S. (2013). *MOOCs and Open Education: Implications for Higher Education*. Retrieved from <http://publications.cetis.org.uk/wp-content/uploads/2013/03/MOOCs-and-Open-Education.pdf>



Anexos

Anexo 1: Tabla de ejemplos de proveedores xMOOC

Proveedor de xMOOC	URL
Coursera	https://www.coursera.org
Edx	https://www.edx.org
Udacity	https://www.udacity.com
Futurelearn	http://futurelearn.com
Codecademy	http://www.codecademy.com
Openlearning	https://www.openlearning.com
NPTEL	http://nptelonlinecourses.iitm.ac.in
Khan Academy	https://www.khanacademy.org
Udemy	https://www.udemy.com
ALISON	http://alison.com

Tabla 10: Ejemplos de Proveedores xMOOC (Littlejohn, 2013)



Anexo 2: Script de Python – Procesamiento de sesiones

```
import psycopg2
```

```
try:
```

```
    connection = psycopg2.connect(user = "adminPython",
```

```
                                   password = "XXXXXXXXXXXX",
```

```
                                   host = "127.0.0.1",
```

```
                                   port = "5432",
```

```
                                   database = "LA")
```

```
    cursor = connection.cursor()
```

```
        #Tabla que contiene todas las marcas de tiempo de las actividades de los
```

```
        #usuarios que al menos terminaron una actividad
```

```
        #Se ordena en base a los usuarios y su marca de tiempo en segundo lugar
```

```
        #para poder recorrerlo en un bucle
```

```
    sql="SELECT * from usuarios_actividades_ts order by 1,3"
```

```
    cursor.execute(sql)
```

```
    data = cursor.fetchall();
```

```
except (Exception, psycopg2.Error) as error:
```

```
    print ("Error while connecting to PostgreSQL", error)
```

```
### Para insertar resultados en la DB
```

```
insert_table_session = """ INSERT INTO user_session (ucchile_user_id, sessions) VALUES  
(%s,%s) """
```

```
insert_table_session_duration = """ INSERT INTO user_session_durations  
(ucchile_user_id, duration) VALUES (%s,%s) """
```

```
def insert_table(sql_insert_query, records):
```



try:

```
connection = psycopg2.connect(user = "adminPython",  
                               password = "XXXXXXXXXXXXX ",  
                               host = "127.0.0.1",  
                               port = "5432",  
                               database = "LA")
```

```
cursor = connection.cursor()
```

```
result = cursor.executemany(sql_insert_query, records)
```

```
connection.commit()
```

```
print(cursor.rowcount, "user_session creado con exito")
```

except (Exception, psycopg2.Error) as error:

```
print("Error al insertar registros en user_session {}".format(error))
```

finally:

```
if (connection):
```

```
    cursor.close()
```

```
    connection.close()
```

Inicio de procesamiento

```
users = []
```

```
current_user_id = data[0][0]
```

```
current_user = []
```

```
for i in data:
```

```
    user_id = i[0]
```

```
    if user_id == current_user_id:
```



```
current_user.append(i)

else:

    users.append(current_user)

    current_user = []

    current_user_id = user_id

    current_user.append(i)

users.append(current_user)

def get_user_sessions(u):

    user_total_session = []

    current_session = []

    previous_time = u[0][2]

    for activity in u:

        time = activity[2]

        #Transformar a minutos los segundos de la sesión

        if (((time - previous_time).total_seconds())/60) < 40:

            current_session.append(activity)

        else:

            user_total_session.append(current_session)

            current_session = []

            current_session.append(activity)

        previous_time = time

    user_total_session.append(current_session)

    return user_total_session;
```



```
### Obtención de sesiones para todos los usuarios

# session :: [(id, item, ts)]

# sessions :: [session] -- per user

sessions = []

for u in users:

    sessions.append(get_user_sessions(u))

def session_duration(session):

    timeDiff = session[len(session)-1][2] - session[0][2]

    return (timeDiff.total_seconds())/60

def user_number_of_session(user):

    return (len(user))

### Creacion de tabla de numero de sesiones

# table_session :: [(user_id, n_sessions)]

table_session = []

for user in sessions:

    user_id = user[0][0][0]

    n_sessions = user_number_of_session(user)

    table_session.append((user_id, n_sessions))

insert_table(insert_table_session, table_session)

### Creación de tabla de duración de sesiones

# table_session_durations :: [(user_id, session_duration)]

table_session_durations = []

for user in sessions:
```



```
user_id = user[0][0][0]

for session in user:

    duration = session_duration(session)

    table_session_durations.append((user_id, duration))

insert_table(insert_table_session_duration, table_session_durations)
```



Anexo 3: Script de Python – Agregar estado en las actividades

```
import psycopg2

def insert_table(sql_insert_query, records):

    try:

        connection = psycopg2.connect(user = "adminPython",

                                       password = "XXXXXXXXXXXXX ",

                                       host = "127.0.0.1",

                                       port = "5432",

                                       database = "LA")

        cursor = connection.cursor()

        result = cursor.executemany(sql_insert_query, records)

        connection.commit()

        print(cursor.rowcount, "log_user creado con exito")

    except (Exception, psycopg2.Error) as error:

        print("Error en log_user {}".format(error))

    finally:

        if (connection):

            cursor.close()

            connection.close()

def get_user_sessions(u):

    user_total_session = []

    current_session = []

    previous_time = u[0][1]
```



```
#contador de sesiones del mismo usuario para agregar de forma dinámica
```

```
cont=1
```

```
#listas auxiliares para cambiar estados de las actividades por usuario
```

```
aux_estado_1 = []
```

```
aux_estado_2 = []
```

```
user_total_session_state = []
```

```
current_session_state = []
```

```
for activity in u:
```

```
    time = activity[1]
```

```
    #obtengo el ID de la actividad por usuario para ver cuando la realizó
```

```
    #cuando la repitió, etc
```

```
    case_id = activity[0]
```

```
    estado_id = activity[3]
```

```
    estado=""
```

```
    if (((time - previous_time).total_seconds())/60) < 40:
```

```
        if (estado_id==1):
```

```
            if (case_id in aux_estado_1):
```

```
                estado="Iniciado - Nuevamente"
```

```
            else:
```

```
                estado="Iniciado"
```

```
                aux_estado_1.append(case_id)
```

```
        else:
```

```
            if (case_id in aux_estado_2):
```




```
estado="Repetido"

else:

    estado="Finalizado"

    aux_estado_2.append(case_id)

current_session.append((activity,estado,cont))

else:

    if (estado_id==1):

        if (case_id in aux_estado_1):

            estado="Iniciado - Nuevamente"

        else:

            estado="Iniciado"

            aux_estado_1.append(case_id)

    else:

        if (case_id in aux_estado_2):

            estado="Repetido"

        else:

            estado="Finalizado"

            aux_estado_2.append(case_id)

user_total_session.append(current_session)

current_session = []

cont=cont+1

current_session.append((activity,estado,cont))

previous_time = time
```



```
user_total_session.append(current_session)
```

```
return user_total_session;
```

try:

```
connection = psycopg2.connect(user = "adminPython",
```

```
    password = "XXXXXXXXXXXXX ",
```

```
    host = "127.0.0.1",
```

```
    port = "5432",
```

```
    database = "LA")
```

```
cursor = connection.cursor()
```

```
sql="SELECT * FROM log_actividades_estado_id"
```

```
cursor.execute(sql)
```

```
data = cursor.fetchall();
```

```
except (Exception, psycopg2.Error) as error :
```

```
    print ("Error while connecting to PostgreSQL", error)
```

```
### Inicio de procesamiento
```

```
#lista de todos los usuarios
```

```
users = []
```

```
current_user_id = data[0][4]
```

```
current_user = []
```

```
for i in data:
```

```
    user_id = i[4]
```

```
    if user_id == current_user_id:
```

```
        current_user.append(i)
```



else:

```
users.append(current_user)
```

```
current_user = []
```

```
current_user_id = user_id
```

```
current_user.append(i)
```

```
users.append(current_user)
```

```
### Obtención de sesiones para todos los usuarios
```

```
# session :: [(case_id, ts, nombreActividad, estadoActividad, user_id)]
```

```
# sessions :: [(case_id, ts, nombreActividad, estadoActividad_id, user_id, estadoActividad, sesion)]
```

```
sessions = []
```

```
for u in users:
```

```
    sessions.append(get_user_sessions(u))
```

```
# # user1_sessions = get_user_sessions(users[2])
```

```
### Para insertar resultados en la DB
```

```
insert_table_session_duration = """ INSERT INTO log_actividades_estado_sesion  
(case_id, time_stamp, actividad, estado_actividad_id, user_id, estado_actividad, sesion)  
VALUES (%s,%s,%s,%s,%s,%s,%s) """
```

```
### Creación de tabla de log de actividades completa
```

```
table_log = []
```

```
for log in sessions:
```

```
    for i in log:
```

```
        for j in i:
```

```
            table_log.append((j[0][0],j[0][1],j[0][2],j[0][3],j[0][4],j[1],j[2]))
```

```
insert_table(insert_table_session_duration, table_log)
```



Anexo 4: Script de Python – Agregar el número de interacciones con las actividades

```
import psycopg2

def insert_table(sql_insert_query, records):

    try:

        connection = psycopg2.connect(user = "adminPython",

                                       password = "XXXXXXXXXXXXX ",

                                       host = "127.0.0.1",

                                       port = "5432",

                                       database = "LA")

        cursor = connection.cursor()

        result = cursor.executemany(sql_insert_query, records)

        connection.commit()

        print(cursor.rowcount, "log_user creado con exito")

    except (Exception, psycopg2.Error) as error:

        print("Error en log_user {}".format(error))

    finally:

        if (connection):

            cursor.close()

            connection.close()

def get_user_sessions(u):

    #u :: [(user_id, ts, nombre_actividad, tipo_actividad, id_actividad, estado_actividad,

    sesion)]

    aux=[]
```



```
previous_session = u[0][6]

aux.append(previous_session)

for activity in u:

    if not (activity[6] in aux):

        aux.append(activity[6])

return sorted(aux)

def get_user_interactions(u):

    #u :: [(user_id, ts, nombre_actividad, tipo_actividad, id_actividad, estado_actividad,
    sesion)]

    sesiones=get_user_sessions(u)

    #Auxiliares para contar las interacciones

    cont_int_vl=0

    cont_int_exam=0

    int_total=0

    #auxiliar para agregar cada nuevo log individual

    log_aux=[]

    #Lista de cada usuario con sus interacciones por sesión

    log_interacciones_por_sesion=[]

    for num_sesion in sesiones:

        for act in u:

            tipo_actividad=act[3]

            sesion=act[6]

            if (num_sesion==sesion):

                if (tipo_actividad=="lecture"):
```



```
cont_int_vl=cont_int_vl+1

elif (tipo_actividad=="exam"):

    cont_int_exam=cont_int_exam+1

#Guardar los registros

int_total=cont_int_vl+cont_int_exam

log_aux=[act[0], num_sesion, cont_int_vl, cont_int_exam, int_total]

log_interacciones_por_sesion.append(log_aux)

#inicializar las variables contadoras y listas auxiliares

log_aux=[]

cont_int_exam=0

cont_int_vl=0

int_total=0

return log_interacciones_por_sesion

#Consulta para acceder al log de actividades #act_id, #ts, #nombre_act, #estado_act(1-2) #user_id

try:

    connection = psycopg2.connect(user = "adminPython",

                                   password = "XXXXXXXXXXXXX ",

                                   host = "127.0.0.1",

                                   port = "5432",

                                   database = "LA")

    cursor = connection.cursor()

    sql="SELECT * FROM log_actividades_estado_sesion_tipo_actividad"

    cursor.execute(sql)
```



```
#data: lista con todo el log de actividades, con el tipo de sesion

data = cursor.fetchall();

except (Exception, psycopg2.Error) as error :

    print ("Error while connecting to PostgreSQL", error)

#### Inicio de procesamiento

#lista de todos los usuarios

users = []

#data:: [(user_id, ts, nombre_actividad, tipo_actividad, id_actividad,
estado_actividad_sesion)]

current_user_id = data[0][0]

current_user = []

#data: lista con todo el log de actividades desde la DB

for i in data:

    user_id = i[0]

    if user_id == current_user_id:

        current_user.append(i)

    else:

        users.append(current_user)

        current_user = []

        current_user_id = user_id

        current_user.append(i)

users.append(current_user)

#### Obtencion de sesiones para todos los usuarios

# sessions :: [(user_id, sesion, num_int_vl, num_int_exam)]
```



```
sessions = []

for u in users:

    #print(u)

    sessions.append(get_user_interactions(u))

### Para insertar resultados en la DB

insert_table_session_duration = """ INSERT INTO log_interacciones_por_sesion_2
(user_id, sesion, num_int_vl, num_int_exam, int_total) VALUES (%s,%s,%s,%s,%s) """

### Creacion de tabla de log de actividades con el numero de interacciones por actividad

table_log = []

for log in sessions:

    for i in log:

        table_log.append(i)

insert_table(insert_table_session_duration, table_log)
```