



UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA INDUSTRIAL

TÁCTICAS DE ESTUDIO EN UN ENTORNO EN LÍNEA DE UN CURSO DE
INGENIERÍA MEDIANTE UN MODELO DE MARKOV OCULTO.

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERO CIVIL INDUSTRIAL.

ALAN FRANCO AMOLEF CÁRDENAS.

PROFESOR GUÍA:
SERGIO CELIS GUZMÁN.

MIEMBROS DE LA COMISIÓN:
CAROLINA SEGOVIA RIQUELME.
FELIPE VILDOSO CASTILLO.

SANTIAGO DE CHILE

2024

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE: Ingeniero Civil Industrial.

POR: Alan Franco Amolef Cárdenas.

FECHA: 2024.

PROFESOR GUÍA: Sergio Celis Guzmán.

TÁCTICAS DE ESTUDIO EN UN ENTORNO EN LÍNEA DE UN CURSO DE INGENIERÍA MEDIANTE UN MODELO DE MARKOV OCULTO.

La presente investigación se enfoca en el análisis de datos obtenidos de un entorno virtual de aprendizaje llamado EOL utilizado por la Universidad de Chile para complementar la educación universitaria presencial con medios digitales audiovisuales y escritos.

Actualmente, se recopilan grandes cantidades de datos de todo tipo, el área de la educación no es la excepción, de ahí que, surge la necesidad de utilizar dicha información con objetivos académicos y prácticos para aportar en la toma de decisiones de rediseño de cursos, conocimiento de los métodos de estudio explicados por las tácticas de aprendizaje. Esto logrado gracias a técnicas del *data science* y una metodología CRISP-DM, especializada en datos y entornos digitales.

El modelo estadístico para la programación y obtención de los resultados es el Modelo de Markov Oculto (HMM) que tiene por objetivo tomar variables observables y encontrar estados ocultos, en este caso, tácticas de aprendizaje, es decir, un conjunto de formas de estudiar y desenvolverse en el entorno virtual EOL, esto se logra ya que los registros contienen información de lo que los estudiantes hacen en la plataforma, lo que ven, ya sea video o lectura, la fecha y hora del evento y la sesión en que se lleva a cabo. Con esto, se crean variables derivadas como un desglose de la información para los estudiantes de la carrera de ingeniería.

Los resultados de esta investigación es que existen seis tácticas de aprendizaje, las que se diferencian en hora del día en que se estudia, días de la semana, el tipo de material consumen y como lo hacen. Se obtiene una matriz de transición que resumen las tácticas y su evolución, lo que ayuda a graficar el cambio en el tiempo y observar el largo plazo, encontrando un vector de probabilidad estacionaria, con ello, se concluye que existen dos tácticas a las que se converge sin importar por cual táctica se comience. Estos resultados pueden apoyar en el entendimiento de las tácticas de estudio y tomar medidas, ya que la convergencia a una o unas tácticas no garantiza aprendizaje efectivo ni representa el objetivo de la casa de estudios o del equipo docente.

Tabla de contenido

Capítulo 1: Contextualización	1
1. Introducción	1
1.1 Universidad de Chile	2
1.2 U-cursos	3
1.3 EOL	5
1.4 Motivación personal	7
2. Capítulo 2: Objetivos e hipótesis	9
2.1 Hipótesis	9
2.2 Objetivo General	9
2.3 Objetivos Específicos	10
3. Capítulo 3: Revisión de literatura	11
4. Capítulo 4: Metodología	17
4.1 Origen de los datos	20
4.2 Variables	21
4.3 Exploración de variables	26
4.4 Preparación de datos	31
5. Capítulo 5: Modelo markov oculto	38
5.1 Supuestos	38
5.2 ¿Por qué usar HMM?	39
5.3 Validación de supuestos del modelo	40
6. Capítulo 6: Resultados	46
6.1 Tácticas	46
6.2 Matriz de transición	48
6.3 Vector de probabilidad estacionario	50
7. Capítulo 7: Discusión	54
7.1 Aprendizajes del presente estudio	55
7.2 Otras aplicaciones	55
8. Capítulo 8: Conclusiones	58
8.1 Trabajo futuro	59
8.2 Sesgo y limitaciones	60
9. Bibliografía	61
10. Anexos	65

Capítulo 1: Contextualización

1. Introducción

En la era digital actual, el campo de la educación ha experimentado una transformación significativa, impulsada por el avance de las tecnologías de la información y la comunicación. Este cambio ha permitido el desarrollo y la aplicación de nuevas herramientas y metodologías para mejorar el proceso educativo. Por ejemplo, la minería de datos se ha aplicado para analizar datos enfocados en el proceso de enseñanza y aprendizaje de los estudiantes. Esto permite identificar factores que influyen en la retención del aprendizaje, rendimiento académico, asistencia y participación extracurricular (Quijije & Maldonado, 2023). Esta capacidad de análisis detallado proporciona a los educadores y administradores información valiosa para tomar decisiones informadas y mejorar la experiencia educativa de los estudiantes.

En este contexto, el desarrollo de plataformas de aprendizaje en línea y la implementación de herramientas analíticas avanzadas han emergido como elementos clave para mejorar el proceso de enseñanza-aprendizaje y adaptarse a las necesidades cambiantes de los estudiantes y de los sistemas educativos en general. Según Stanojević, Cenić y Cenić (2018), esta evolución ha permitido integrar los medios didácticos modernos, como los ordenadores y los programas educativos, que potencian la adquisición activa de conocimientos y la creatividad de los alumnos, permitiéndoles aprender y explorar de forma autónoma. De esta manera, las nuevas tecnologías y métodos educativos trabajan conjuntamente para ofrecer una educación más personalizada y efectiva.

Este trabajo, se enmarca en el proyecto FONDEF ID22i10048, cuyo objetivo principal es mejorar el proceso de enseñanza-aprendizaje y el diseño instruccional de cursos en línea mediante el desarrollo de nuevas analíticas de aprendizaje avanzadas. En colaboración con la Vicerrectoría de Tecnologías de la Información (VTI) de la Universidad de Chile, esta investigación se centra en las tácticas de estudio y posibles estrategias de los estudiantes. Los resultados serán un aporte al diseño y la posterior implementación en cursos en línea que la Universidad de Chile considere rediseñar o implementar.

Teniendo como base que existe una gran diversidad de métodos de estudio, este trabajo se propone investigar los diferentes enfoques de aprendizaje en un entorno virtual. Para ello, se analizan los datos recopilados durante el semestre de otoño de 2019 en la Facultad de Ciencias Físicas y Matemáticas de la Universidad de Chile, proporcionados por la plataforma de Educación en Línea (en adelante EOL) de la VTI.

A través del análisis de datos anonimizados y la aplicación de técnicas de análisis estadístico, precisamente el Modelo de Markov Oculto, se buscará identificar patrones y tendencias en el comportamiento de estudio de los estudiantes en periodos específicos y a través de las semanas.

Los hallazgos de este estudio no solo contribuirán al avance del conocimiento en el campo de la educación en línea, sino que también proporcionarán información valiosa para el diseño, rediseño y la implementación de estrategias educativas más efectivas y centradas en el estudiante. Al comprender mejor las tácticas de estudio de los estudiantes y su relación con el entorno educativo digital, se podrán desarrollar intervenciones y políticas más adecuadas para promover un aprendizaje significativo y duradero en entornos virtuales.

Para la conceptualización y comprensión más clara de este proyecto, se detallan algunos conceptos, entidades y funcionamientos relevantes que permiten guiar y establecer las consideraciones que se han tomado para la ejecución y desarrollo de la presente investigación

1.1 Universidad de Chile

La importancia de la Universidad de Chile en esta investigación es por el hecho de ser precursora del proyecto general y de incluir a estudiantes en apoyo de este. Además, la entidad facilitadora de los datos es dependiente de esta casa de estudios.

La Universidad de Chile es una de las instituciones de educación superior más antiguas y prestigiosas de Chile, estando dentro de las 500 mejores universidades del mundo. Fue fundada el 17 de septiembre de 1842 y es una de las más grandes del país, contando con un aproximado de 46.937 alumnos de pregrado y postgrado.

La Universidad de Chile cuenta con una amplia oferta de programas académicos en diversas áreas, incluyendo ciencias sociales, humanidades, ciencias naturales, ingeniería, medicina y artes. Además, tiene una destacada investigación científica y cuenta con una importante influencia en el desarrollo cultural y político de Chile, siendo así, la número 1 en investigación básica y aplicada de Chile.

Entre sus logros destacados se encuentran el haber sido la primera universidad en establecer la carrera de ingeniería en Chile, así como la formación de destacados intelectuales y líderes en distintas áreas del conocimiento, por ejemplo, 21 de los ex presidentes de la nación.

Cuenta con varias sedes y campus distribuidos en distintas locaciones y comunas de Santiago de Chile, en particular, este proyecto se desarrolla en la Facultad de Ciencias Físicas y Matemáticas (en adelante, FCFM), campus de las carreras profesionales de ingeniería y ciencias. Además, el proyecto está a cargo de la Vicerrectoría de Tecnologías de la

Información (en adelante, mencionada como VTI). Cuya función principal es liderar y gestionar los aspectos relacionados con la tecnología de la información en la universidad. Algunas de las funciones específicas de la VTI son: gestión de infraestructura tecnológica, desarrollo y soporte de sistemas, seguridad de la información, gestión de servicios tecnológicos y apoyo y capacitación tecnológica.

Los datos utilizados para esta investigación han sido extraídos a partir del semestre de otoño del año 2019 de la FCFM. Estos datos han sido anonimizados, con la finalidad de resguardar la información de cada uno de los estudiantes considerados, y han sido facilitados por la VTI, extraídos de la plataforma EOL, que es el sitio desarrollado para la Educación en Línea por la Universidad de Chile.

1.2 U-cursos

U-cursos es una plataforma educativa que la Universidad de Chile ha implementado para mejorar la experiencia de aprendizaje de sus estudiantes. Esta plataforma ofrece un espacio en línea donde los profesores pueden gestionar sus cursos de manera eficiente. A través de U-cursos, los profesores tienen la capacidad de cargar materiales de clase, como lecturas, videos y tareas, organizando la estructura del curso de una manera accesible para los estudiantes.

La utilidad de U-cursos radica en su capacidad para facilitar la comunicación y la interacción en línea. Los estudiantes pueden acceder a los recursos del curso en cualquier momento y desde cualquier lugar, lo que les brinda una mayor flexibilidad en su aprendizaje. Además, la plataforma ofrece foros de discusión que permiten a los estudiantes interactuar entre sí y con los profesores, fomentando el debate y la colaboración.

Los beneficios de U-cursos son notables. La flexibilidad que brinda permite a los estudiantes organizar su tiempo de estudio de manera más eficiente y acceder a los materiales de clase cuando mejor les convenga. La plataforma simplifica la entrega de tareas, la retroalimentación y el envío de comentarios, lo que ahorra tiempo y reduce la necesidad de trámites administrativos.

Figura 1: Visualización de inicio de U-cursos.

The screenshot shows the 'U-cursos' dashboard. At the top, there is a search bar and a user profile section with the name 'Nombre del estudiante'. Below this is a navigation menu with icons for Blog, Calendario, Canales, Correo, Cursos, Exámenes, Historial, Horario, Perfil, and Tareas. The main content area is titled 'Tareas' and 'Resumen de tareas para tus cursos activos'. It features a donut chart showing task completion status: 'En Plazo' (15.7%), 'Entrega...' (35.3%), and 'Conada'. Below the chart is a table of tasks:

Nº	Tarea	Fecha Limite	Estado
1	Tarea 4	31/08/2023 23:59 22 horas	En Plazo
2	Tarea 3	01/09/2023 23:59 1 día	En Plazo
3	Tarea 2	04/09/2023 23:59 4 días	En Plazo
4	Tarea 1	24/08/2023 23:59	Entregada 24 de Agosto a las 23:21 hrs.

Dentro de U-cursos, en la página inicial, Figura 1, los estudiantes pueden ver los cursos del semestre actual o anteriores, pueden ver su horario de clases, el horario y la sala correspondiente, también pueden ver el calendario, correos, pendientes, entre otros.

Figura 2: Opciones de los cursos.

The image displays a grid of 20 icons representing various course management options, arranged in four rows of five:

- Acta
- Administrar
- Asistencias
- Calendario
- Clase Virtual
- Correo
- EOL
- Encuestas
- Estadísticas
- Foro
- Historial
- Horario
- Integrantes
- Material Al...
- Material Do...
- Notas
- Novedades
- Presentació
- Tareas
- Tests

Dentro de cada curso al que el estudiante asiste, tiene las opciones presentadas en la Figura 2, donde tiene la posibilidad de interactuar con el equipo docente, hacer consultas a sus pares, entregar tareas, responder test online, ver sus notas, leer y/o descargar material de estudio, conocer su asistencia (en cursos que pidan asistencia), entre otras opciones. Lo que aquí aplica es el ícono EOL, que redirige al estudiante a la plataforma EOL con material complementario u obligatorio según el curso.

1.3 EOL

La plataforma de Educación en Línea (EOL) de la Universidad de Chile es un espacio virtual donde se ofrecen cursos, programas y recursos educativos en formato digital. Está diseñada para brindar acceso a la educación a distancia a través de Internet, permitiendo a estudiantes y profesionales aprender y mejorar sus habilidades sin la necesidad de estar físicamente presentes en un aula tradicional. Esta plataforma y sistema de trabajo, no busca necesariamente ser un remplazo de la educación presencial, sino que es una herramienta que podría facilitar el acceso en el caso de personas que por trabajo o distancia tienen dificultades para asistencia tradicional, o simplemente para complementar el trabajo en aula, pues brinda varias herramientas para trabajo y desarrollo académico.

En la plataforma EOL, los usuarios pueden encontrar una variedad de cursos en línea en diferentes áreas del conocimiento, como ciencias, humanidades, tecnología, negocios y más. Estos cursos suelen contar con materiales didácticos como videos que pueden ser visualizados, pausados o adelantados, también lecturas y test de alternativas y ejercicios interactivos en línea que al finalizar dan la retroalimentación respectiva con el puntaje correspondiente.

Los participantes pueden acceder a los contenidos en horarios flexibles, lo que brinda la oportunidad de adaptar el aprendizaje a sus propias agendas. Esta opción de aprendizaje no reemplaza las cátedras tradicionales, sino que las complementa y da la posibilidad a los estudiantes de consumir mayor contenido e información.

Los y las estudiantes pueden acceder a esta plataforma mediante un ícono presente en U-cursos.

Cabe mencionar que la modalidad EOL se está implementando gradualmente, actualmente la plataforma de Educación Online aloja cursos de pregrado, postgrado y educación continua de la Universidad de Chile, y cursos para externos dictados por la Universidad.

Figura 3: Visualización de un curso de EOL.

The screenshot shows the top navigation bar with the logo 'vti eol Educación Online' and links for 'Mis Cursos' and 'Explorar Cursos'. Below this is a blue header with the text 'Facultad de Ciencias Físicas y Matemáticas: Código del curso' and 'Nombre del curso'. A secondary navigation bar contains 'Curso', 'Progreso', 'Discusión', and 'Wiki'. The main content area displays 'Nombre del curso' followed by a list of modules: '¡Comencemos!', 'Módulo 1 | Título Tema 1', 'Módulo 2 | Título Tema 2', 'Módulo 3 | Título Tema 3', 'Módulo 4 | Título Tema 4', 'Módulo 5 | Título Tema 5', 'Módulo 6 | Título Tema 6', 'Módulo 7 | Título Tema 7', 'Módulo 8 | Título Tema 8', and 'Módulo 9 | Título Tema 9'. An 'Expandir todo' button is located in the top right of the list.

Figura 4: Desglose de partes del curso.

This screenshot shows the same course interface as Figure 3, but with 'Módulo 1 | El pro Tema 0' expanded. Under this module, 'Sesión 1 | Tema 1' is also expanded, revealing a list of three chapters: '1. Capítulo 1', '2. Capítulo 2', and '3. Capítulo 3'. 'Sesión 2 | Tema 2' remains collapsed. The rest of the course structure, including the navigation bars and other modules, is identical to the previous figure.

Al ingresar al ícono de EOL, los estudiantes pueden interactuar con los recursos disponibles puestos a su disposición (Figura 3), el curso está dividido en módulos, donde cada uno abarca un tema específico del syllabus, además, cada módulo está subdividido en sesiones (Figura 4), que, a su vez, están compuestas de temas particulares, los que son abarcados mediante videos educativos, material de lectura y test online para medir la comprensión del contenido.

1.4 Motivación personal

Como estudiante universitario, auxiliar y ayudante de ramos matemáticos en diferentes universidades y además como profesor particular de los mismos, he conocido diferentes perfiles de estudiantes, quienes han requerido metodologías específicas en cada uno de los casos, o al menos, consideraciones a tener en cuenta al momento de la enseñanza-aprendizaje, pues este es un sistema complejo que requiere trabajo de ambas partes.

Fuera de las diferentes formas de aprendizaje que existen (Hernández & Cabrera, 2021), también se presenta una diferencia en la temporalidad que destina cada quien para sus estudios, pues hay quienes tienen un estudio constante y van a la par con las clases, otras personas realizan estudios intermitentes, asimismo, otra proporción pospone su estudio para los últimos días previo a la instancia evaluativa. Las distintas maneras de estudiar también están destinadas a lograr diferentes objetivos y metas, hay técnicas que permiten aprobar la asignatura y otras técnicas que están destinadas a la reprobación, aunque depende particularmente de la persona, ya que podría tener conocimientos previos que implican menos estudio. Las capacidades personales del estudiante son igualmente fundamentales, de la misma forma es un factor determinante el entorno, la motivación, la disponibilidad, entre otros.

Conocer más sobre este ámbito ayudaría a los equipos docentes a diseñar los cursos pensando en diferentes métodos de estudio y apoyar el aprendizaje de los cursos dictados, después de todo, actualmente existe un aumento explosivo en la recolección de datos de todo tipo, en particular, se tienen los datos de navegación en la plataforma educativa EOL. Estos datos reflejan el comportamiento de los estudiantes, siendo la huella del proceso educativo e información objetiva de este proceso.

Por otro lado, los métodos de estudio podrían estar relacionados con variables observables, estos métodos de estudio se le llamarán “tácticas” para los fines de esta investigación.

Una táctica de aprendizaje se refiere a las estrategias y enfoques utilizados para facilitar el proceso de adquisición de conocimientos y habilidades por parte de los estudiantes. Las buenas prácticas de los estudiantes universitarios, como aprender activamente y optimizar el tiempo, están relacionadas con sus tácticas de aprendizaje, como los enfoques superficiales, profundo y de logro (Gozalo, León & Mendo 2020). Estas tácticas también existen en los

entornos virtuales, por ejemplo, estudiar todos los días en la tarde durante 2 horas sería una táctica, ver todo el contenido el día antes de la clase podría ser otra táctica, esto significa que una táctica es una manera específica de interactuar con la plataforma digital.

A partir de esto, nos cuestionamos, ¿las diferentes tácticas de estudio estarán vinculadas a un intervalo horario o a días específicos? De existir horarios de conexión para cada táctica, se esperaría que las personas que estudian constantemente tengan horario de “oficina”, mientras que los estudiantes que consumen el material con poca antelación tengan extensos horarios de estudio y a altas horas de la madrugada. También es válido cuestionarse: ¿todos aprenden con la misma táctica o tácticas? Es claro que deben existir diferentes métodos de aprendizaje y consumo de información, pero ¿Cuántas y cuáles son dichas tácticas?

Capítulo 2: Objetivos e hipótesis

2.1 Hipótesis

Hipótesis 1: Dado los datos observables de navegación de los estudiantes en la plataforma EOL, existen comportamientos no observables que se pueden encontrar mediante algoritmos de *data science* con el objetivo de comprender el comportamiento de estudio en línea.

Hipótesis 2: Existen intervalos horarios diferenciados para cada tipo de táctica de estudio adoptada por los usuarios en la plataforma EOL. Además, se anticipa que estas tendencias variarán entre los días de semana y los fines de semana, entonces, ¿estos intervalos temporales reflejan diferencias en las tácticas de estudio? Pues, podrían variar según la semana del semestre, el contexto de cada estudiante o sus necesidades y/o restricciones personales.

Hipótesis 3: Existe una táctica de estudio que es más frecuentemente utilizada por los estudiantes y se mantiene constante a lo largo del tiempo. Se postula que aquellos estudiantes que no emplean esta táctica al principio tenderán a implementarlo a medida que transcurran los días o semanas y se vean en la necesidad de utilizarla. Esto por probar nuevas maneras de estudiar o replicar lo que hacen sus pares.

2.2 Objetivo general

El objetivo general de esta investigación es identificar y posteriormente analizar las tácticas de estudio empleadas por los estudiantes en la plataforma de Educación en Línea (EOL) utilizada por la Universidad de Chile, aplicando técnicas de análisis de datos y modelos estadísticos. Se busca comprender las diferentes tácticas que los estudiantes emplean para abordar el proceso de estudio, con el fin de exponer información relevante que pueda ser usada para planificar y diseñar cursos de manera más efectiva, así como para implementar cambios que se ajusten mejor a las necesidades y preferencias de los estudiantes.

2.3 Objetivos específicos

1. Analizar los datos e identificar patrones de exploración web en la plataforma EOL, utilizada por estudiantes de la Facultad de Ciencias Físicas y Matemáticas, para identificar patrones recurrentes de acciones y eventos realizados por los usuarios.
2. Implementar de manera justificada un modelo estadístico previamente validado con los datos de navegación para modelar las secuencias de acciones observables, descubrir posibles estados ocultos (tácticas de estudio), clasificarlos para analizarlos por día o semana y resumir la información extraída.
3. Interpretar los resultados obtenidos de los análisis para comprender cómo las tácticas surgen, evolucionan a través del tiempo y convergen a un método único.
4. Expresar los aprendizajes y hallazgos de la investigación, qué relevancia presenta en el área de estudio y plantear como sugerencia de trabajo futuro cómo poder continuar con la investigación para descubrir nuevos patrones, nuevas formas de estudiar y diferentes formas de implementar este modelo.

Capítulo 3: Revisión de literatura

La analítica de aprendizaje (*learning analytics* en inglés) es un campo de estudio dedicado a la recopilación, análisis y aplicación de datos sobre el aprendizaje y la enseñanza. Emplea técnicas analíticas y algoritmos de minería de datos para comprender el comportamiento de los estudiantes, prever el rendimiento académico y personalizar la experiencia educativa. Según Ferreira y Andrade (2013), el objetivo de este es proporcionar información significativa a educadores y estudiantes para mejorar la efectividad del proceso educativo y al mismo tiempo ofrecer nuevas formas para que los profesores comprendan a los estudiantes y hagan un uso efectivo de los recursos.

Esta información y sus resultados pueden ser presentados mediante tablas, gráficos, reportes, entre otros, así, se puede mostrar al equipo docente y también los estudiantes el desempeño y participación en actividades en línea, como el uso de plataformas educativas. Esto permite identificar áreas de mejora y personalizar el aprendizaje (Villa et al., 2022) con el objetivo de tener mejores versiones del curso en el futuro y entregar una experiencia educativa mucho más significativa a quienes utilizan esta metodología para su formación.

En Chile, la educación universitaria no es obligatoria, por lo que quienes ingresan, es por motivos y necesidades personales, de ahí que, al asumir que son seres racionales, buscan la aprobación de las asignaturas, el aprendizaje y finalizar los estudios en el plazo correspondiente. Estas asignaturas tienen evaluaciones de distintos tipos, la que los estudiantes deben aprobar para avanzar su malla curricular.

Para lograrlo, los estudiantes universitarios utilizan una variedad de métodos de aprendizaje y de estudio, así como los profesores de cátedras utilizan diferentes maneras de transmitir el conocimiento para que se logre el aprendizaje, lo anterior ha sido investigado y es el espíritu de esta memoria.

La educación superior está en constante evolución, buscando siempre mejorar los métodos de enseñanza para optimizar el aprendizaje de los estudiantes. En este enfoque, Gargallo y Jiménez (2017) plantean que se ha investigado la percepción del contexto de aprendizaje por parte de los estudiantes. Así mismo, el estudio se centró en la percepción del contexto del aprendizaje, se examinó la incidencia de métodos centrados en el aprendizaje en estudiantes universitarios del primer curso de la Universidad Católica de Valencia. Se encontraron mejoras significativas en las habilidades de los alumnos y en su percepción del entorno de aprendizaje diseñado por los profesores, concluyendo que afecta positivamente. Estos hallazgos subrayan la importancia de diseñar ambientes de aprendizaje que no solo transmitan conocimientos, sino que también motiven y capaciten a los estudiantes para aprender de manera autónoma y colaborativa.

Con el avance de la tecnología, estas formas de enseñar y estudiar se han ido modificando, los recursos web son una herramienta al alcance de la mano para quienes tienen el acceso. Es decir, las innovaciones en la enseñanza o las diferentes formas de compartir el conocimiento están ligadas a plataformas web que cada universidad posee o las plataformas que la universidad incentiva en usar. Las plataformas virtuales de educación en universidades españolas se han convertido en un recurso esencial para el correcto desarrollo de los estudios y la optimización del aprendizaje de los estudiantes. Estas plataformas garantizan la accesibilidad a su contenido web y recursos de aprendizaje, lo que es fundamental para una educación personalizada (Gutiérrez, Solera & García, 2015).

Relacionado con lo anterior, Consuegra, Sucre y Bustamante (2023) nos dicen que el avance de la tecnología en la educación superior ha sido significativo y ha transformado la forma en que se enseña y se aprende. A través de la implementación de herramientas tecnológicas en la labor docente, se ha observado una mejora en el rendimiento académico de los estudiantes y en el desarrollo de las clases impartidas en programas de pregrado y postgrados. Estas herramientas tecnológicas influyen positivamente en el proceso de enseñanza y aprendizaje, facilitando la adquisición y conclusión de nuevos aprendizajes.

En el contexto actual, la tecnología juega un papel crucial en la transformación de la educación superior. En el contexto de la sociedad de la información, las plataformas web de las universidades se han convertido en un recurso importante para la difusión y promoción de las casas de estudio superiores. Estas plataformas no solo sirven para atraer a nuevos estudiantes, sino que también se integran en los procesos centrales de producción de conocimiento, enseñanza-aprendizaje y gestión administrativa de las universidades (Torres, 2010). La digitalización de estos procesos ha facilitado una mayor eficiencia y accesibilidad en la administración y el intercambio de información.

Además de las plataformas web, la integración de nuevas metodologías educativas ha sido esencial para aprovechar al máximo las tecnologías disponibles. El uso de estrategias didácticas metodológicas que integran las tecnologías de la información y la comunicación (TIC) ha revolucionado la forma en que los estudiantes aprenden y los profesores enseñan. La incorporación de las TIC ha permitido un mayor acceso a la información, una interacción más dinámica en el aula y una mejora general en el proceso de enseñanza-aprendizaje. Algunas estrategias como el aula invertida, el aprendizaje colaborativo, el aprendizaje basado en proyectos, el aprendizaje autónomo, el aprendizaje por competencias y la gamificación han sido fundamentales en este avance tecnológico en la educación superior (Suasnabas, Carrasco, Morán, & Medranda, 2023), además, estas innovaciones no solo enriquecen la experiencia educativa, sino que también preparan a los estudiantes para un mundo laboral cada vez más digitalizado.

Retomando el gigante avance de la tecnología, la incorporación de TIC y plataformas web para apoyar el aprendizaje, se puede decir que la educación en línea es un complemento valioso para la educación presencial, realizando una correcta combinación entre

presencialidad y trabajo autónomo online se puede maximizar el aprendizaje y la retención de contenidos. Para su implementación y continuidad se debe considerar que la actitud de los estudiantes hacia la educación en línea es fundamental, aquellos con experiencia previa en tecnología y cursos en línea, que reciben apoyo de tutores preparados en sistemas fáciles de usar, tienden a tener una actitud más positiva y sienten menos estrés a la hora de hacer una actividad en línea (Hernández, Fernández & Pulido, 2018).

También se deben tener en cuenta las posibilidades de los estudiantes para participar o desarrollar actividades en línea, no es trivial que todos tienen un espacio para estudiar tranquilamente en sus hogares o una biblioteca cercana a sus domicilios. Nuñez (2022), explica que las condiciones de participación en cursos en línea están relacionadas y son dependientes al acceso y la calidad del internet, contar con un dispositivo adecuado y tener un espacio apropiado para realizar las actividades a distancia.

El avance de la tecnología en la educación superior ha sido significativo y ha transformado la forma en que se enseña y se aprende. Actualmente hay plataformas en línea que permiten hacer cálculos complejos para apoyar la resolución de problemas matemáticos, hay graficadores, simuladores de diversas áreas, como de medicina, asistencia social, psicología, encuestas, test, resúmenes, videos explicativos de personas que comparten su conocimiento, etc.

Todas estas nuevas formas de enseñar guían a los estudiantes a nuevas formas de estudiar; esto requiere un recurso de tiempo dedicado a ello que el estudiante lo debe autogestionar, ya que fuera de la sala de clases el estudiante es responsable de su propio aprendizaje: repasar contenidos, lecturas complementarias, visitar las plataformas indicadas, recomendadas o sugeridas. Debido a que existen diferentes formas de estudiar o autogestionar el tiempo, Espinoza et al. (2023) desarrollan un estudio donde se menciona que la duración del sueño en días de descanso típico o fin de semana fue mayor que en días de trabajo típico o entre semana, lo que sugiere que los estudiantes pueden aprovechar los fines de semana para descansar más y, por ende, estudiar menos. Este estudio, muestra también que la media de horas de sueño fue mayor en días de descanso que en días de trabajo típico entre semana en estudiantes que utilizan plataformas educativas. Lo que provocó que alrededor de un tercio de la población tuvo un *jetlag* social de más de dos horas, esto se refiere a un cambio temporal en los patrones de sueño, cuyos efectos son similares en el cuerpo como viajar a una zona horaria diferente.

Algunas de las innovaciones en la enseñanza y el aprendizaje están relacionadas con la participación en equipos de trabajo mediante el uso de herramientas web que lo facilitan, algunas de las herramientas usadas son *Google Classroom*, videoconferencias, recursos didácticos, acceso a páginas web, videos, preguntas dirigidas, participación en chats, pizarras digitales (Gómez y Miranda, 2022). Además, sus usos influyen en el rendimiento académico al combinar educación presencial y en línea. ya que los estudiantes pueden acceder a estos recursos en cualquier momento.

Romero, Flores y Espinoza (2018), explican que la evolución tecnológica y la integración de las Tecnologías de Información y Comunicación han permitido ampliar la cobertura de la educación de calidad a través de entornos virtuales, facilitando el aprendizaje autónomo y la inclusión de segmentos de la población antes excluidos, como trabajadores a tiempo completo en el sector productivo.

Un elemento nuevo en educación son los LMS (*Learning Management Systems*), en educación, es una herramienta de e-learning utilizada para mejorar la experiencia de aprendizaje de los estudiantes y construir su comprensión de ciertos temas (Kasim & Khalid, 2016).

Debido a la importancia del *learning analytics* en la educación universitaria, Salazar, García, et al. (2017) pronostican que el futuro de los LMS en la educación superior formará una parte integral y en constante evolución del proceso educativo. Podrían avanzar hacia una mayor personalización del aprendizaje, permitiendo a los educadores adaptar los contenidos y actividades según las necesidades individuales de los estudiantes, lo que mejoraría la efectividad de la enseñanza además de que se espera que se integren tecnologías emergentes como la inteligencia artificial, el aprendizaje automático y la realidad virtual para ofrecer experiencias de aprendizaje más interactivas y efectivas.

Toda esta nueva tecnología y la rápida adaptabilidad a ella tiene como una parte impulsora a la pandemia que suprimió las fronteras entre lo íntimo y público, generando incluso procesos de negociación entre los miembros de la familia sobre el uso de dispositivos. Estas actividades educativas, familiares y sociales incrementaron el tiempo de conexión de jóvenes universitarios (Olmedo, 2023).

Estas tecnologías no solo ayudan al aprendizaje de los estudiantes y que puedan autogestionar sus horas de estudio, sino que también ayudan al entendimiento de esas formas de estudiar. Cada universidad tiene sus plataformas propias, donde los estudiantes pueden navegar y ejecutar las acciones que la página les provee. A partir de ahí, la casa de estudios puede descargar bases de datos con información del comportamiento de los estudiantes. Estas bases de datos luego se pueden trabajar para obtener conclusiones y eventualmente hacer cambio, ya sea, en el material a utilizar, el tipo de asignaciones o consignas e incluso el plazo de anterioridad con el que deberían entregarse los recursos o tareas. Al utilizar estas bases de datos, deben ser anonimizadas. López (2019), explica que, al anonimizar los datos, se eliminan los identificadores personales de los estudiantes, lo que protege su privacidad y cumple con las regulaciones sobre protección de datos. El uso de datos anonimizados genera confianza entre los estudiantes, ya que saben que su información personal está protegida y se asegura el cumplimiento de las leyes y regulaciones sobre protección de datos personales.

Luego de ser anonimizado, se obtienen beneficios adicionales para los encargados del análisis, modelado o cualquier otro trabajo sobre la base de datos ya que, al no tener acceso a información personal identificable, los investigadores pueden realizar análisis más

objetivos sin sesgos o influencias externas, los datos anonimizados pueden compartirse con otros investigadores o instituciones para fomentar la colaboración y el avance del conocimiento (Baltazar, Martínez y Sámano, 2021)

Esta información en correctos equipos de trabajo tecnológico puede generar grandes cambios educacionales y ayudar a mejorar la educación universitaria de varias formas ya que aplicando técnicas de *data science*, se pueden analizar datos sobre el rendimiento de los estudiantes para identificar factores que impactan la calidad de la educación, como el nivel de formación, condiciones contractuales y características de los docentes, esto permite tomar medidas para mejorar la calidad de la enseñanza.

Gómez, Laverde y Díaz (2016), explican que también es posible realizar métricas y predicciones, clúster y tendencias, con esto se puede por ejemplo hacer una detección temprana de riesgo de deserción mediante el monitoreo de indicadores como asistencia, participación y resultados parciales, el *data science* puede ayudar a identificar tempranamente estudiantes en riesgo de desertar para brindarles apoyo oportuno (Romero, Flores y Espinoza, 2018). Hay que mencionar que estas técnicas son mayormente utilizadas en entornos universitarios, aunque pueden ser extrapoladas a la educación básica y media del país, sería un avance positivo para el rendimiento académico y la educación de los ciudadanos.

Conocer las formas de estudiar y de navegar por la plataforma educativa es relevante por el motivo que los profesores universitarios esperan y desean ciertos comportamientos y actitudes de sus alumnos para facilitar el proceso de enseñanza-aprendizaje. Un aula cómoda para el equipo docente, con estudiantes deseosos de aprender y de utilizar las tecnologías disponibles, contribuye significativamente a este objetivo.

En el entorno universitario, es fundamental que los estudiantes se integren activamente en el proceso educativo. Los docentes valoran que los estudiantes participen activamente en discusiones, hagan preguntas y se involucren en las actividades del aula. Esto demuestra interés y compromiso con el aprendizaje (Chocarro, Sobrino y González, 2014). Además, esta participación beneficia tanto al estudiante como al grupo, enriqueciendo el ambiente de la clase y fomentando un entorno colaborativo donde todos pueden aprender unos de otros. Por lo tanto, una actitud proactiva y comprometida es clave para maximizar el potencial de aprendizaje y éxito en la universidad.

Es esencial que los estudiantes adopten una actitud responsable hacia su educación. Los profesores esperan que los alumnos asuman responsabilidad por su propio aprendizaje, cumplan con las tareas y lecturas asignadas, y desarrollen autonomía para regular su proceso de aprendizaje. Esta responsabilidad personal no solo ayuda a los estudiantes a adquirir conocimientos de manera más efectiva, sino que también los prepara para enfrentar futuros desafíos profesionales. La capacidad de gestionar su propio aprendizaje y cumplir con las

expectativas académicas es crucial para el éxito tanto en la universidad como en la vida profesional.

También se requiere la parte aplicada de los conocimientos. Es crucial que los estudiantes puedan poner en práctica sus aprendizajes teóricos en situaciones concretas. Por su parte, Boeta y Pinto (2017) explican que la finalidad de los profesores es que los alumnos puedan aplicar y transferir los conocimientos adquiridos a situaciones reales y contextos relevantes, demostrando comprensión profunda de los temas. Esta habilidad no solo muestra una sólida comprensión de los contenidos, sino que también prepara a los estudiantes para enfrentar y resolver problemas en sus futuras profesiones, haciendo que la transferencia de conocimiento sea un objetivo esencial en la educación superior.

La relación entre docentes y estudiantes se fortalece cuando existe una comunicación abierta y constructiva. Los docentes valoran recibir retroalimentación de los estudiantes sobre su enseñanza, ya sea a través de evaluaciones o interacciones, para mejorar continuamente sus prácticas pedagógicas (Bailey y Flores, 2018). Este intercambio de opiniones permite a los profesores ajustar y perfeccionar sus métodos, asegurando que se adapten mejor a las necesidades del alumnado. Al recibir y actuar sobre esta retroalimentación, se crea un ambiente educativo dinámico y en constante mejora, beneficiando tanto a los estudiantes como a los profesores.

Capítulo 4: Metodología

La metodología a utilizar para cumplir los objetivos de estudio y responder a las preguntas de investigación es CRISP-DM, (*Cross-Industry Standard Process for Data Mining*) se trata de una metodología popular para el desarrollo de proyectos de minería de datos.

Sobre esta metodología, Chuquival, Galindos, Maquera, Palza y Mamani (2011) destacan lo beneficiosa que es debido a su estructura robusta y bien definida para guiar proyectos de minería de datos de manera efectiva. CRISP-DM proporciona una estructura clara y bien definida que guía a los investigadores a través de las diferentes etapas de un proyecto de minería de datos, desde la comprensión del negocio y los datos hasta la implementación de soluciones. es flexible y se puede adaptar a diferentes proyectos y situaciones, lo que la hace ampliamente aplicable en diversos contextos de minería de datos. sigue un enfoque iterativo que permite a los equipos revisar y ajustar sus procesos a medida que avanzan en el proyecto, lo que facilita la mejora continua y la adaptación a los cambios y pone un fuerte énfasis en comprender los objetivos y necesidades del negocio, lo que garantiza que los resultados de la minería de datos estén alineados con los objetivos comerciales y generen valor.

Proporciona un marco estructurado que guía a lo largo de las diferentes etapas del proceso de minería de datos, desde la comprensión del problema hasta la implementación de la solución. Estas seis fases principales de la metodología CRISP-DM pasan a tener las definiciones detalladas a continuación a partir del trabajo titulado *Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en línea: un análisis en minería de datos* de los autores Mancilla, Leal, Sánchez y Vidal (2020), además, se explica cómo fueron empleadas y ayudaron la construcción de la presente investigación:

1. Comprensión del problema: En esta fase, se define el objetivo del proyecto y se comprenden los requerimientos y limitaciones de este. También se realiza una exploración inicial de los datos disponibles.

Esta etapa comenzó con una reunión inicial con el profesor guía y con la VTI, donde se presentó un proyecto de trabajo mayor que tenía la opción de incorporar a estudiantes tesistas. Se logró un acuerdo con la investigación general dando paso a la integración del equipo y este trabajo. Como hito inicial, se recibió una base de datos con la que se trabajó. Se conversaron los lineamientos del proyecto, los objetivos, los recursos de personal, el apoyo de la universidad y de los diferentes departamentos e incluso de colaboración con otra universidad.

Como un segundo paso, se entendieron los objetivos principales del proyecto completo para dar origen al objetivo general y específico de este trabajo en particular que, al concluir, pretender ser un aporte al proyecto completo.

Una exploración superficial de la base de datos ayudo a conocer que se encontraban en archivo .CSV, cuantas filas y columnas existen, los tipos de datos, entre otros factores que ayudan a tener un contacto inicial con los datos para el trabajo venidero.

En esta etapa también se hace una revisión de la literatura, se leen *papers*, artículos, otras tesis y memorias de temas similares para conocer como abordan el problema, como tratan los datos, que conclusiones obtuvieron o cuales fueron sus limitaciones.

2. Comprensión de los datos: Aquí se recopilan y exploran los datos relevantes para el proyecto. Se realiza un análisis exploratorio de los datos, se identifican patrones y se evalúa su calidad.

En esta fase, la base de datos es estudiada con mayores detenimientos, se realiza una exploración de los datos, se ordenan, se visualizan gráficos de variables, estadísticas de las filas agrupadas según distintas variables, se observan variables importantes como la hora, la fecha, los tipos de eventos, también se procura tener la información anonimizada, mediante reuniones se evalúa la calidad de las variables o su representatividad para lograr cumplir los objetivos.

3. Preparación de los datos: En esta etapa se lleva a cabo la limpieza, transformación y selección de los datos necesarios para el análisis. Se realizan tareas como la eliminación de valores atípicos, la imputación de datos faltantes y la normalización de variables.

Se revisan los datos de cada columna, se buscan datos faltantes y se eliminan datos atípicos, por ejemplo. Este paso ayudó a eliminar a un integrante de la base de datos, ya que era el profesor de cátedra, esto se evidencio gracias a que tenía registros antes de iniciar oficialmente el semestre incluso también después de haber terminado el mismo.

En esta etapa se tomó la fecha para crear nuevas variables como el día de la semana, si es semana o fin de semana, se usó la hora para crear la variable momento del día para saber si ocurrió en la mañana, tarde o noche. Estas variables se integraron como prefijo a los nombres de los eventos, también se creó la variable de identificación individual que es una cadena de caracteres que permite identificar a un individuo dentro de la base de datos, pero no en la vida real.

4. Modelado: En esta fase se seleccionan y aplican técnicas de modelado de datos para construir y evaluar modelos predictivos o descriptivos. Se pueden utilizar diversas técnicas, como regresión, clasificación, *clustering*, entre otras. En este proyecto se implementará el Modelo de Markov Oculto.

La etapa de modelado es fundamental, ya que entrega los resultados que buscamos según los datos disponibles; el modelo de Markov oculto es usado por motivos justificados en el capítulo de modelado de Markov oculto. Aquí se preparan los datos en un formato adecuado para Markov oculto, aunque antes de ello, se debe verificar que se cumplan todos los supuestos del modelo, pues, de no cumplirse, el modelo dará un resultado, pero no será correcto, representativo o importante para poder extraer conclusiones valiosas. En ese sentido, se deben estudiar las restricciones del modelo y los supuestos, luego con las variables preparadas, revisar que se cumpla todo lo necesario. En continuidad a esto, se deben preparar los datos en el formato que Markov oculto los solicita para trabajarlos y entregar los resultados del modelo, que en este caso es una matriz de transición que también puede expresarse como un grafo.

5. Evaluación: Aquí se evalúa el rendimiento y la calidad de los modelos construidos. Se realiza una validación cruzada y se comparan los resultados obtenidos con los objetivos establecidos en la fase de comprensión del problema.

El modelo de Markov oculto solicita en un inicio ingresar la cantidad de nodos que se desean, para escoger el óptimo se ejecuta el modelo simulando usar de 1 a 10 nodos. Para cada una de las elecciones se calcula la métrica AIC y BIC, estos valores son graficados junto a la cantidad de nodos y se escoge el que resulte óptimo para la finalidad. Esta es la forma de evaluar el haber escogido la mejor segmentación de los datos. Este paso es crucial para tener resultados significativos y relevantes, ya que, sin una argumentación sólida, los resultados podrían ser ambiguos o poco estables.

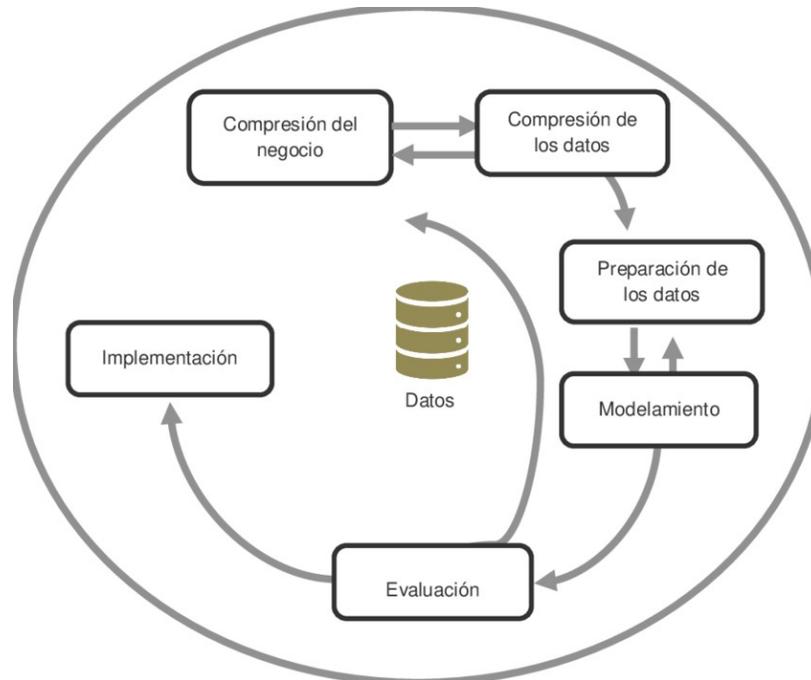
6. Implementación: En esta etapa, se implementa la solución en un entorno de producción o de servicio. Esto implica poner en práctica los resultados obtenidos y asegurarse de que sean utilizados de manera efectiva para abordar el problema planteado.

Es importante destacar que esta metodología es iterativa, lo que quiere decir que se puede volver a cualquier fase de esta, reorganizar ideas o reformular procesos ya realizados ajustándolo según lo que se requiera.

Para completar los objetivos de este trabajo, se va a prescindir de la etapa 6, la implementación, ya que los objetivos de investigación y las hipótesis se responden al obtener los resultados del modelo ejecutado.

Un diagrama de esta metodología se resume en la Figura 5.

FIGURA 5. Imagen extraída de *Factores asociados al éxito de los estudiantes en modalidad de aprendizaje en línea: un análisis en minería de datos*, Mancilla et al (2020)



4.1 Origen de los datos

La VTI de la UCH tiene acceso a los datos de navegación de los estudiantes en la plataforma EOL por el hecho de ser los administradores de dicho entorno virtual, estos datos son de carácter confidencial además de estar anonimizados, esto significa que no es posible reconocer a un usuario en particular.

Además, EOL funciona con código abierto, esto significa que se puede acceder al código fuente para que pueda ser estudiado, visto o modificado, se facilita el acceso a la comunidad de software quienes pueden incluir modificaciones, mejoras, incluir funcionalidades nuevas, así se masifica el software entre la comunidad y gradualmente se implementan mejoras gracias a los aportes computacionales de terceros.

Por otro lado, esta información es tratada de manera ética, con objetivos académicos e investigativos con la finalidad de contribuir a la educación en principio universitaria, pero con la posibilidad de extenderse a otros niveles e incluso a otros servicios o industrias.

Debido a que los datos son de los patrones de comportamiento de estudio y de las acciones llevadas a cabo por estudiantes reales, es que no se procederá a la individualización exacta del curso con el objetivo de mantener anonimizado el comportamiento de estudiantes individuales y el del curso completo.

Sin perjuicio de lo anterior, es posible mencionar que los datos trabajados son del primer semestre del año 2019. Este curso, que considera una única sección con duración semestral se impartió en formato presencial, incluyendo la plataforma EOL como un complemento de las clases. El curso era de carácter obligatorio de la malla curricular de ingeniería por lo que todos los estudiantes deben aprobarlo para poder avanzar académicamente.

4.2 Variables

La base de datos cuenta con 13 columnas para cada fila, las variables de las columnas se resumen en la Tabla 1, además, cada una de estas variables se explica en detalle.

Tabla 1: variables disponibles en la base de datos.

Nombre de usuario
Tipo de evento
Dirección IP
Agente
Página de navegación
Referencia previa
Lenguaje aceptado
Evento
Contexto
Tiempo
Página
Nombre
Sesión

Las variables disponibles y su descripción son como siguen:

1.- Nombre de usuario (*username*)

Esto se encuentra anonimizado, los nombres de usuarios son códigos de caracteres, esto permite identificar a una persona dentro de la base de datos, pero no identificarla en la realidad.

2.- Tipo de evento (*event_type*)

La información aquí capturada sigue una estructura de registro de eventos. En este esquema, cada vez que un estudiante realiza una acción, se genera una serie de atributos que se detallan

en la Tabla 2. Asimismo, en la Tabla 3 se presentan las categorías de eventos que se registran de forma inherente en los módulos, la conexión entre Tabla 2 y Tabla 3 es que los tipos de eventos que se registran es uno de los atributos que se registran en el momento que un estudiante explora la plataforma EOL. Al observar ambas tablas, se puede notar que esta información ofrece una visión integral sobre el comportamiento de los estudiantes en la plataforma. Se comprenden las actividades que llevan a cabo y el orden en que las realizan, el tiempo dedicado a cada actividad y los intervalos entre estas, su relación con el contenido y cómo interactúan entre sí.

3.- Dirección IP (*ip*)

Es una dirección de Protocolo de Internet (IP) perteneciente a una red privada según las convenciones del sistema de direccionamiento IP. Las direcciones IP son números que se asignan a dispositivos y equipos en una red para identificarlos y permitir la comunicación dentro de esa red y con otras redes.

En la base de datos se presenta como un número de 10 cifras que identifica al dispositivo utilizado conectado a la plataforma.

4.- Agente (*agent*)

La información aquí proporcionada se refiere a la cadena de agente de usuario que generalmente es enviada por un navegador web al acceder a un sitio web. Esta cadena de texto contiene información sobre el navegador y el sistema operativo que está utilizando el usuario para acceder al sitio.

Los servidores web pueden utilizar esta información para adaptar el contenido o la funcionalidad del sitio según las capacidades del navegador y el sistema operativo del usuario, es decir, la visualización se adapta, ya sea, si se está viendo en un computador, Tablet, celular u otros.

5.- Página de navegación (*host*)

La solicitud HTTP tiene el host "eol.uchile.cl" en todas las filas, está indicando que la comunicación se está llevando a cabo con el servidor de la plataforma de Educación en Línea de la Universidad de Chile. Esto es importante para dirigir la solicitud al servidor correcto y obtener la respuesta deseada, en este caso, relacionada con la educación en línea ofrecida por la Universidad de Chile mediante EOL.

6.- Referencia previa (*referer*)

La variable "*Referer*" (también conocida como "*Referer Header*") en una solicitud HTTP se utiliza para indicar la URL de la página web desde la cual se originó la solicitud actual.

Esto puede ser útil para los servidores web y aplicaciones para rastrear el origen de las solicitudes y brindar un contexto más completo sobre la actividad del usuario en el sitio. En este caso, las solicitudes están relacionadas con el contenido del curso en la plataforma de Educación en Línea de la Universidad de Chile.

7.- Lenguaje aceptado (*accept_language*)

La información de esta variable indica el idioma en que el usuario prefiere recibir las respuestas o si tienen cierto grado de flexibilidad o indiferencia en cuanto al idioma.

8.- Evento (*event*)

Esta variable describe una acción o cambio en la plataforma EOL. En este caso, está relacionado con la navegación o interacción con un bloque de contenido específico en un curso. Los campos proporcionados dan información sobre el estado previo y actual, la pestaña actual, la ubicación del widget y un identificador único para el bloque de contenido.

9.- Contexto (*context*)

Contiene información sobre el usuario involucrado, la organización, el curso y la ubicación dentro de la plataforma donde ocurrió la acción. Estos detalles permiten entender el contexto en el cual tuvo lugar el evento en cuestión.

10.- Tiempo (*time*)

Es una marca de tiempo que indica cuándo ocurrió un evento o una acción en la plataforma. La marca de tiempo está en formato ISO 8601, que es un estándar internacional para representar fechas y horas de una manera legible por máquinas y humanos.

Representa la fecha en formato año-mes-día, la hora en formato horas:minutos:segundos.milisegundos y el desplazamiento de la zona horaria en relación con la hora UTC (Tiempo Universal Coordinado) lo que también se conoce como hora estándar en Greenwich. Esta marca de tiempo proporciona información precisa sobre cuándo tuvo lugar el evento en cuestión.

11.- Página (*page*)

Representa la URL de la página web específica en la plataforma EOL que el usuario estaba viendo. Proporciona información sobre el curso, la sección de contenido y otros detalles relevantes para entender el contexto en el que se encuentra el usuario.

12.- Nombre (*name*)

Es un identificador para un evento o acción que ha ocurrido en la plataforma de Educación en Línea (EOL) de la Universidad de Chile. Es utilizado para identificar interacciones y acciones específicas que los usuarios llevan a cabo.

La información contenida en esta columna es igual a la información contenido en la columna `event_type`.

13.- Sesión (*session*)

Es un identificador único de sesión, una sesión generalmente se refiere a la interacción continua de un usuario con la plataforma durante un período determinado de tiempo.

Es un valor hexadecimal que representa un identificador único, se utiliza este tipo de identificadores para rastrear y distinguir diferentes sesiones de usuarios.

Cada vez que un usuario se conecta o interactúa con la plataforma, se genera una nueva sesión con un identificador único.

Tabla 2: Datos que componen los eventos registrados.

Atributo	Descripción
<i>TIMESTAMP</i>	Fecha y hora en que se registra el evento
SESIÓN	ID de sesión del usuario generado por Open edX
REFERENCIA	Vínculo a la página específica donde se generó el evento
TIPO EVENTO	Nombre del tipo de evento que se registró
USUARIO	Código de usuario anonimizado por razones de seguridad

Tabla 4: Tipos de eventos que se registran en los módulos.

Tipo de evento	Descripción
ADD A POST	Usuario inicia una conversación en un foro
ADD USER	Usuario agregado
CLICK BOOKMARK	Usuario usa un marcador que había sido previamente agregado manualmente
CLICK COURSE TOOL HEADING	Usuario selecciona el link "curso" que muestra la estructura completa de contenidos
CLICK NAVIGATION BAR	Usuario selecciona una pestaña (unidad) en la barra de navegación que muestra las unidades de una subsección específica
CLICK RESUME	Usuario ingresa al curso en la misma parte en que lo había dejado anteriormente
LOAD VIDEO	Cuándo un video es cargado completamente en la plataforma y está listo para visualizarse
NEXT PAGE	Usuario avanza de unidad mientras navega por el curso (botón "Siguiente")
PAGE CLOSE	Usuario cierra la página
PAUSE VIDEO	Usuario pausa el video
PLAY VIDEO	Usuario inicia reproducción de un video
PREVIOUS PAGE	Usuario retrocede de unidad mientras navega por el curso (botón "Anterior")
RESET PROBLEM	Usuario reinicia las respuestas a un problema
SAVE ANSWER OF A PROBLEM	Usuario guarda las respuestas de un problema
SEEK VIDEO	Usuario busca dentro de un video (avanza o retrocede en la barra de reproducción)
SELECT LINK	Usuario selecciona cualquier hyperlink dentro del contenido del curso
SERVER GRADES SUBMITTED	Registro cuando el usuario graba y envía exitosamente su respuesta a un problema
SERVER PROBLEM CHECKED	Registro cuando el usuario verifica un problema
SHOW ANSWER OF A PROBLEM	Usuario selecciona "mostrar respuesta" de un problema
SPEED CHANGE VIDEO	Usuario cambia velocidad de reproducción de un video
SUBMIT PROBLEM	Usuario envía su respuesta a un problema
USER ENROLLEMENT	Usuario es inscrito exitosamente en el curso
VIDEO COMPLETED	Usuario termina de ver completamente un video
VIEW FORUM	Usuario revisa una conversación de un foro

4.3 Exploración de variables

Vemos a continuación, análisis descriptivos de la base de datos, agregación de variables y características para comprenderlas de forma individual y las interacciones relevantes entre sí.

4.3.1 Tipos de eventos

En la Tabla 4 se observan los eventos ocurridos en el curso, se ejecutan 17 eventos de los 24 disponibles para la plataforma.

Cada uno de los 17 tipos de eventos, son acciones que los estudiantes realizaron en la plataforma EOL mientras interactuaban con ella, cuando un estudiante requiere visualizar un video, se guardará en la base de datos el evento “play_video”, si el estudiante pausa la reproducción o cambia la velocidad entonces se ejecutará el evento “pause_video” y “speed_change_video”, respectivamente. Así ocurre para cada evento que el estudiante accione mientras se encuentra en una sesión de trabajo en la plataforma.

Tabla 4: tipos de eventos desplegados en la base de datos

Tipo de evento	
1	seq_next
2	edx.ui.lms.sequence.next_selected
3	page_close
4	load_video
5	seq_prev
6	edx.bi.course.upgrade.sidebarupsell.displayed
7	edx.ui.lms.link_clicked
8	seq_goto
9	edx.course.home.resume_course.clicked
10	play_video
11	seek_video
12	pause_video
13	stop_video
14	edx.ui.lms.sequence.previous_selected
15	problem_show
16	speed_change_video
17	edx.course.tool.accessed

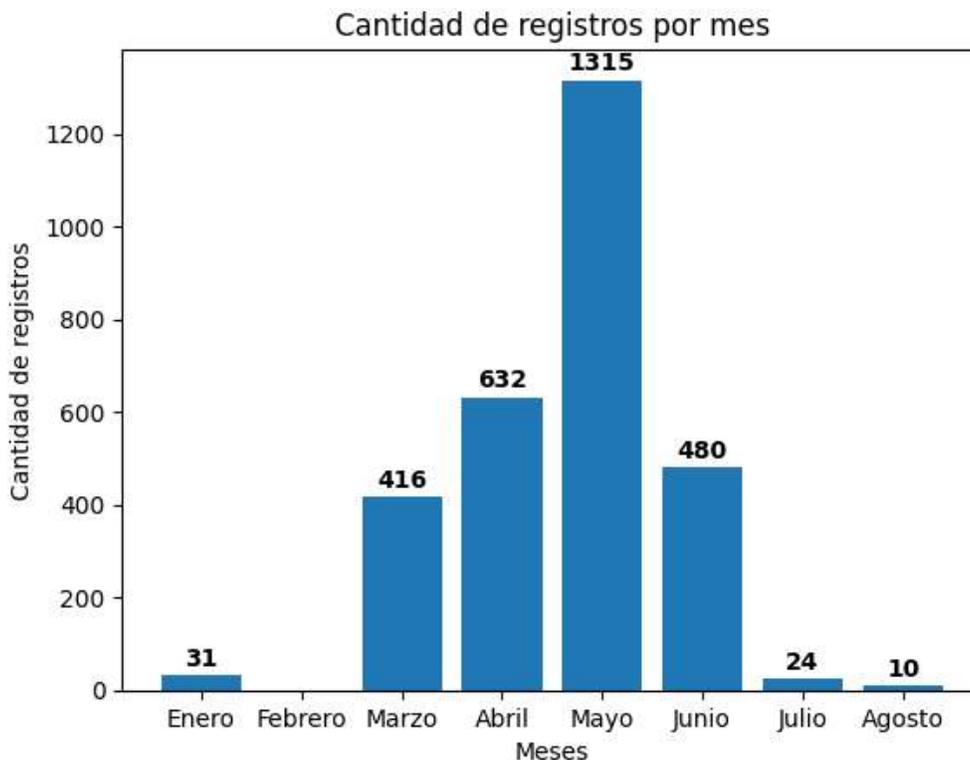
4.3.2 Registros por mes

En la Figura 6 se ve la cantidad de acciones que ocurrieron cada mes del semestre. Las acciones de enero tienen que ver con la configuración de la plataforma, cargar material docente de manera previa al inicio de las clases, exploración de la plataforma, entre otras actividades realizadas por el equipo docente. Luego, en febrero, la UCH se encuentra en periodo de vacaciones, por lo que es normal y coherente con la realidad que no exista ningún registro ese mes, es un periodo de cese de funciones.

Para los demás meses (marzo a julio) se ve una especie de distribución con una forma de campana, lo que no debe confundirse con distribución normal, ya que no necesariamente siguen una distribución normal y pueden presentar características distintas de una distribución estándar.

Finalmente, agosto es el equipo docente revisando el material o estudiantes descargando información para su biblioteca digital personal.

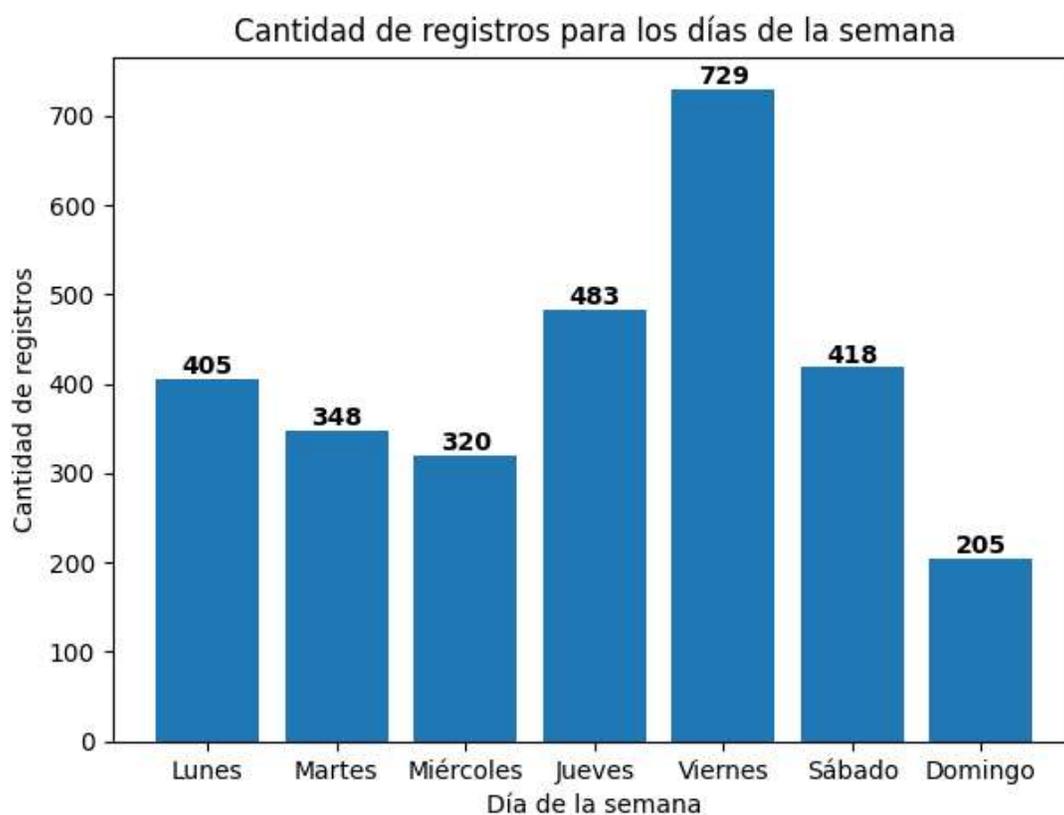
Figura 6: frecuencia de registros (acciones en la plataforma) acumulados por mes.



4.3.3 Registros por día

Se observa en la Figura 7, que las acciones en la plataforma EOL se concentran en la semana, es decir, entre lunes y viernes, esto representa un 78.58% del total. El fin de semana, sábado y domingo, representa el 21.42% restante. Se puede inferir que los estudiantes tienen una preferencia para acceder a la página durante la semana. Se observa también que el acceso de jueves y viernes representa un 53.04% del acceso durante la semana y un 41.68% del total considerando semana y fin de semana. La cantidad de estudiantes considerados para este análisis es el total del curso, 60 estudiantes.

Figura 7: frecuencia de registros acumulados por día de la semana.

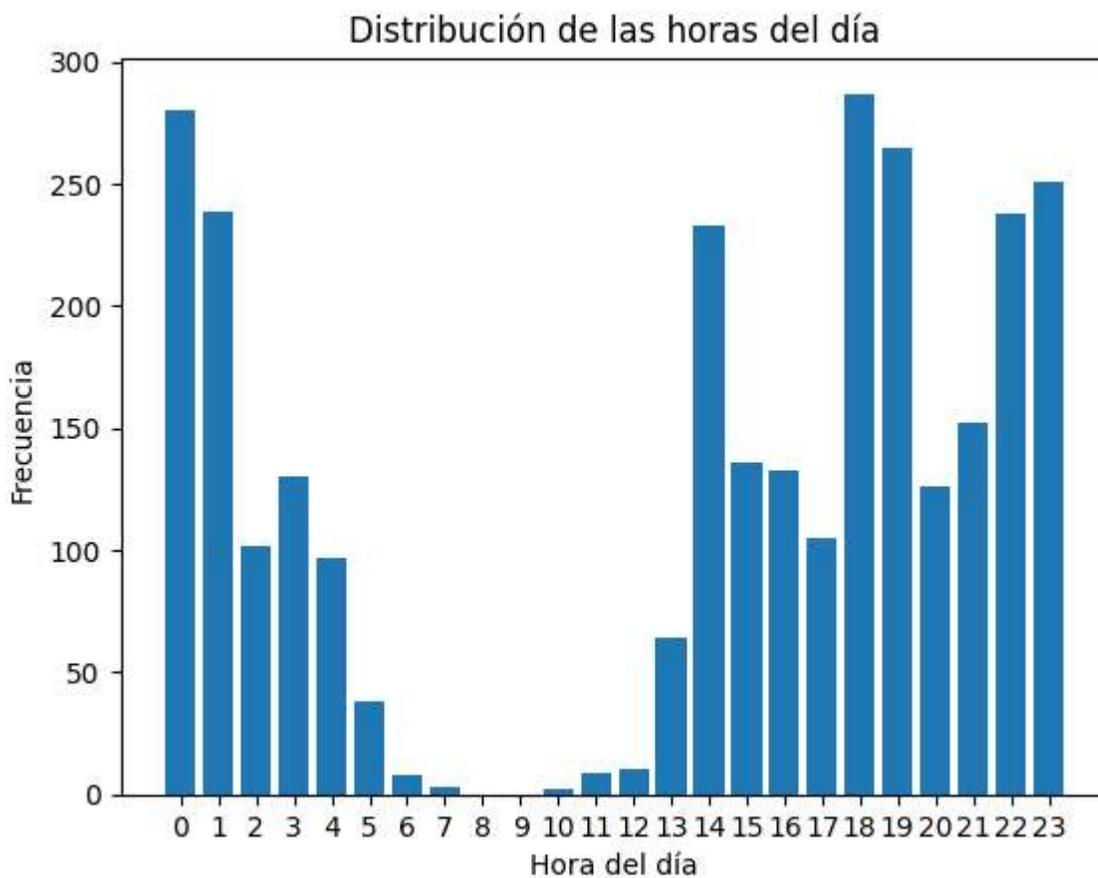


4.3.5 Registros por hora

En la distribución de las horas del día de la Figura 8, vemos cuántos registros se realizan por intervalos de hora con amplitud de una hora. El motivo de observar la frecuencia por hora es para conocer si existen horas de trabajo más concurridas durante el día.

Se ven 3 picos importantes, a las 14, 18 y 22 horas, también hay un valle en el acceso a EOL entre las 6 y 12 horas. Esto motiva a creer que hay horarios establecidos en que los estudiantes estudian o hacen otras actividades como dormir, descansar o de ocio, pudiendo convertirse en una variable clave en el modelado como se verá más adelante.

Figura 8: distribución de horas del día de estudio.



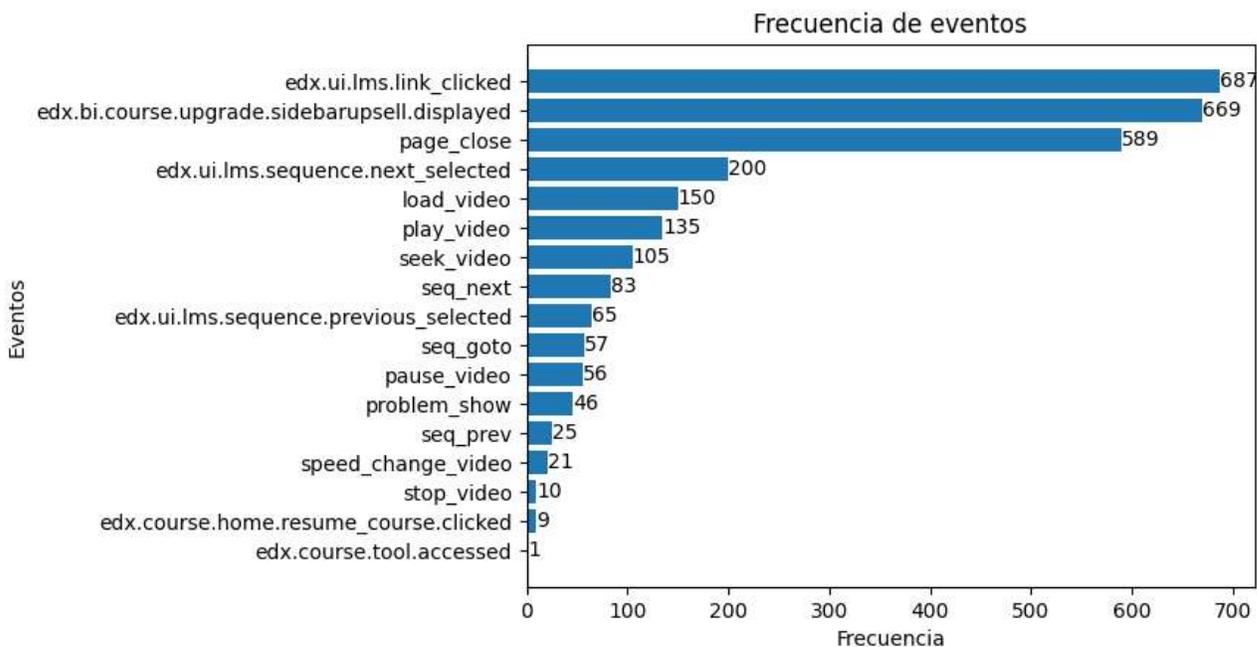
4.3.6 Frecuencia de eventos

En la Figura 9, se aprecia la frecuencia de eventos ocurridos en la plataforma de estudios, hay eventos que ocurren más que otros, como los son los eventos relacionados con la lectura, la realización de clics sobre elementos diferentes a videos o cambios de configuración.

Con menor frecuencia, ocurren eventos relacionados con la visualización de videos, estos son: cargar video, reproducir video, ver video desde un tiempo específico, pausar video, modificar velocidad, entre otros.

Esto también hace inferir que hay diferentes formas de estudiar, ya sea ver material audiovisual como videos, realizar ejercicios de lectura, incluso es posible de desglosar aún más, para quienes consumen el material audiovisual de una manera tradicional o de una manera no tradicional como modificando la velocidad o saltándose algunas partes del contenido.

Figura 9: frecuencia de eventos



4.4 Preparación de datos

4.4.1 Sufijo de día y hora

Después de la explotación de los datos, y basándose en las hipótesis de investigación, es necesario agregar los datos por hora del día y por día de la semana. Esto se logra con la columna “time”, la que contiene información precisa de cuando se produjo el evento.

El día de 24 horas, se dividió en tres bloques, mañana, tarde y noche. Para saber en cuál de los bloques se produjo el evento, se agrega un sufijo al nombre del evento, precedido por un *slash* que separa al evento y el sufijo del bloque horario.

El sufijo se asigna de la siguiente manera según la hora:

Mañana: para eventos ocurridos entre [6, 12)

Tarde: para eventos ocurridos entre [12, 21)

Noche: para eventos ocurridos entre [21, 6)

Finalmente, la simbología añadida es la siguiente:

/m: para eventos que ocurren en la mañana

/t: para eventos que ocurren en la tarde

/n: para eventos que ocurren en la noche

Luego, corresponde conocer si el evento ocurre durante la semana (lunes a viernes) o el fin de semana (sábado a domingo). Para ello se procede a hacer una categorización análoga a la anterior.

El sufijo se asigna de la siguiente manera según el día:

Semana: si el día es: {lunes, martes, miércoles, jueves, viernes}

Fin de semana: si el día es: {sábado, domingo}

Finalmente, la simbología añadida es la siguiente:

/s: para eventos que ocurren en la semana

/f: para eventos que ocurren el fin de semana

Así, para ejemplificar, si el evento se llama: evento.alpha, y ocurre un miércoles a las 17 horas, entonces se convertiría en: evento.alpha/t/s. De otra forma, otro evento llamado evento.betha que ocurra un sábado a las 23 horas, se renombraría a evento.betha/n/f.

Luego de realizar la conversión anterior, los tipos de eventos únicos aumentan a 67.

4.4.2 Fecha de inicio y semanas

Como se vio en la exploración de datos, hay eventos que tuvieron lugar el enero, además de que febrero no tiene actividad. De ahí que se decide filtrar la base de datos con una fecha inicial desde el 25 de febrero del 2019, el motivo es porque es un lunes y también la semana en que comienza marzo.

A partir de esa fecha, se crean 17 semanas de igual cantidad de días, cada una es una semana completa de lunes a domingo de inicio de semestre hasta el fin. La definición de las semanas ayuda para analizar el comportamiento según avanza el semestre.

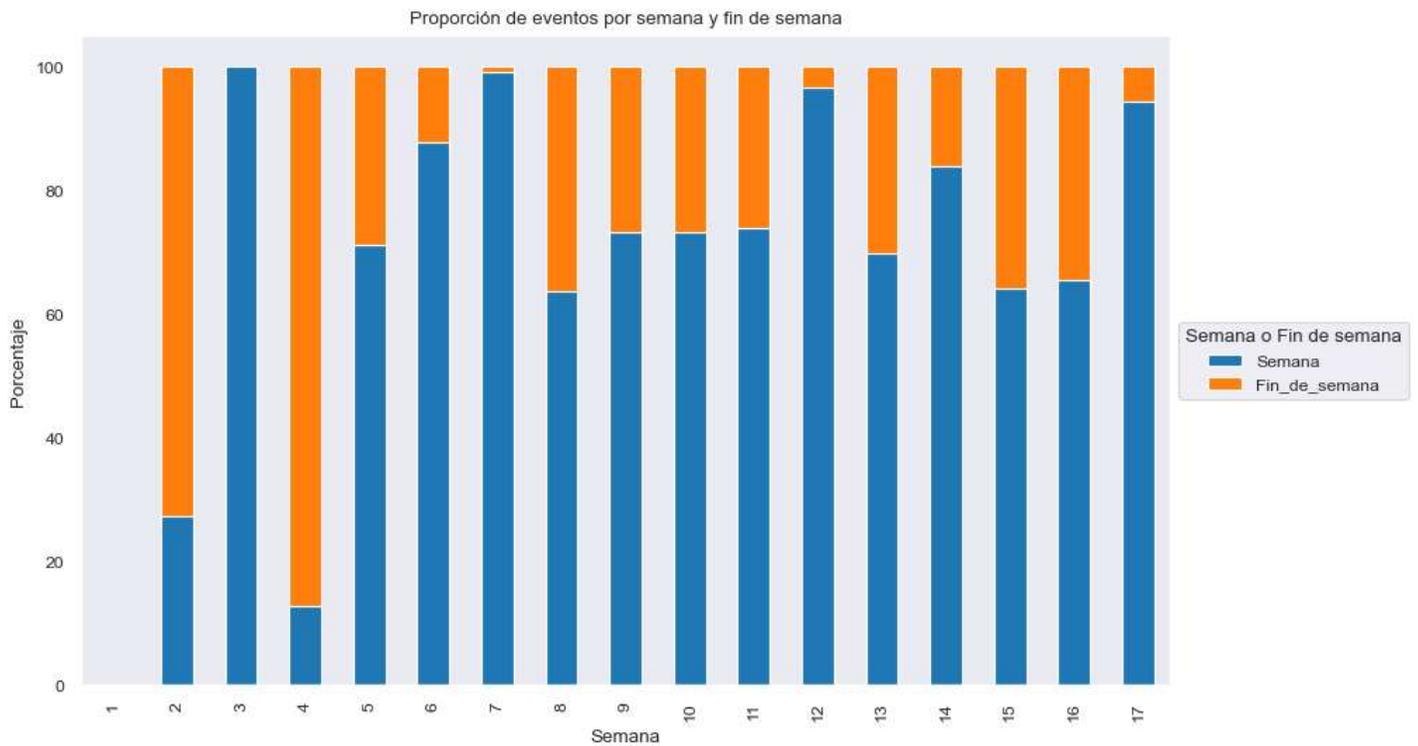
Tabla 5: distribución de semanas de estudio.

Semana	Intervalo
Semana 1	[2019-02-25 , 2019-03-04)
Semana 2	[2019-03-04 , 2019-03-11)
Semana 3	[2019-03-11 , 2019-03-18)
Semana 4	[2019-03-18 , 2019-03-25)
Semana 5	[2019-03-25 , 2019-04-01)
Semana 6	[2019-04-01 , 2019-04-08)
Semana 7	[2019-04-08 , 2019-04-15)
Semana 8	[2019-04-15 , 2019-04-22)
Semana 9	[2019-04-22 , 2019-04-29)
Semana 10	[2019-04-29 , 2019-05-06)
Semana 11	[2019-05-06 , 2019-05-13)
Semana 12	[2019-05-13 , 2019-05-20)
Semana 13	[2019-05-20 , 2019-05-27)
Semana 14	[2019-05-27 , 2019-06-03)
Semana 15	[2019-06-03 , 2019-06-10)
Semana 16	[2019-06-10 , 2019-06-17)
Semana 17	[2019-06-17 , 2019-06-24)

Luego de tener definidas las semanas en la Tabla 5, esta información se integra a la base de datos, se añade la columna “semana” y cada fila tiene un numero entre 1 y 17 que indica la semana a la que pertenece.

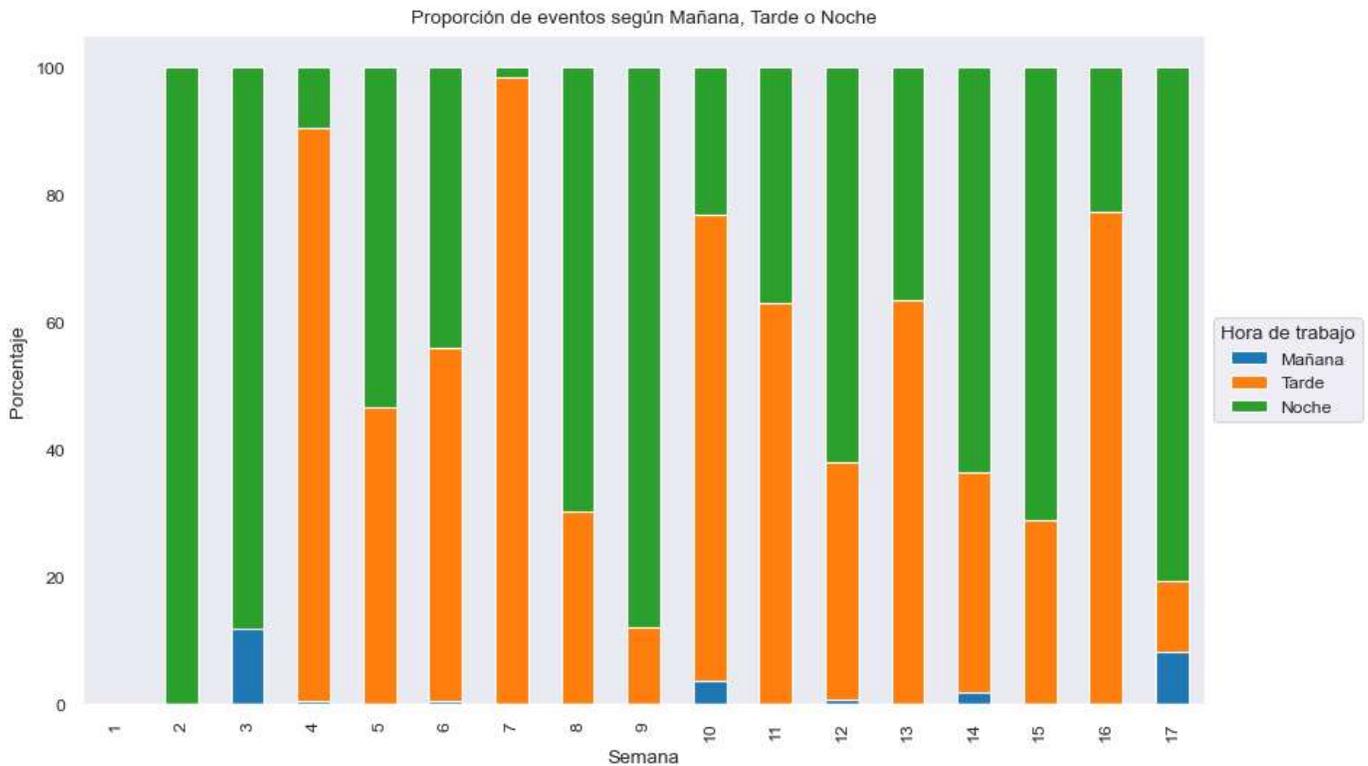
Esta nueva variable nos ayuda por ejemplo a visualizar como varia la cantidad de eventos entre semana o fin de semana a través de las semanas. En la Figura 10, se observa que el trabajo realizado durante la semana es mayor que el trabajo realizado en el fin de semana en 14 de las 17 semanas. Se muestra el porcentaje que representa Semana y Fin_de_semana para cada una de las semanas como proporción del total.

Figura 10: gráfico de barras apiladas al 100% sobre eventos en semana o fin de semana.



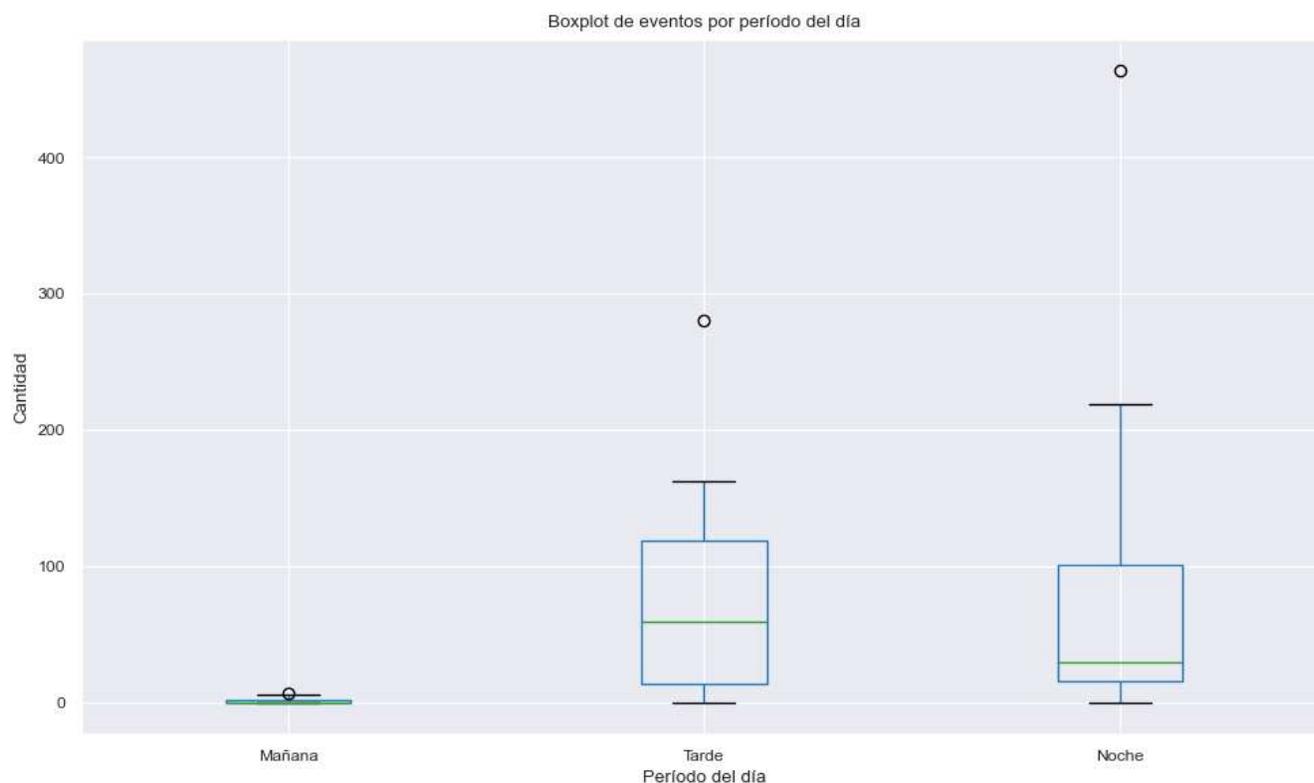
Con respecto al trabajo según la hora del día, se ve en la Figura 11, que la hora favorita para el ingreso a la plataforma es durante la Noche, luego, en segundo lugar, se prefiere el horario de Tarde y por último un horario de Mañana. Esto también se observa en el gráfico agregado de la Figura 8, de ahí que, la información agregada y desagregada es coherente.

Figura 11: gráfico de barras apiladas al 100% según hora de trabajo del día.



El contar la cantidad de registros de eventos que ocurren en las diferentes horas del día y agruparlos por periodo del día, se ve en el *boxplot* de la Figura 11, que los eventos en el periodo de Mañana con los menores, mientras que para Tarde y Noche están en rangos semejantes.

Figura 11: *boxplot* de periodo de trabajo según hora.



4.4.3 Sesiones

La columna "*session*" contiene un código que identifica la sesión de trabajo de una persona en particular. Esto significa que, si una persona realiza varias acciones dentro de una misma sesión de trabajo, todas esas acciones tendrán el mismo código de sesión. Sin embargo, si dos personas diferentes ingresan a la plataforma al mismo tiempo, cada una de ellas tendrá su propio código de sesión, aunque estén accediendo simultáneamente.

Por ejemplo, imagina que Michelle inicia sesión a las 10:00 a.m. y realiza varios eventos dentro de esa sesión, todas esas acciones estarán etiquetadas con el mismo código de sesión, digamos "ABC123". Luego, Hugo inicia sesión también a las 10:00 a.m., pero sus eventos estarán etiquetados con un código de sesión diferente, como "XYZ789". Aunque ambos ingresaron a la plataforma a la misma hora, tienen códigos de sesión distintos porque son usuarios diferentes.

Se utiliza este código junto con el nombre de usuario para saber que número de sesión es, esto implica que cuando una persona ingresa por primera vez a la plataforma, y realiza por ejemplo 4 eventos, se registrara con el mismo nombre de usuario para cada fila, la hora correspondiente de cada evento y tendrán el mismo código de sesión, además, esta será la

sesión número cero de la persona, luego, cuando vuelva a acceder, será la sesión número uno y así en orden correlativo, entre sesión y sesión ocurre que la persona no está activa en la plataforma, o sea, no tiene abierta la página, cuando vuelve a ingresar se traducirá en otra sesión, pudiendo ser dentro del mismo día, misma semana, incluso en la misma hora si llegase a ocurrir que está en EOL, lo cierra y vuelve a entrar, como por ejemplo, reiniciar el dispositivo usado para conectarse, reinicio de internet, etc.

Capítulo 5: Modelo de Markov Oculto

Un Modelo Oculto de Markov (HMM por sus siglas en inglés) es un modelo estadístico que presupone que el sistema que se está modelando sigue un proceso de Markov con parámetros desconocidos. Su propósito principal es identificar estos parámetros ocultos a partir de los parámetros observados. Estos parámetros obtenidos pueden ser utilizados para realizar análisis posteriores, como por ejemplo en aplicaciones de reconocimiento de patrones.

En un modelo de Markov normal, es decir, no el modelo oculto de Markov, se tienen los estados o nodos que son visibles al igual que las probabilidades de transición ubicadas en los arcos. En un modelo oculto de Markov, solo se tiene la información, pero no las probabilidades de transición ni los nodos.

Por ejemplo, no es necesario que le preguntes directamente a una persona si está feliz, triste o enojada. Estos estados emocionales pueden ser inferidos observando variables físicas que son observables, como la posición de las cejas, la respiración, el tono y volumen de la voz, la humedad de los ojos y el contacto visual. Por lo tanto, una persona no tiene que comunicarnos directamente su estado emocional; podemos deducirlo basándonos en estas variables observables. Además, incluso si nunca se ha visto a una persona expresar estas emociones específicas, un modelo oculto de Markov podría identificar, a partir de todas las observaciones, que existen tres tipos distintos de emociones. Luego, corresponde al programador asignarles nombres a estos estados.

5.1 Supuestos

El modelo HMM requiere que se cumplan los siguientes supuestos para poder concluir correctamente: Es necesario que existan estados ocultos, estos estados no son directamente observables, pero se pueden inferir que existen y que representar a un conjunto de características. Deben existir variables medibles y visibles, son las que ayudaran a modelar y encontrar los estados ocultos. La probabilidad de ir de un estado a otro depende únicamente del estado actual y no está condicionado por estados anteriores o la historia completa de su recorrido, además de no depender de sus estados anteriores, no depende de las observaciones en otros estados.

Finalmente, las probabilidades de emisión son constantes a lo largo del tiempo.

5.2 ¿Por qué usar HMM?

El uso de un modelo de Markov oculto (HMM) es beneficioso en diversas aplicaciones debido a su capacidad para modelar secuencias de datos y capturar patrones ocultos.

Los HMM pueden ser útiles para modelar secuencias de datos educativos, como el comportamiento de los estudiantes en plataformas de aprendizaje, la progresión en la comprensión de lectura, o el desarrollo del pensamiento crítico y creativo (Hernández & Soberanes, 2023). Esto se debe a la capacidad de los HMM para manejar y analizar grandes volúmenes de datos complejos.

Asimismo, a partir de los autores se entiende que al aplicar un HMM a datos educativos, se pueden identificar patrones ocultos en el comportamiento de los estudiantes, las interacciones en plataformas de aprendizaje, o el progreso en habilidades específicas. Esta identificación de patrones permite a los educadores y desarrolladores de plataformas de aprendizaje personalizar y mejorar sus métodos y herramientas de enseñanza.

Hernández González & Soberanes (2023) explican que los HMM pueden ayudar a predecir el rendimiento académico, identificar áreas de mejora en la comprensión de lectura o en el pensamiento crítico, y analizar patrones de interacción de los estudiantes con los sistemas de gestión del aprendizaje. Esto facilita la toma de decisiones informadas para optimizar el proceso educativo y apoyar mejor a los estudiantes en su trayectoria de aprendizaje. Gonzaga (2022)

Los modelos de Redes Neuronales, como las Redes Neuronales Recurrentes (RNN) y las Redes Neuronales Convolucionales (CNN), son similares al HMM en el sentido de que también pueden modelar secuencias de datos y capturar dependencias temporales (Chuquival, Galindos, Maquera, Palza & Mamani, 2011), La razón por la que trabajamos con el modelo HMM es que son preferidos sobre las RNN o CNN. Esto se debe a que estos modelos son más interpretables, lo que significa que es más fácil entender y explicar cómo están tomando decisiones. Esto es especialmente útil en el contexto escolar, donde la transparencia es esencial para corregir y mejorar con precisión la estrategia de enseñanza.

Además, en problemas de secuencia con dependencias temporales moderadas, tienden a ser más simples y consumen menos recursos computacionales. Los HMM tienen una trayectoria precisa de modelado de datos educativos como el comportamiento de los estudiantes y la progresión a lo largo de habilidades. Así, crear un modelo basado en HMM tenía sentido debido a su rastro exitoso y garantizado. Todo este historial proporciona una base de confianza y un punto de partida sólido para nuevas investigaciones e implementaciones. A pesar de los requisitos de grandes cantidades de datos que requieren las RNN y las CNN para entrenarse de manera efectiva, los HMM pueden aprender de conjuntos de datos pequeños o medianos, lo cual es un escenario común en la mayoría de las aplicaciones educativas.

También existen las cadenas de Markov, pero debido a que hay información no observable, como el nodo o división a la que pertenecen, no es posible de extraer la utilidad de este modelo sin agregarle la componente oculta. Por otro lado, existe un modelo de Cadenas de Markov Ocultas Generalizadas (G-HMM), es una extensión del HMM que permite una mayor flexibilidad al modelar secuencias con estructuras más complejas y múltiples tipos de observaciones (Moreno, Sukei, Olmos & Artés, 2022) a pesar de la existencia de este otro modelo, está destinado para datos con una mayor complejidad.

Por ello, el modelo HMM es idóneo para el tratamiento de los datos compartidos por la VTI, en la etapa del entendimiento del problema de la metodología CRISP-DM se asigna como el modelo electo para trabajar.

5.3 Validación de supuestos del modelo

a) Existencia de estados oculto

Dado que hay información observable sobre las formas de estudiar, y la exploración de las variables muestra que hay diferentes alternativas para navegar en la plataforma, diferencias como de día, semana, hora, formas de reproducir los videos o, por el contrario, navegar y hacer acciones no relacionadas con los videos, es que se presume que hay cúmulos o estados que representan cierta forma de estudio. Estos estados que están por determinarse más adelante en cantidad y características representan la táctica de estudiar de grupos de personas.

Se realiza un *test* de Chi-cuadrado entre las variables categóricas: {evento, semana}, esto se realiza para saber si existe o no asociación entre estas dos variables, la existencia del tal asociación significa que dichas variables tienen una tendencia o cierto patrón.

La aplicación del *test* da como conclusión que se rechaza la hipótesis nula, esto quiere decir que hay asociación significativa entre las variables, observable mediante el estadístico valor p . Esto se puede interpretar como que, dependiendo de la semana de estudio, se realizan diferentes combinaciones de eventos.

Hay que aclarar, además, que el *test* Chi-cuadrado indica si hay asociación o no entre dos variables categóricas, pero no indica que tipo de asociación existe, el resultado de este *test* se tiene en la Tabla 6.

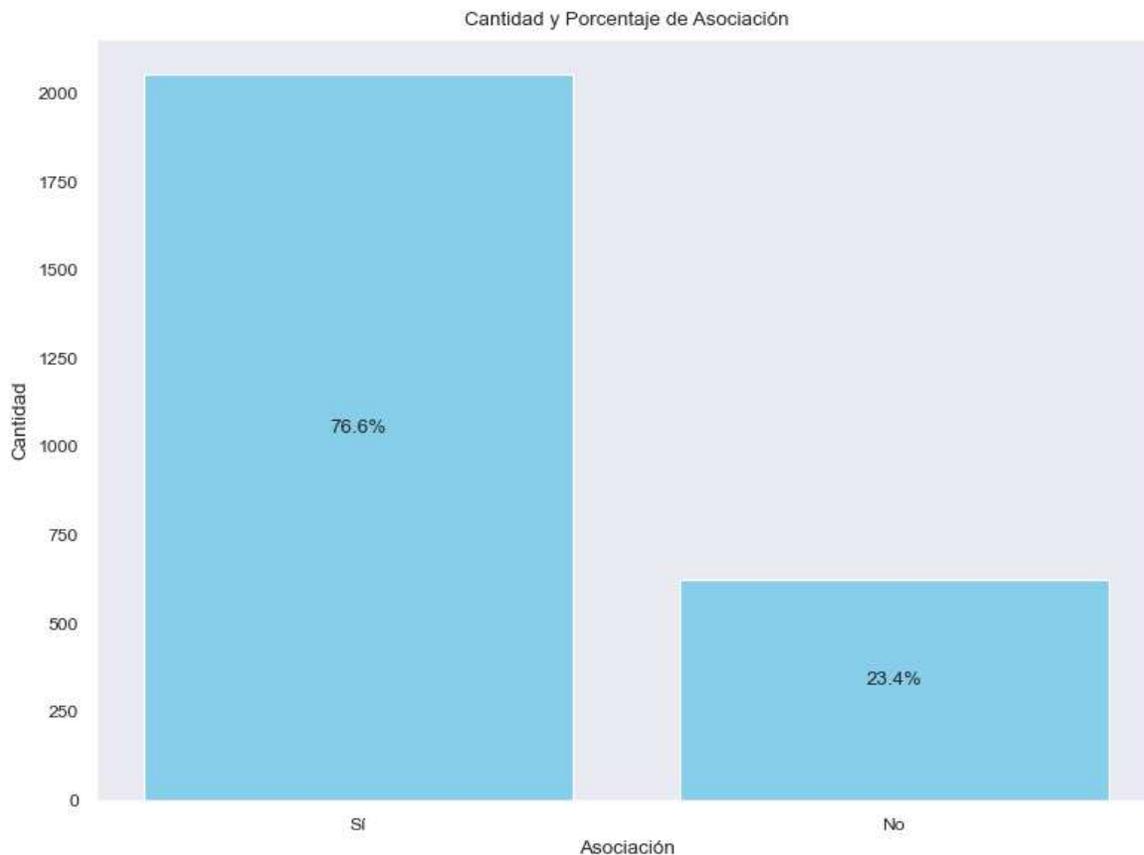
Tabla 6: resultado estadísticas del *test* Chi-cuadrado.

Resultados del test Chi-cuadrado:
Estadísticas de Chi-cuadrado: 4918.89991975048
Valor p: 0.00000
Grados de libertad: 990

La siguiente variante del *test* se realiza persona por persona, se agrupan los datos por 'usuario', para cada grupo de 'usuario' se realiza el test de Chi-cuadrado para las variables 'evento' y 'semana'. Luego, basado en el valor p del *test*, se determina si hay asociación o no. Lo anterior se presenta en la Figura 13 donde se aprecia que el 76.6% de los usuarios tienen una asociación significativa entre las variables evento y semana.

Con respecto al 23.4% de los usuarios que no tienen dicha asociación, también es posible de categorizar con algún método de estudio, ya que el método de estudio va cambiando a través del tiempo.

Figura 13: fracción de usuarios que tiene alta asociación entre evento y semana.



Ya que se muestra la existencia de estados ocultos no observables, se procede a dar por hecho que se cumplen los demás supuestos que son asumidos, como que los datos son finitos, que la probabilidad de transición de un estado a otro no depende de la secuencia completa de los estados anteriores, que las observaciones son independientes entre sí, también se asume que los estados y transiciones serán constantes, es decir que, al ejecutar nuevamente el modelo, dará los mismos resultados.

b) Modelado con HMM

En primer lugar, se crea una lista con todos los eventos existentes en la base de datos después de haber agregado sus respectivos sufijos según el día y hora.

Una vez conocidos los eventos a clasificar, se procede a crear los métodos de estudio, 18 posibles combinaciones que se formulan según indicadores de hora y día.

c) Métodos de estudio según eventos observables

La creación de estos métodos se generó de forma visual y manual, las características de cada evento fueron entendidas y luego clasificadas según sus atributos de hora, día y forma de estudio según una división lógica y lo suficientemente diferentes entre sí como para poder encontrar formas de estudio y analizarlas posteriormente.

Se definen los métodos de estudio, cada método representa a un conjunto de eventos, estos son los mismos eventos a los que se les agregaron los sufijos para categorizarlos por hora del día y día de la semana. Estos métodos son excluyentes entre sí, una persona no puede estar usando 2 o más métodos al mismo tiempo porque dependen de una temporalidad y de acciones, que, si bien puede hacer la misma acción en otro momento, la estaría haciendo en otro tiempo.

Los métodos 1 al 6 tienen relación con lecturas, navegación y cualquier acción que no esté vinculada con videos.

Método 1: personas que realizan los eventos durante la mañana y en la semana.

Método 2: personas que realizan los eventos durante la mañana y en el fin de semana.

Método 3: personas que realizan los eventos durante la tarde y en la semana.

Método 4: personas que realizan los eventos durante la tarde y en el fin de semana.

Método 5: personas que realizan los eventos durante la noche y en la semana.

Método 6: personas que realizan los eventos durante la noche y en la semana.

Los métodos 7 al 12 tienen relación con la visualización de videos, los eventos aquí considerados son la visualización tradicional de videos, es decir, cargar video, reproducir y continuar.

Método 7: personas que ingresan a la plataforma para ver videos de una manera tradicional durante la mañana y en la semana.

Método 8: personas que ingresan a la plataforma para ver videos de una manera tradicional durante la mañana y en el fin de semana.

Método 9: personas que ingresan a la plataforma para ver videos de una manera tradicional durante la tarde y en la semana.

Método 10: personas que ingresan a la plataforma para ver videos de una manera tradicional durante la tarde y en el fin de semana.

Método 11: personas que ingresan a la plataforma para ver videos de una manera tradicional durante la noche y en la semana.

Método 12: personas que ingresan a la plataforma para ver videos de una manera tradicional durante la noche y en el fin de semana.

Finalmente, los métodos 13 al 18 tienen relación con la visualización de videos, pero de una manera no tradicional, es decir, realizando acciones como pausar el video, cambiar la velocidad de reproducción, no verlo desde el principio, sino que verlo desde un minuto en específico saltándose alguna parte o partes, también cambiar de página o volver atrás.

Método 13: personas que ingresan a la plataforma para ver videos de una manera no tradicional durante la mañana y en la semana.

Método 14: personas que ingresan a la plataforma para ver videos de una manera no tradicional durante la mañana y en el fin de semana.

Método 15: personas que ingresan a la plataforma para ver videos de una manera no tradicional durante la tarde y en la semana.

Método 16: personas que ingresan a la plataforma para ver videos de una manera no tradicional durante la tarde y en el fin de semana.

Método 17: personas que ingresan a la plataforma para ver videos de una manera no tradicional durante la noche y en la semana.

Método 18: personas que ingresan a la plataforma para ver videos de una manera no tradicional durante la noche y en el fin de semana.

La elección de estos métodos tiene sentido al comprenderse como la división de las 3 principales acciones que un estudiante puede realizar en la plataforma, Consumir material de lectura, consumir contenido audiovisual de una manera tradicional o consumir contenido audiovisual de una manera no tradicional, donde cada una de estas divisiones se subdivide en el horario del día y si es durante la semana o no.

Luego de tener los métodos, se leen uno a uno los eventos o acciones que realizaron los estudiantes. En base a los métodos definidos, se clasifica cada evento dentro de un método, esto es revisado cuidadosamente mediante programación Python e inspección visual para evitar asignaciones incorrectas, esto se hace en la asignación de métodos.

d) Asignación de métodos

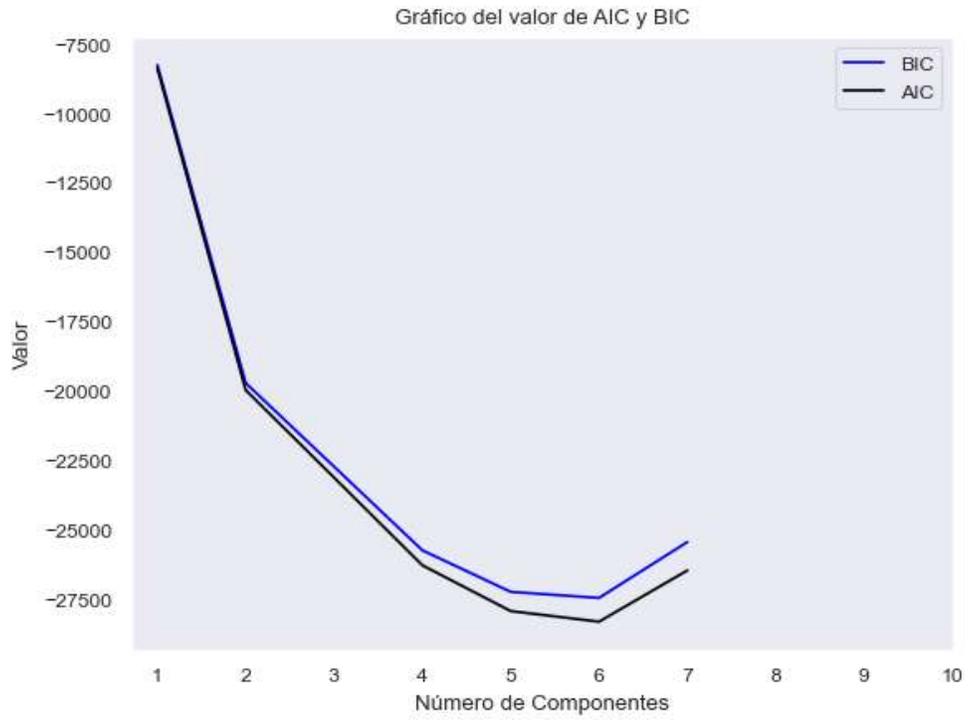
Después de definir los métodos de estudio y los eventos que contiene cada uno de ellos. Se crea una nueva columna llamada: “*action*” en la base de datos original.

Entonces, la instrucción es que, en cada fila, se debe leer el evento que se ejecuta en la plataforma, y luego asignarle el nombre del método que lo contiene, lo mismo se hace para todas las filas, indicando a que método de estudio corresponde el evento que se gatilló.

e) Número de nodos

Para determinar el número de nodos o tácticas de estudios que existen, se utiliza el método del AIC y BIC. Esto se logra ejecutando el modelo HMM en Python y dejando el número de componentes como una variable, dicho número de nodos es relacionado con el AIC y BIC generando un gráfico que permite decidir de manera visual cuál es el número óptimo de estados. Según el grafico en la Figura 14, el menor valor de AIC y BIC se obtiene al escoger 6 componentes, de ahí es posible interpretar que la cantidad de nodos óptima es 6. Es decir, se determinarán 6 tácticas de estudio para el curso analizado.

Figura 14: valor de AIC y BIC según número de nodos.



Capítulo 6: Resultados

Luego de completar con éxito las etapas previas, se obtienen los resultados de la investigación. Se logran encontrar 6 tácticas de estudio que difieren en el día que se estudia, en la hora del día y en la forma que interactúan con los recursos de aprendizaje de lectura y material audiovisual. También se obtiene una matriz de transición entre tácticas que facilita la visualización de evolución de una táctica a otra o la permanencia en la misma, se explora la existencia de un vector de probabilidad estacionario, esto es, la convergencia de las tácticas a medida que avanza el tiempo.

6.1 Tácticas

Luego de escoger el número óptimo de tácticas según las métricas de AIC y BIC, se obtiene una tabla resumen con las Tácticas como filas, que son los estados ocultos no observables y los métodos como columnas. Cada táctica está compuesta por un porcentaje de cada método.

Las tácticas encontradas utilizan 15 de los 18 métodos definidos, el motivo, según las características del modelo, es que en las filas de la base de datos en la columna eventos no eran significativos en comparación con las otras formas de estudiar o navegar en la página.

Para presentar este resumen de tácticas empleadas se eliminan los métodos que fueron usados en porcentajes menores al 5%, esto porque representan acciones o eventos ejecutados escasamente, así, estas celdas no son consideradas. Esta eliminación tuvo lugar solo en las tácticas 1 y 2.

Al prescindir de las entradas menores a 5% la suma de las filas deja de ser 100%, eso es un problema para cuando se deseen realizar más análisis, además de que no cuadrarían los valores, por eso, cada entrada se multiplica por $100 / (\text{suma de la nueva fila})$, esto reescala los valores para que la fila sume 100%, dicho procedimiento se realiza para cada fila donde fueron eliminados elementos.

La tabla definitiva con la información limpia y reescalada se presenta en la Tabla 7.

Tabla 7: tácticas y métodos con porcentajes, limpiado y reescalado.

Táctica	Métodos														
	Método 1	Método 3	Método 4	Método 5	Método 6	Método 9	Método 10	Método 11	Método 12	Método 13	Método 14	Método 15	Método 16	Método 17	Método 18
1	-	6.41	21.80	53.85	6.40	-	-	-	-	-	-	-	11.54	-	-
2	7.85	33.33	-	13.72	-	-	-	-	-	11.76	5.89	15.67	-	5.89	5.89
3			57.32	-	-	-	13.41	-	-	-	-	-	29.27	-	-
4				-	62.86	-	-	-	10	-	-	-	-	-	27.14
5		58.44		-	-	9.09	-	-	-	-	-	32.47	-	-	-
6				60	-	-	-	11.25	-	-	-	-	-	28.75	-

En la Tabla 7 se puede observar la composición de cada una de las 6 tácticas encontradas. La táctica 2 es la que combina más métodos de estudio, siendo 8 en total en diferentes porcentajes. Por el contrario, las tácticas 3, 4, 5 y 6 son las que usan menos combinación de métodos, siendo 3, en diferentes porcentajes cada una mientras que la táctica 1 utiliza 5 métodos.

Luego, de la Tabla 7, se desprende una valiosa información. Cada táctica se analiza junto con los métodos que emplea, se procesa la información y se obtiene como resultado la Tabla 8, la que desglosa cada táctica según su composición del Periodo del día, Parte de la semana y Material.

Tabla 8: resumen de tabla de tácticas y métodos con porcentajes.

Táctica	Periodo del día			Parte de la semana		Material		
	Mañana	Tarde	Noche	Semana	Fin de semana	no video	video tradicional	video no tradicional
1	-	39.75%	60.25%	60.26%	39.74%	88.46%	-	11.54%
2	25.5%	49%	25.5%	54.89%	45.11%	54.9%	-	45.10%
3	-	100%	-	-	100%	57%	13%	29%
4	-	-	100%	-	100%	63%	10%	27%
5	-	100%	-	100%	-	58%	9%	32%
6	-	-	100%	100%	-	60%	11%	29%

Se observa de la Tabla 8 un resumen del comportamiento de las tácticas escogidas por los estudiantes para consumir el material digital de EOL.

Para el periodo del día, se tiene que solo la Táctica 2 tiene un estudio en la mañana, todas las demás tácticas encontradas realizan las actividades en el horario de tarde o de noche. Existiendo tácticas centradas exclusivamente en la tarde o en la noche, como lo son las tácticas, 3 y 5 para la tarde y 4 y 6 para la noche. Los estudiantes que ocupan esas tácticas ingresan a la plataforma únicamente en esos horarios.

Para el periodo de la semana en que ingresan, es relevante notar que 4 de las 6 tácticas ingresan a la plataforma en la semana o en el fin de semana de forma excluyente, es decir, existe un motivo por el cual, 2 grupos de estudiantes solo hacen uso de EOL el fin de semana. La universidad y el equipo docente podría tener en cuenta esta información para planificar las evaluaciones, las pruebas o diferentes actividades, ya que las personas con estudios de fin de semana no podrían enterarse de publicaciones de la semana y viceversa; no se puede esperar que los estudiantes ingresen todos los días ni en todos los periodos del día, un plan de horario programado de subir material y de revisarlo sería beneficioso para la organización. De todas formas, por factores externos, un grupo solo consumirá la información el fin de semana.

Con respecto al material, se observa que siempre la visualización no tradicional tiene más uso que la visualización tradicional, incluso las tácticas 1 y 2 no ven videos de forma

tradicional. Esto podría significar que el material es muy rápido o complejo de entender, de ahí que los estudiantes deben pausar más veces el video o, por el contrario, podría ser muy lento y los estudiantes adelantan la velocidad. Para ambos casos, es necesario conocer la apreciación de los estudiantes sobre el material audiovisual para ajustarse a sus necesidades y que no tengan la necesidad de ajustar los parámetros de la reproducción para adaptarse, aunque es positivo que exista la posibilidad de que cada persona pueda ajustar características de la reproducción para ver la información audiovisual. De aquí se puede desprender la necesidad de conocer cómo les gustaría que fuera el contenido en cuanto a rapidez, explicación, detalles, interacción, etc. para ajustar el contenido a los estudiantes o dar más opciones de ajuste al material para que cada persona pueda consumirlo a su medida.

Finalmente, los estudiantes consumen información en formato texto en mayor proporción que el material audiovisual, esto podría estar condicionado por el objetivo del curso, por sesgos de los estudiantes, por los recursos disponibles, ya que, si hay 8 lecturas y 2 videos, evidentemente el material más consumido será el de lectura por la composición de los recursos, de ahí que se debería contrastar con la cantidad de información de cada tipo y normalizar para poder comparar. Por ello, no se puede concluir que la mayor información consumida es de lectura. Si se puede concluir que el material audiovisual es consumido de una manera no tradicional, lo que no implica que sea perjudicial para el estudiante.

6.2 Matriz de transición

Esencialmente, una matriz de transición es crítica para mostrar la dinámica del sistema en la forma en que se mueve de un estado a otro en el tiempo. A través de esto, es posible prever los próximos estados y analizar cómo se comporta a medida que avanza considerando las probabilidades de transición. Este es un punto vital para nosotros, pues permite comprender sus movimientos y sus potenciales cambios en el futuro. Adicionalmente, proporciona a los interesados predicciones sobre su estado futuro. Esto es crucial para la toma de decisiones y la planificación estratégica. Puede guiar los posibles escenarios y resultados en función de las estadísticas proporcionadas por las probabilidades de transición. Con esta información, las partes interesadas podrán planificar mejor.

Asimismo, la matriz de transición también se utiliza para analizar la estabilidad y la convergencia de un sistema. Permite determinar si un sistema alcanza un estado estable a largo plazo o si experimenta cambios significativos en sus estados a lo largo del tiempo. Este análisis es vital para asegurar que el sistema funcione de manera consistente y predecible.

Finalmente, en aplicaciones prácticas, como en la optimización de procesos o en el control de sistemas, la matriz de transición es crucial para diseñar estrategias eficientes. Permite identificar las transiciones más probables y optimizar el rendimiento del sistema. De esta

manera, se pueden implementar mejoras continuas y ajustes precisos para alcanzar los objetivos deseados.

La matriz de transición resultante muestra la probabilidad de estar en un estado (Táctica) i e ir a un estado (Táctica) j . Es una matriz cuadrada con entradas entre 0 y 1, sus filas suman 1. Las entradas destacadas son el destino con mayor probabilidad de esa fila.

Es importante ya que el siguiente estado solo depende del estado actual, incluso existiendo la probabilidad de mantenerse en el estado por otro periodo de tiempo. Se presenta la matriz de transición en la Figura 15.

Figura 15: matriz de transición entre tácticas de estudio.

		Tácticas					
		1	2	3	4	5	6
Tácticas	1	0,00	0.20	0.20	0,00	0.6	0,00
	2	0.07	0,00	0.13	0,00	0.33	0.47
	3	0,00	0,00	0.54	0.36	0,00	0.10
	4	0.05	0.07	0.16	0.52	0.09	0.11
	5	0.02	0.04	0.02	0.02	0.64	0.26
	6	0,00	0.05	0,00	0.06	0.27	0.62

Cuando una persona comienza usando la táctica 1, en el siguiente periodo se espera que se cambie a la táctica 5, que según la matriz de transición es donde tiene más probabilidades de dirigirse. Si comienza en la táctica 2, se espera que transite hacia la táctica 6. Para las tácticas 3, 4, 5 y 6, se espera que se mantengan en dicha táctica ya que la mayor probabilidad es la de permanecer en ese estado.

Lo anterior no implica que no pueda cambiarse a una táctica diferente a la de mayor probabilidad de la fila correspondiente, de ahí que son probabilidades.

De esta matriz de transición se observa el cambio para el siguiente periodo. Para conocer donde estará en 2 o más periodos, se deben usar las propiedades de Markov que dice que se debe tomar la potencia de la matriz y así saber estadísticamente cómo evolucionan los estados en que se encuentran en este caso los estudiantes según la táctica con la que hayan iniciado.

6.3 Vector de probabilidad estacionario

Dado que una matriz de transición es la evolución de un estado a través de tiempo, es plausible considerar la existencia de un comportamiento del largo plazo, e este caso, se llama vector de probabilidad estacionaria. Para encontrarlo, se debe asegurar su existencia mediante el cumplimiento de supuestos.

Supuestos

1. Una única clase recurrente

También conocida como una única clase recurrente cerrada, es un concepto utilizado en la teoría de Markov y en la teoría de grafos para describir un conjunto de estados en un sistema que está interconectado de tal manera que una vez que se entra en cualquiera de los estados de la clase, es posible moverse a cualquier otro estado dentro de la misma clase con una cierta probabilidad en un número finito de pasos.

Para verificar este supuesto se utiliza la librería NetworkX de Python, donde Peñafiel, Burgos, Zavala & Solís (2023) explican que el creador de NetworkX, una biblioteca de Python para el análisis de redes es el Laboratorio de Investigación en Ciencias de la Computación (LINC) de la Universidad de California en Los Ángeles (UCLA), que toma todos los pares de nodos de la matriz de transición y verifica si existe un camino finito para llegar de un nodo a otro.

En este caso, la matriz de transición cumple el supuesto de ser una única clase recurrente.

2. Aperiodicidad

Cuando una matriz de transición es aperiódica, significa que no hay un patrón regular o predecible en la secuencia de estados que el sistema atraviesa. En otras palabras, no hay un período fijo en el que el sistema se repita en un ciclo constante. Esta propiedad es fundamental porque asegura que el sistema pueda alcanzar cualquier estado desde cualquier otro estado en un número finito de pasos, lo que implica una mayor flexibilidad y complejidad en el comportamiento del sistema a lo largo del tiempo.

Para verificar este supuesto, se determinan longitudes de los ciclos en los que se podría entrar, para cada uno de ellos se calcula el máximo común divisor (MCD). Si el MCD de las longitudes de los ciclos para cada estado es igual a 1, significa que no hay un período común entre los ciclos para ningún estado. Esto sugiere que el sistema no se repite en un patrón

periódico y, por lo tanto, es aperiódico. Si, en cambio, para al menos un estado, el MCD no es igual a 1, indica que existe un período común entre los ciclos para al menos un estado, lo que sugiere que el sistema podría repetirse en un patrón periódico y, por lo tanto, no es aperiódico.

De la matriz de transición, se concluye usando el método del MCD que la matriz es aperiódica.

Al cumplir ambas condiciones, se concluye que existe un vector de probabilidad estacionario.

3. Probabilidad estacionaria

Se obtienen los valores y vectores propios de la matriz de transición traspuesta, los valores propios son los valores de las probabilidades estacionarias.

4. Vector de probabilidad estacionaria

[0.0166561, 0.04516968, 0.07874141, 0.12130289, 0.36903639, 0.36909353]

Este vector se interpreta como que en el largo plazo, sin importar en el nodo (táctica) que se comience, se convergerá al nodo i con la probabilidad en la posición i , donde $i \in \{1, 2, 3, 4, 5, 6\}$.

5. Evolución de matriz de transición

Se toma la matriz de transición y se analiza cómo es su evolución a medida que avanza el tiempo de una manera discreta, la matriz obtenida es el estado inicial, y el vector de probabilidad estacionaria es la evolución de la matriz en el largo plazo, a pesar de ello, entre estos periodos, existen pasos intermedios.

Para visualizarlo, la matriz se eleva a n , donde $n \in \{1, 2, 3, \dots, 30\}$, luego se crean 36 listas, llamadas $listaij$, donde $i \in \{1,2,3,4,5,6\}$ y $j \in \{1,2,3,4,5,6\}$. $listaij$ contiene los valores de la matriz en la posición ij de las 30 matrices resultantes de la potenciación. Estas listas son graficadas y muestran la evolución de las probabilidades de transición del estado i al estado j según avance de paso.

Para ilustrar un ejemplo, se toma como referencia la imagen del Anexo A. Aquí se ve como evoluciona la probabilidad de llegar a las tácticas 1 a 6 comenzando de la 1. En el paso 0 (eje de la abscisa) se observa que comenzar con la táctica 1 y continuar con la misma táctica en el paso siguiente tienen una probabilidad de 0%, comenzar con la táctica 1 y cambiar a la táctica 4 o 6 también es 0% probable, por otro lado, comenzar en esta táctica y cambiar a la táctica 2 o 3 en el siguiente paso tiene una probabilidad de 20%, por último, cambiar a la táctica 5 es lo más probable con un valor de 60%. A medida que transcurre el tiempo o los pasos, estas probabilidades se modifican, observándose que, en el largo plazo, las tácticas de destino más probables son la 5 y 6. Esta información se extrae de la matriz de transición.

El análisis de los demás anexos es análogo, los gráficos de estas evoluciones se encuentran en los anexos A al L.

Del anexo A al F, se observa cómo cambia la probabilidad de comenzar en un estado específico y terminar en otro a medida que transcurren los pasos. Se observa que sin importar en el estado que se comience, se converge a los estados 5 y 6 con igual probabilidad considerando una aproximación por redondeo con dos decimales, es decir, 36.90%, esta probabilidad a la que se converge es exactamente la misma que se obtiene mediante el vector de probabilidad estacionaria.

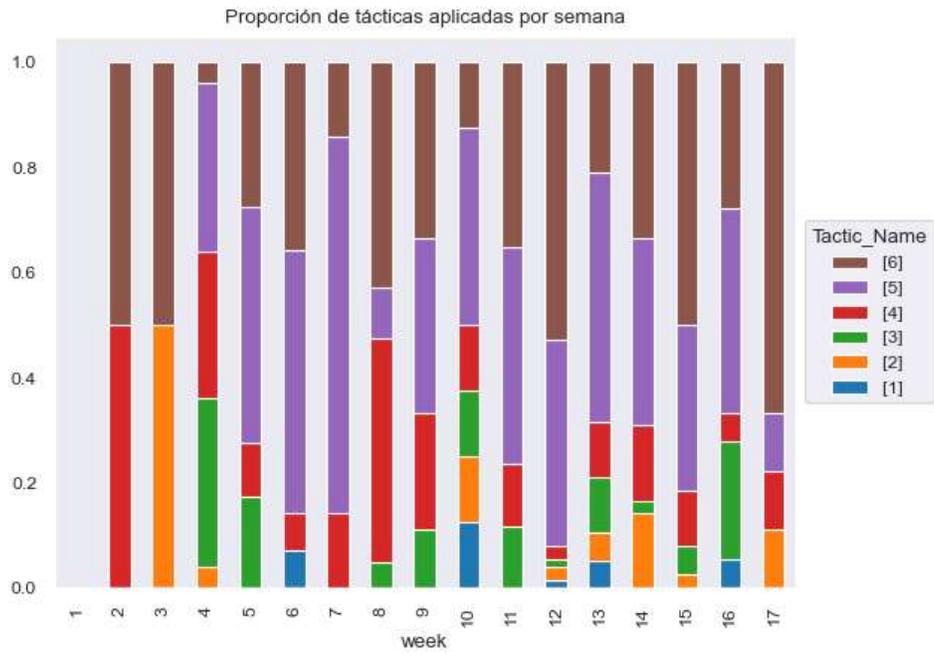
De los anexos G al L, se observa el proceso inverso, se muestra la evolución de comenzar en cualquier estado y la convergencia hacia un estado en particular. Si se comienza en cualquiera de los 6 estados, se ve la convergencia de llegar al estado j , además, se observa que en el anexo G, anexo H, anexo I, anexo J, anexo K, anexo L, las curvas convergen a: 1.7%, 4.5%, 7.9%, 12.1%, 36.9%, 36.9%, respectivamente. Que es exactamente el resultado del vector de probabilidad estacionaria.

6. Tácticas por semana

Semana a semana, los estudiantes ingresan a la plataforma, a medida que se integran con sus pares pueden intercambiar información o métodos de estudio, así mismo, cada estudiante puede observar sus resultados y sus objetivos para cambiar el método de estudio o mantenerlo. Es ser objetivos el pensar que distintos estudiantes tienen diferentes metas de aprendizaje y de calificaciones, por lo que las formas de estudiar reflejan intereses personales de cada estudiante, intereses que son desconocidos, al menos no se pueden obtener mediante los actuales datos de EOL.

En la Figura 16, se presenta un gráfico de barras apiladas al 100% con la proporción de tácticas según la semana. Es claro que la táctica 5 y 6 son las más aplicadas. Con la ayuda de la matriz de transición se puede conocer la probabilidad de estar en la táctica i y cambiar a la táctica j .

Figura 16: proporción de tácticas empleadas por semana.



Capítulo 7: Discusión

Se satisface la Hipótesis 1 que plantea que un modelo HMM permite descubrir patrones de navegación ocultos de los estudiantes con los datos disponibles antes expuestos, además, estas tácticas fueron clasificadas satisfactoriamente, las que representaban formas de estudiar dependientes del día, de la hora y de las acciones llevadas a cabo.

Se vio en la tabla de resumen, Tabla 9 y Tabla 10, que existen periodos del día que son representados por tácticas diferentes, y que también estas tácticas cambian según el transcurso de las semanas, visto en Figura 10. Estos resultados confirman la Hipótesis 2, que sostiene que existen horarios diferenciados para cada táctica y varían entre semanas.

Para la Hipótesis 3, se confirma que existe no uno, sino dos tácticas más usadas por los estudiantes y que se mantienen en el tiempo, incluso que las diferentes formas de estudiar convergen a ellas. Estas son las tácticas 5 y 6, las que coinciden en que el estudio se realiza durante días hábiles, alrededor del 60% del contenido con el que interactúan no está relacionado a video, y cuando lo hacen, lo consumen de una manera no tradicional. Estas tácticas difieren en el periodo del día, la táctica 5 es en un horario de tarde, mientras que la táctica 6 se despliega en un horario de noche. Estas tácticas ocurren en el largo plazo, aunque en un semestre, el largo plazo podría considerar después de 10 semanas por el hecho de que ya se habrían ejecutado múltiples inicios de sesión.

En añadidura de lo anterior, se responden las preguntas planteadas en el ítem de motivación y se procede a responder. No, no todos los estudiantes aprenden igual, no todos los estudiantes estudian igual, esto a pesar de que están en el mismo curso con el mismo material disponible, acceden, lo manipulan y lo reciben de diferentes maneras. Para las formas de aprender, según los datos proporcionados por EOL, son las 6 tácticas encontradas. Seguramente, en otros cursos, con otros requisitos, existirán otras tácticas o formas de estudiar.

7.1 Aprendizajes del presente estudio

Uno de los principales aprendizajes derivados de esta investigación es la comprensión de la amplia gama de tácticas de estudio empleadas por los estudiantes. Desde estudiar en diferentes momentos del día hasta interactuar de manera única con el material educativo, este estudio ha revelado la diversidad de enfoques que los estudiantes adoptan para aprender o para cumplir con las tareas del curso.

Observar cómo los estudiantes combinan, ajustan y cambian sus métodos de estudio resalta la importancia de la flexibilidad en el proceso educativo. Este aprendizaje destaca la necesidad de adaptarse a las circunstancias individuales y a las demandas académicas en constante evolución, es un hecho que los métodos de estudios no son una regla para todos y todas, cada persona tiene sus formas personales de interiorizar mejor el contenido.

La forma en que los estudiantes interactúan con el material educativo, especialmente con los videos, refleja un uso creativo y adaptativo de los recursos disponibles. Este aprendizaje destaca la importancia de proporcionar acceso a una variedad de herramientas en complemento con las tradicionales, quiz, puzzles, juegos, trabajo en equipo, aula invertida, investigación, mini documentales, entre otras tantas posibilidades, esto por el hecho de que hay diversas formas de aprender y de trabajar, y por qué no, tal vez las personas que reprueba una asignatura en particular es porque el material disponible no se adaptó a las necesidades de aprendizaje de la persona.

En este caso las tácticas o formas de aprendizajes convergen a dos, que además son similares, solo difieren en el periodo del día en que se realizan, siendo la táctica 5 en la tarde y la táctica 6 en la noche, esto no implica que sean métodos efectivos o que sean respetuosos con la salud mental, solo se puede entender que son métodos de estudio y estadísticamente se terminan usando en mayor medida que los demás.

Por último, entender el comportamiento de una muestra mediante datos objetivos es diferente que hacer el estudio en base a datos auto informados mediante encuestas, entrevistas o test porque los datos objetivos muestran la realidad que a veces las personas pueden olvidar, confundir o darla a conocer con sesgos personales o del entorno.

7.2 Otras aplicaciones

Este trabajo puede ser extendido a otras disciplinas o industrias, esto por el hecho de que, al tener variables observables de una persona, empresa o grupo de interés, es posible encontrar estados ocultos no observables y así clasificarlos, evaluarlos, potenciarlos o reducirlos, a

continuación, se presentan algunos ejemplos en los que el modelo HMM podría ser de utilidad:

Tácticas de enseñanza:

El modelo HMM se puede utilizar en la enseñanza, ya vimos que ayuda a entender el comportamiento de estudio, aunque se puede expandir aún más con métricas de origen presencial y las calificaciones como, por ejemplo:

1. **Participación en el aula virtual:** Interacción con materiales de curso, participación en discusiones, envío de tareas, etc.
2. **Rendimiento académico:** Notas, calificaciones en exámenes, participación en actividades extracurriculares.
3. **Comportamiento de estudio:** Tiempo dedicado al estudio, métodos de estudio utilizados, horarios de estudio.
4. **Retroalimentación del profesor:** Comentarios sobre el desempeño del estudiante, áreas de mejora identificadas por el profesor.

Mediante el uso de HMM, las instituciones educativas podrían identificar patrones de comportamiento de los estudiantes que indiquen niveles de compromiso, comprensión del material o incluso posibles problemas académicos. Esto podría ayudar a los educadores a adaptar sus enfoques de enseñanza, brindar apoyo personalizado a los estudiantes y mejorar la efectividad del proceso educativo.

Para las ventas, el modelo HMM (modelos ocultos de Markov) puede ayudar a las empresas a descubrir qué les gusta comprar a los clientes observando sus compras anteriores, el uso de sitios web o aplicaciones, la edad, el sexo y la actividad en las redes sociales. Esto permite modificar los aspectos de marketing y ventas para generar más ventas y mantener contentos a los clientes al descubrir qué quieren o no quieren comprar o la secuencia en la que compran, también se podría personalizar la experiencia de venta en la página web o las redes sociales.

También puede contribuir a estrategias de campañas políticas. Conociendo como interactúan en redes sociales: debates, compartir, comentar, encuestas y sondeos de opinión, lo que la gente piensa sobre la política, como quién les gusta, qué les preocupa, por quién quieren votar.

Si se usa HMM en política, se puede descubrir cómo vota o actúa la gente, si les gustan diferentes cosas, si van a eventos políticos o si hablan en línea. Los partidos políticos podrían mejorar sus planes de campaña, dividir a los votantes en grupos más pequeños y enviarles mensajes que sean adecuados para ellos.

Dentro de otras aplicaciones esta la negociación, conociendo el comportamiento de las personas dentro de una negociación como las estrategias usadas, el tono de voz, el lenguaje corporal, el historial de otras negociaciones que hayan resultado exitosas, se podría segmentar en casos exitosos y no exitosos para luego implementar o fomentar las actitudes que llevaron a la aceptación de la negociación.

Esto podría ayudar a identificar patrones de comportamiento que conduzcan a acuerdos exitosos, así como a anticipar posibles obstáculos o conflictos. Esto podría permitir a las empresas desarrollar estrategias de negociación más efectivas, mejorar la gestión de riesgos y aumentar la probabilidad de alcanzar acuerdos mutuamente beneficiosos.

De lo anterior, es posible concluir que este modelo se puede aplicar en diversas áreas entregando resultados relevantes para la toma de decisiones.

Capítulo 8: Conclusiones

El enfoque principal de mi investigación es descubrir las diversas tácticas de estudio empleadas por los estudiantes. Para lograr esto, utilicé una base de datos que incluye columnas como nombre, tipo de evento, fecha con día, hora, segundo, mes, y la cantidad de semana. Esta información me permitió clasificar los eventos según si ocurrieron durante la semana (lunes a viernes) o durante el fin de semana (sábado y domingo), así como también categorizarlos en períodos del día: Mañana, Tarde y Noche.

Luego, utilicé estos datos para entrenar un modelo de Markov oculto, asegurándome previamente de que el conjunto de datos cumpliera con los supuestos del modelo. Este modelo me permitió identificar los diferentes métodos de estudio utilizados por los estudiantes, así como las probabilidades de transición entre ellos, lo que se tradujo en una matriz de transición. Analizando esta matriz y tomando potencia de la misma, pude visualizar cómo evolucionan las tácticas de estudio a lo largo del tiempo. Además, identifiqué el vector de probabilidad estacionaria, indicando las tácticas de estudio más frecuentes y estables.

Mis hallazgos revelaron que las tácticas de estudio más comunes incluyen el estudio en horarios vespertinos o nocturnos durante la semana, y una tendencia a trabajar más durante la semana. También se observó que los estudiantes tienden a consumir material de estudio no relacionado con videos, o a hacerlo de manera no convencional, como modificando la velocidad de reproducción o saltando partes del contenido. Además, se evidenció que los estudiantes tienen la capacidad de adaptar y combinar sus métodos de estudio según sus necesidades y restricciones personales.

Estos resultados se presentan en un gráfico de barras apiladas, que muestra cómo cambian las tácticas de estudio semana a semana. En resumen, mi investigación concluye que hay dos métodos de estudio principales hacia los cuales convergen los estudiantes: el estudio en la plataforma educativa durante las tardes o noches de la semana, y el uso de materiales de estudio no relacionados con videos o con un enfoque no tradicional hacia los videos.

8.1 Trabajo futuro

Después de obtener y analizar los resultados correspondientes, se propone como trabajo futuro para extender esta investigación, explorar la relación entre las tácticas de estudio identificadas y otros factores relevantes, tales como el rendimiento académico, el grado de satisfacción del estudiante y el impacto en la retención estudiantil. La investigación destaca la relación entre los estilos de aprendizaje y el rendimiento académico de los estudiantes universitarios. Identificar y comprender los estilos de aprendizaje de los estudiantes puede influir directamente en su desempeño académico (Paba, Lara & Palmezano, 2008). Por ello es crucial reconocer que el simple hecho de que una táctica sea utilizada con frecuencia no garantiza un aprendizaje efectivo o saludable. Además, el comportamiento de navegación antes o después de una evaluación puede variar, lo que sugiere que los estudiantes pueden sentirse cómodos con su método de estudio o estar explorando otras formas de aprendizaje.

Es esencial determinar si estos métodos de estudio están alineados con las expectativas del equipo docente y de la universidad. Si los estudiantes están consumiendo el material en horarios, días o formas diferentes a los esperados, podría indicar que los incentivos no están alineados con las prácticas educativas tradicionales. Los hábitos de estudio de los estudiantes, como el tiempo dedicado al estudio y la asistencia deben ser evaluados para saber si están alineados con las expectativas institucionales y si tienen una relación positiva con el rendimiento académico (Quezada & Elizabeth, 2015). Por lo tanto, es necesario un análisis detallado de estas tácticas por equipos multidisciplinarios para identificar qué métodos de estudio son saludables y beneficiosos, o al menos no perjudiciales para los estudiantes en aspectos como los ciclos del sueño, los descansos y las actividades extracurriculares, entre otras variables.

Este análisis permitiría implementar conferencias, foros o debates sobre los métodos de estudio, con el fin de determinar cuáles deben ser reducidos, mantenidos o potenciados para promover un aprendizaje saludable entre los estudiantes, previniendo así posibles conflictos de salud mental, estrés o atrasos en las actividades académicas.

Otro aspecto por considerar es la implementación de métricas en la plataforma U-cursos, con visualización tanto para el equipo docente como para los estudiantes, para que puedan evaluar cómo están estudiando y qué resultados están obteniendo con esos métodos. Además, sería relevante conocer qué métodos aplican los estudiantes con las mejores calificaciones y cuáles deben evitarse, ya que pueden estar correlacionados con el fracaso académico o futuros conflictos.

Finalmente, se sugiere analizar la evolución de los métodos de estudio a lo largo de la historia de un curso, comparando las versiones de cada semestre o evaluando cómo cambian según el curso o la malla curricular. Comparar los métodos de estudio de los estudiantes semestre a semestre y a lo largo de su carrera académica proporcionaría información valiosa sobre la

efectividad y relevancia de dichos métodos en relación con sus calificaciones. Esto, a su vez, podría incitar a la universidad a considerar iniciativas para promover cambios en este ámbito.

8.2 Sesgo y limitaciones

Este estudio presenta varios sesgos que podrían influir en la interpretación de los resultados.

Un sesgo significativo está relacionado con la ubicación geográfica de la universidad. Dado que esta institución se encuentra en la capital del país, la mayoría de los estudiantes provienen de esa área, lo que puede resultar en métodos de estudio que difieren notablemente de aquellos utilizados por estudiantes de regiones más remotas o extremas del país.

Otro sesgo importante es la homogeneidad en cuanto a la edad y la especialidad de los participantes. Todos los estudiantes incluidos en el estudio tienen entre 18 y 23 años y están matriculados en la Facultad de Ciencias Físicas y Matemáticas, donde se enfoca en la enseñanza de ingeniería y ciencias. Esto puede limitar la generalización de los resultados a estudiantes de otras edades o de diferentes facultades y carreras.

Además, la Universidad de Chile es una institución pública que requiere altos puntajes en una prueba de selección rigurosa para el ingreso. Esto sugiere que los estudiantes admitidos pueden tener un nivel base diferente en comparación con aquellos de otras universidades que no imponen requisitos de ingreso tan estrictos.

Una limitación adicional del estudio se refiere a los datos recopilados. Dado que el año 2019 marcó la implementación inicial de la plataforma EOL, es posible que los datos obtenidos reflejen un uso más limitado o una funcionalidad menos desarrollada en comparación con la situación actual. Por lo tanto, para una evaluación más precisa y relevante, se recomienda replicar el estudio con datos más recientes, lo que permitiría considerar las nuevas funciones y actualizaciones incorporadas en la plataforma de estudio.

Bibliografía

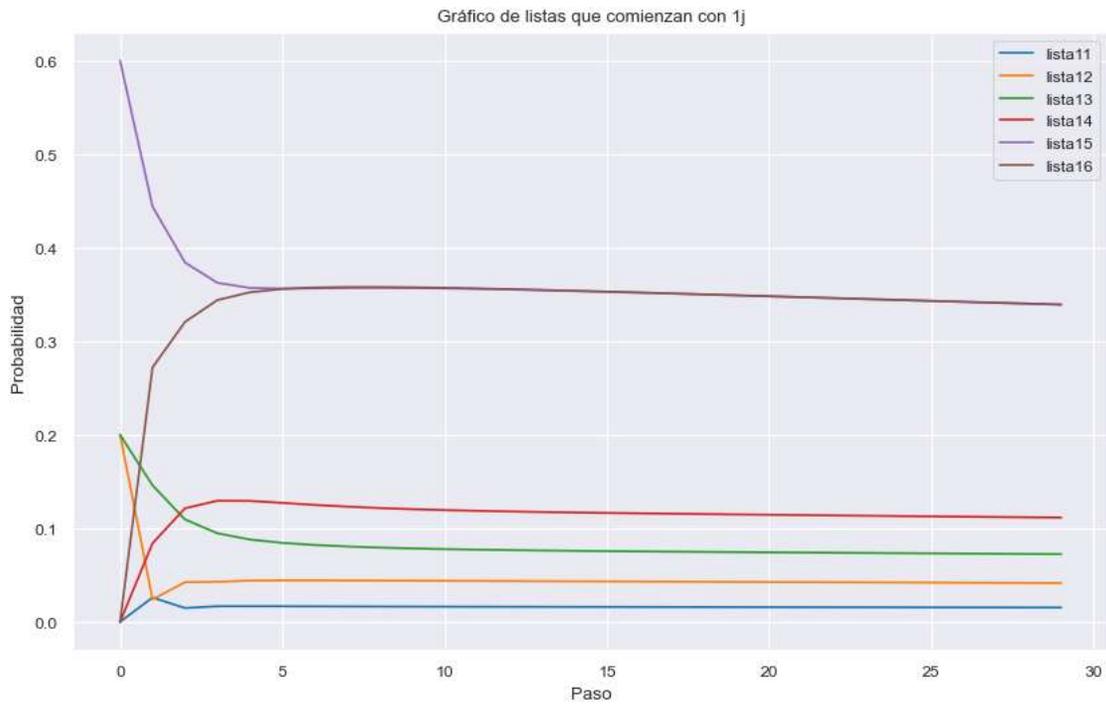
- Acevedo, L., & Andrés, C.G. (2019). Propuesta Pedagógica para Desarrollar Competencias de Investigación en el Manejo de la Información y el Uso de Bases de Datos Especializadas en Estudiantes Universitarios.
- Aguirre Castillo, C. (2016). Superación académica en primer año de Ingeniería y Ciencias: mecanismos de permanencia y mejoramiento académico. Disponible en <https://repositorio.uchile.cl/handle/2250/139152>
- Arregui, D.A., Godoy Mena, M.J., & Marrero Fernández, A. (2023). Análisis de la Ley de Educación Superior de Argentina en relación a las tendencias actuales. *Revista San Gregorio*.
- Baltazar Vilchis, C.A., Martínez Garduño, Y., & Sámano Ángeles, A. (2021). La importancia de la analítica de datos en el seguimiento a estudiantes para el logro de certificaciones profesionales de TI: Estudio de caso. *Tecnología Educativa Revista CONAIC*.
- Bramucci, R., y Gaston, J. (2012). Sherpa. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12 (May), 82.
- Catrileo Herrera, F. (2017). Explorando la relación entre la actividad en línea y el rendimiento académico de estudiantes en primer año de ingeniería. Disponible en <https://repositorio.uchile.cl/handle/2250/146713>
- Celis, S., Moreno, L., Poblete, P., Villanueva, J., y Weber, R. (2015). Un modelo analítico para la predicción del rendimiento académico de estudiantes de ingeniería. *Revista Ingeniería de Sistemas*, 29 (1), 5-24.
- Consuegra, D., Sucre, A.J., & Bustamante, E. (2023). uso de la tecnología en la educación superior en el Centro Regional Universitario de Los Santos. *Cuadernos Nacionales*.
- Delgado Batista, C.A., Sánchez Rodríguez, M., & Ramos Sánchez, E. (2023). IMPLEMENTACIÓN DE UNA PLATAFORMA VIRTUAL EDUCATIVA PARA EL CENTRO EDUCATIVO BILINGÜE FEDERICO ZÚÑIGA FELÍU. *Actas del II Congreso de Creatividad e Innovación en Educación (CIE-2023)*.
- Espinoza Gallardo, A.C., Martínez-Vázquez, Y.V., Zepeda-Salvador, A.P., Martínez-Moreno, A.G., & Vázquez-Cisneros, L.C. (2023). Uso de pantalla y duración de sueño en estudiantes universitarios. *Journal of Behavior and Feeding*.
- Fernández, E.G., Ignacio, C.J., Rodríguez, H., & Pérez, G. (2014). The integration of educational influences of the university and the role of the family in the professional formation of higher education.

- Galindo, L.M., Reyes Martínez, P.G., & González, F. (2022). Escenarios para la transición energética a una economía carbono neutral en América Latina y el Caribe: algunos hechos estilizados. *Sobre México Temas de Economía*.
- Godoy, V.L., Morales, K.F., & Pulido, J.E. (2018). La actitud hacia la educación en línea en estudiantes universitarios. *Revista de Investigación Educativa*.
- Gómez-Calcerrada, S.G., Hernández, E.S., & González, J.M. (2015). Una aproximación a la realidad de las plataformas virtuales de las universidades españolas: el primer reto para una educación personalizada en personas con discapacidad motórica o visual.
- Gómez Miranda, P., & Jiménez García, M. (2022). Rendimiento académico de estudiantes universitarios al final de la transición de la educación presencial a la educación en línea por el covid-19. *RIDE Revista Iberoamericana para la Investigación y el Desarrollo Educativo*.
- Gozalo, M., León-del-Barco, B., & Mendo-Lázaro, S. (2020). Good Practices and Learning Strategies of Undergraduate University Students. *International Journal of Environmental Research and Public Health*, 17. <https://doi.org/10.3390/ijerph17061849>.
- Hernández González, L., & Soberanes Martin, A. (2023). Modelo de obtención de datos de los principales Sistemas de Gestión del Aprendizaje. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*.
- Ibarra, M. d. C., y Michalus, J. C. (2010). Análisis Del Rendimiento Academico Mediante Un Modelo Logit * Academic Performance Analysis Through a Logit Model. *Revista Ingeniería Industrial*, 47-56.
- Ibarra de Leguizamon, A.C. (2023). Factores Que Intervienen En La Formación De Hábitos De Estudio En Estudiantes De La Carrera Licenciatura En Educación Escolar Básica. *Ciencia Latina Revista Científica Multidisciplinar*.
- López, B.G., Félix, E.G., Bertomeu, I.M., & Garcés, A.B. (2015). Métodos innovadores y enfoques de aprendizaje en estudiantes universitarios.
- López, B.G., Rodríguez, M.A., Hervás, N.M., Beut, J.A., & Pérez, C.P. (2017). MÉTODOS CENTRADOS EN EL APRENDIZAJE, IMPLICACIÓN DEL ALUMNO Y PERCEPCIÓN DEL CONTEXTO DE APRENDIZAJE EN ESTUDIANTES UNIVERSITARIOS. *Educacion Xx1*, 20, 161-187.
- López, B.G., Esteban, P.R., Mateo, P.M., Peleato, I.V., & Rodríguez, M.A. (2015). Métodos centrados en el aprendizaje, estrategias y enfoques de aprendizaje en estudiantes universitarios. *Revista De Educacion*, 229-254.

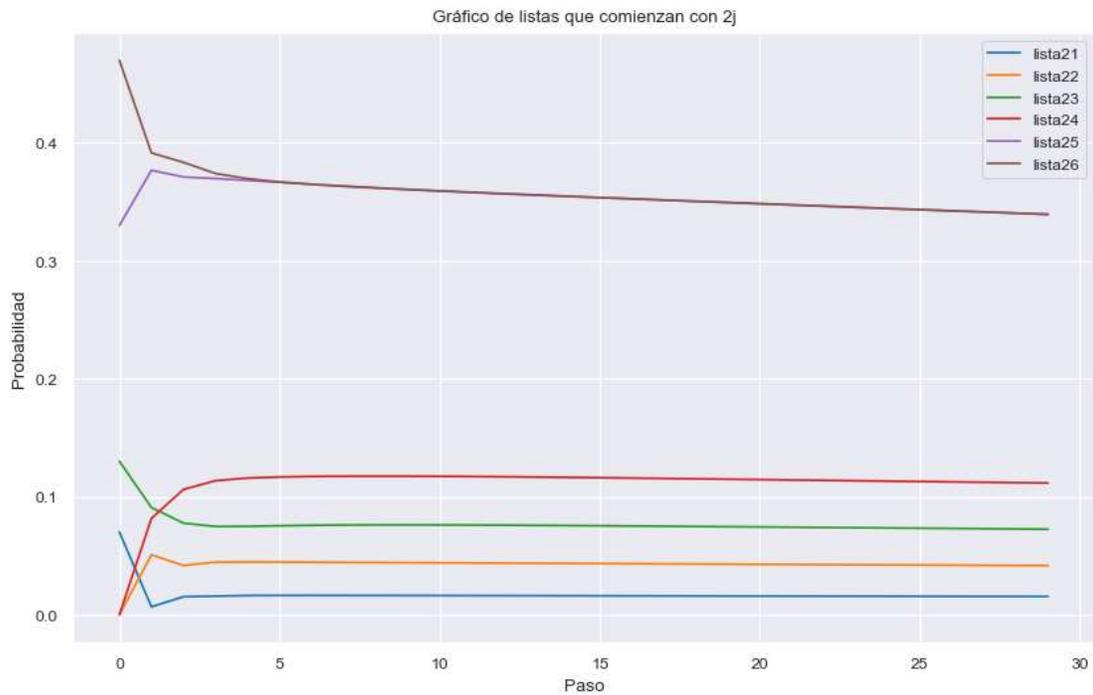
- López Lara, Y., & De la Fuente Valdez, S.M. (2023). Comunicación y clima académico en alumnos de primer ingreso. Caso de estudio: Universidad Autónoma De Nuevo León. *Multidisciplinas de la Ingeniería*.
- Lugo, G.M., López, R.I., Balderrama, J.A., & Ballesteros, L.C. (2017). Identificación automática de estilos de aprendizaje en estudiantes de educación superior. *Educación Superior*.
- Medeiros, E.S., & Leandro, R.A. (2019). Modelo oculto de Markov para imputação de genótipos de marcadores moleculares. *REVISTA BRASILEIRA DE BIOMETRIA*.
- Muñoz Díaz, J. (2017). Estudio de la relación entre la actividad online y el rendimiento académico de estudiantes de ingeniería y educación en dos sistemas de gestión del aprendizaje. Disponible en <https://repositorio.uchile.cl/handle/2250/146689>
- Núñez Ramírez, M.G. (2022). Representaciones sociales de la educación en línea en época de COVID 19 en estudiantes universitarios. *Dilemas contemporáneos: Educación, Política y Valores*.
- Ochoa, M.Á., Figueroa, J.F., & Zallas, F.A. (2021). Tendencias actuales en ambientes de enseñanza-aprendizaje-conocimiento del uso de Tecnología de Información y Comunicación en entornos educativos presenciales y a distancia en las Instituciones de Educación Superior.
- Olmedo-Neri, R.A. (2023). Comunicación en pandemia. Implicaciones tecnológicas y simbólicas en la experiencia de jóvenes universitarios. *Universitas*.
- Paba Barbosa , Carmelina, Lara Gutiérrez, Rosa María , Palmezano Rondón Annie Karina . Estilos de Aprendizaje y rendimiento académico en estudiantes Universitarios. Duazary [en línea]. 2008, 5(2), 99-106[fecha de Consulta 19 de Mayo de 2024]. ISSN: 1794-5992. Disponible en: <https://www.redalyc.org/articulo.oa?id=512156328004>
- Peñafiel-Vera, J., Burgos-Santana, A., Zavala-Alcívar, A.X., & Solís-Bravo, H. (2023). Análisis de las variables de un modelo de logística inversa aplicable a Pymes mediante Analytic Network Process. *Revista Científica y Arbitrada del Observatorio Territorial, Artes y Arquitectura: FINIBUS*.
- Quijije Quiroz, H.B., & Maldonado Zúñiga, K. (2023). Técnica de minería de datos para procesos educativos en estudiantes con necesidades educativas especiales basado en un modelo predictivo. *Revista Científica Arbitrada Multidisciplinaria PENTACIENCIAS*.
- Ríos, J.J., Rojas, H.L., & Niño, Á.D. (2017). Tendencias actuales de la educación superior en Colombia. *Revista CIFE "Lecturas de Economía Social"*, 18, 19-42.

- Rojas, E.T. (2017). Los Sitios Web de las Universidades Privadas de Santiago en el marco de la Sociedad de la Información.
- Romero Ochoa, M.A., Flores Figueroa, J., & Espinoza Zallas, F.A. (2018). Tendencias actuales en ambientes de enseñanza-aprendizaje-conocimiento del uso de Tecnología de Información y Comunicación en entornos educativos presenciales y a distancia en las Instituciones de Educación Superior. *Revista de Investigación Académica Sin Frontera: División de Ciencias Económicas y Sociales*.
- Schiappacasse Valenzuela, M. (2019). Sistema de recomendación para alumnos de primer año basado en sistemas de gestión del aprendizaje. Disponible en <https://repositorio.uchile.cl/handle/2250/170991>
- Suasnabas Pacheco, L.S., Carrasco Cachinelli, C.G., Morán Alvarado, M.D., & Medranda Valencia, R.B. (2023). Uso de las estrategias didácticas metodológicas, utilizando las TIC en la educación superior. *RECIAMUC*.
- Torres Velazquez, A.M., & Zanatta Colín, M.E. (2023). Diseño y validación de una escala de medición para explorar los saberes estructurantes en la formación integral de estudiantes de nivel universitario. *Revista Educación*.
- Villa-Esparza, J.I., Díaz-Levicoy, D., & Salcedo, A. (2023). Actividades sobre gráficos estadísticos en la evaluación diagnóstico integral de aprendizajes. *REVISTA DE INVESTIGACIÓN*.

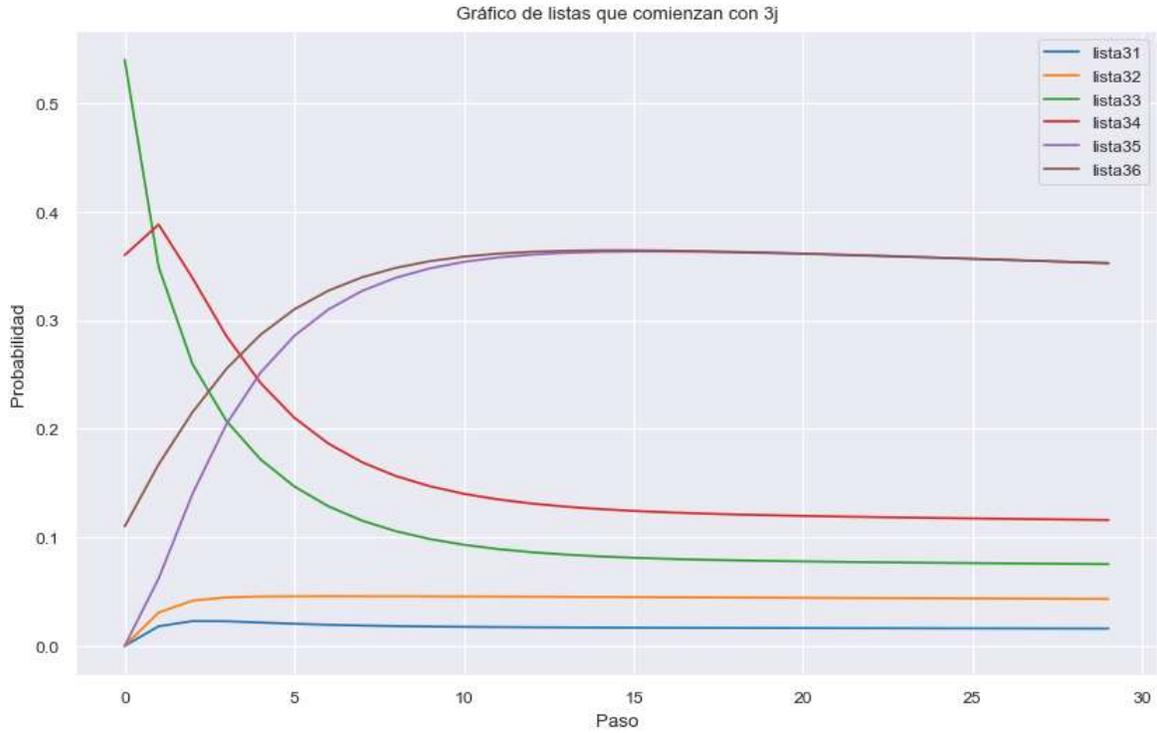
Anexos



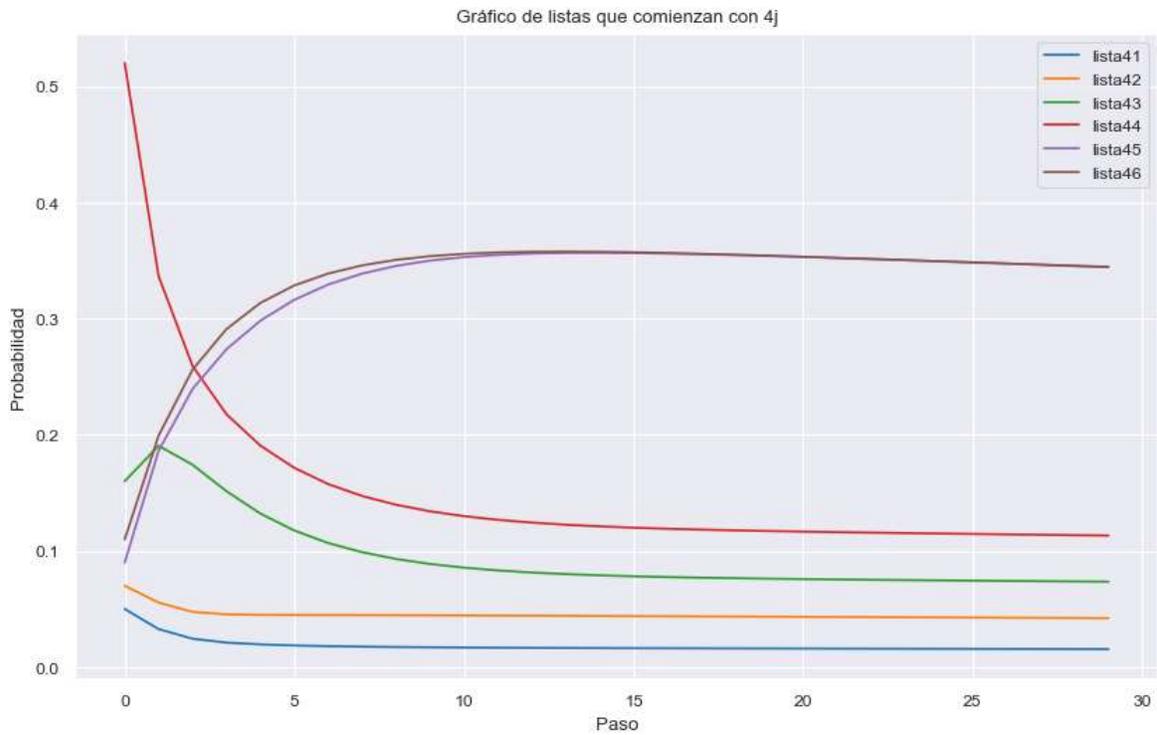
Anexo A: evolución de probabilidades de comenzar en el estado 1 y converger al estado j.



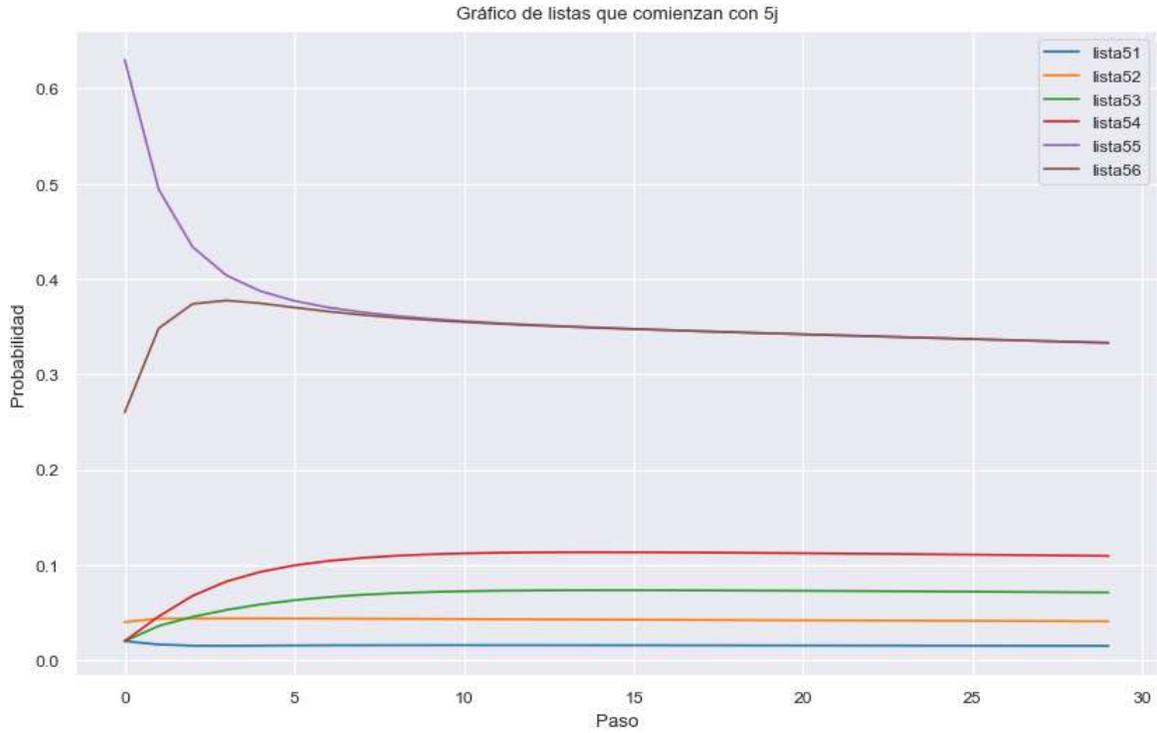
Anexo B: evolución de probabilidades de comenzar en el estado 2 y converger al estado j.



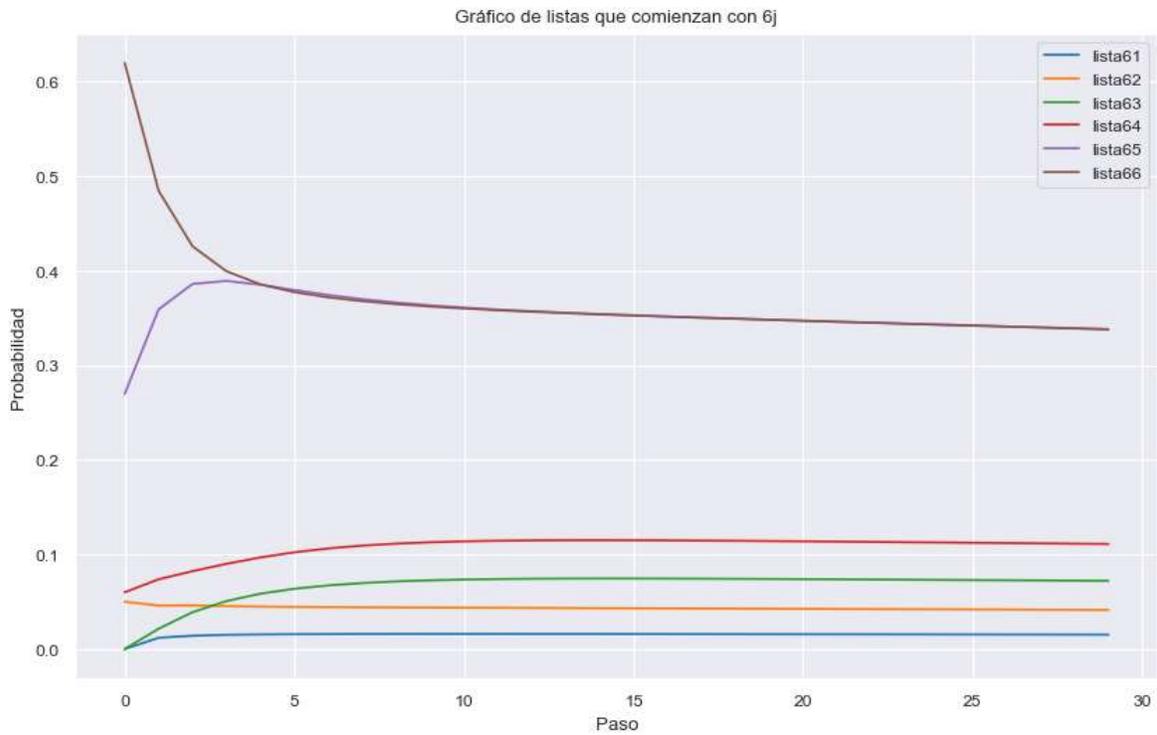
Anexo C: evolución de probabilidades de comenzar en el estado 3 y converger al estado j.



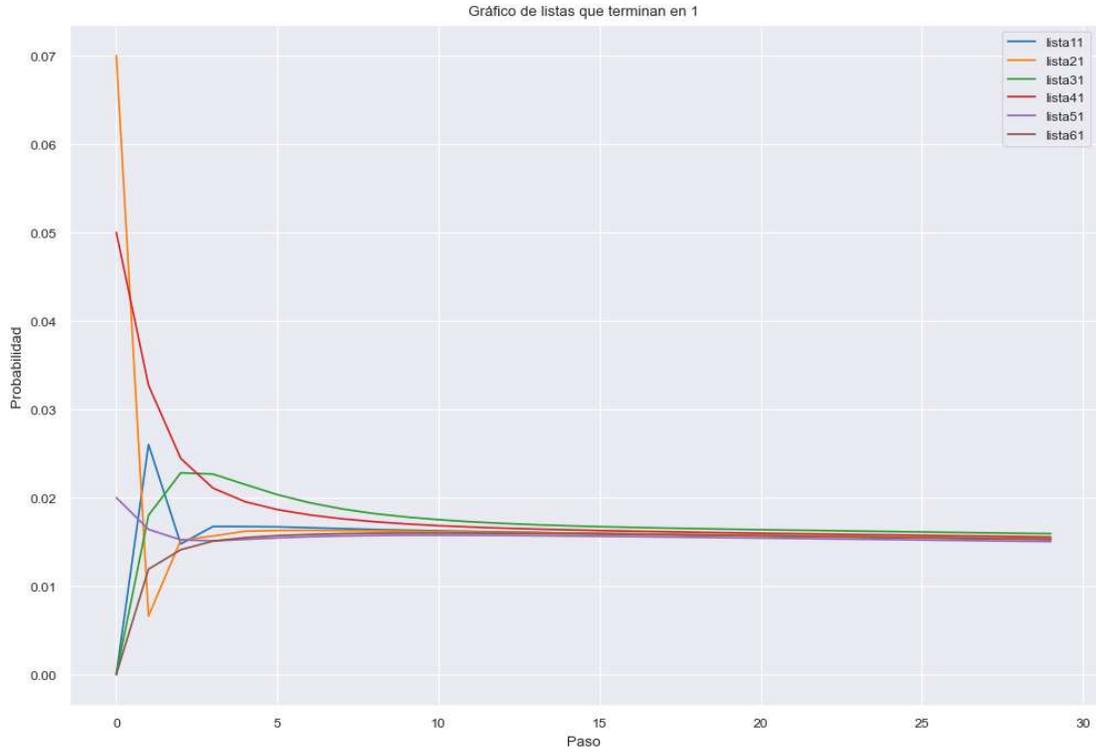
Anexo D: evolución de probabilidades de comenzar en el estado 4 y converger al estado j.



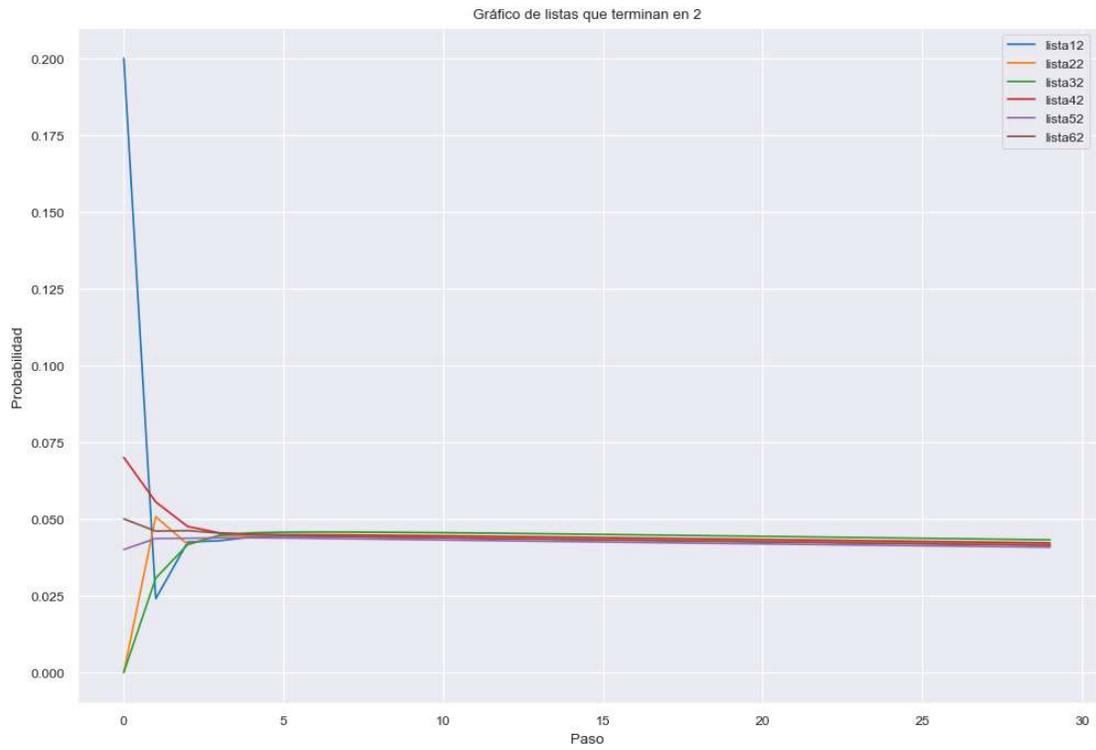
Anexo E: evolución de probabilidades de comenzar en el estado 5 y converger al estado j.



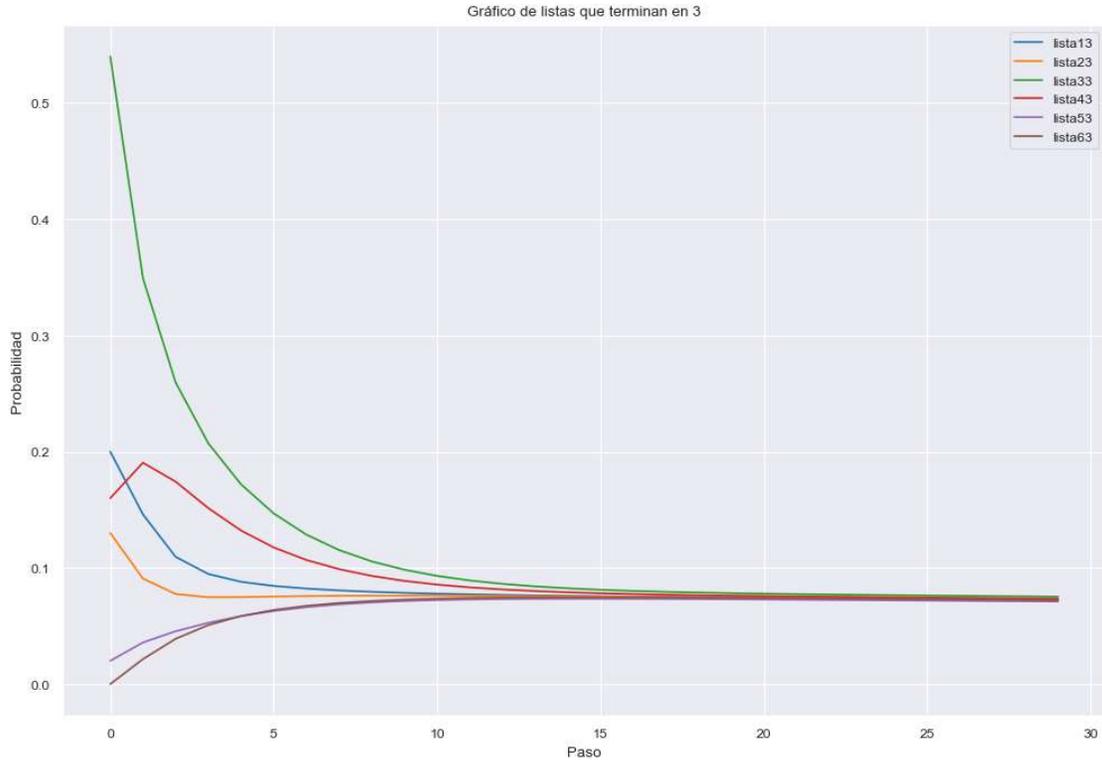
Anexo F: evolución de probabilidades de comenzar en el estado 6 y converger al estado j.



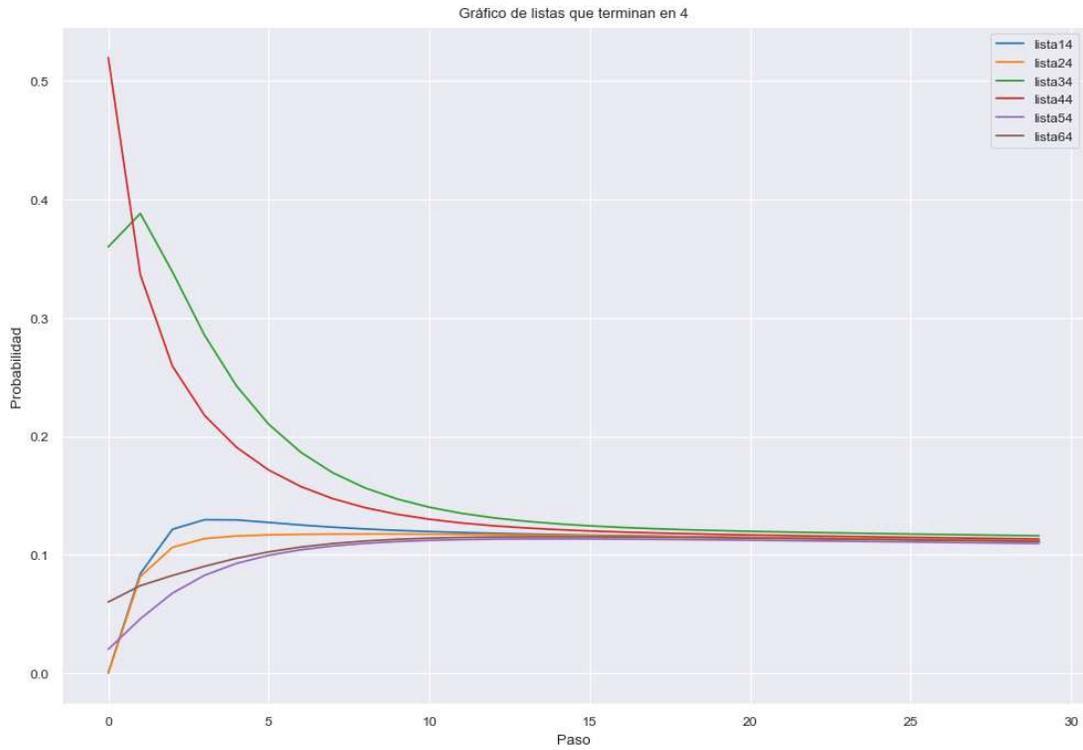
Anexo G: evolución de probabilidades de comenzar en el estado i y converger al estado 1.



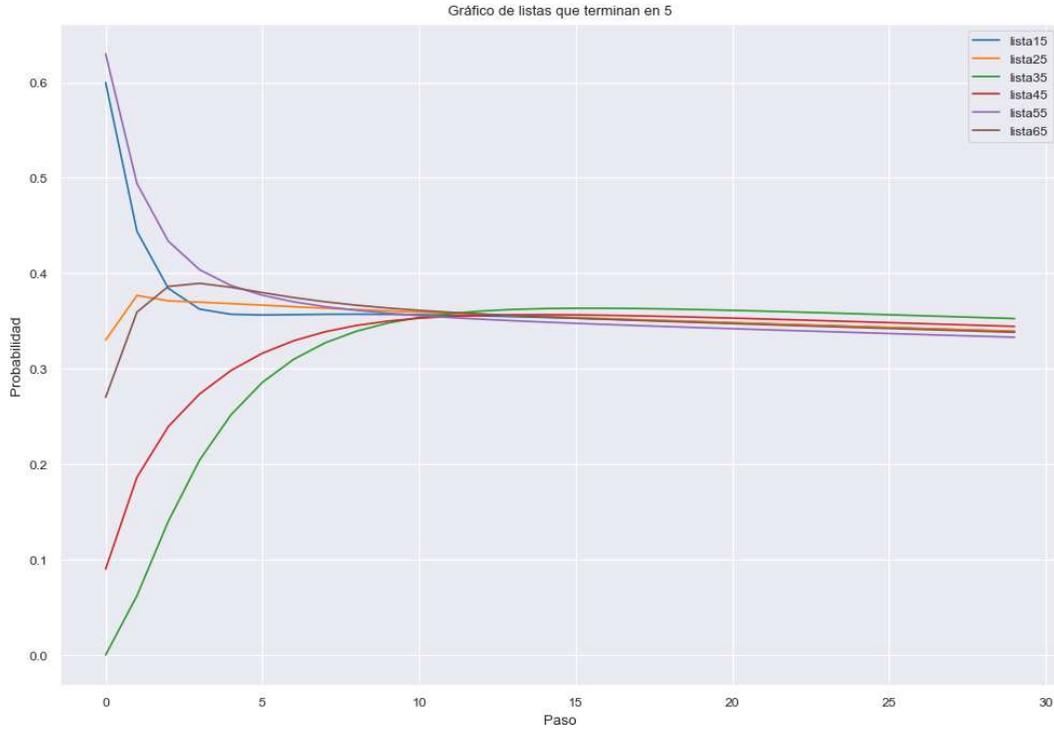
Anexo H: evolución de probabilidades de comenzar en el estado i y converger al estado 2.



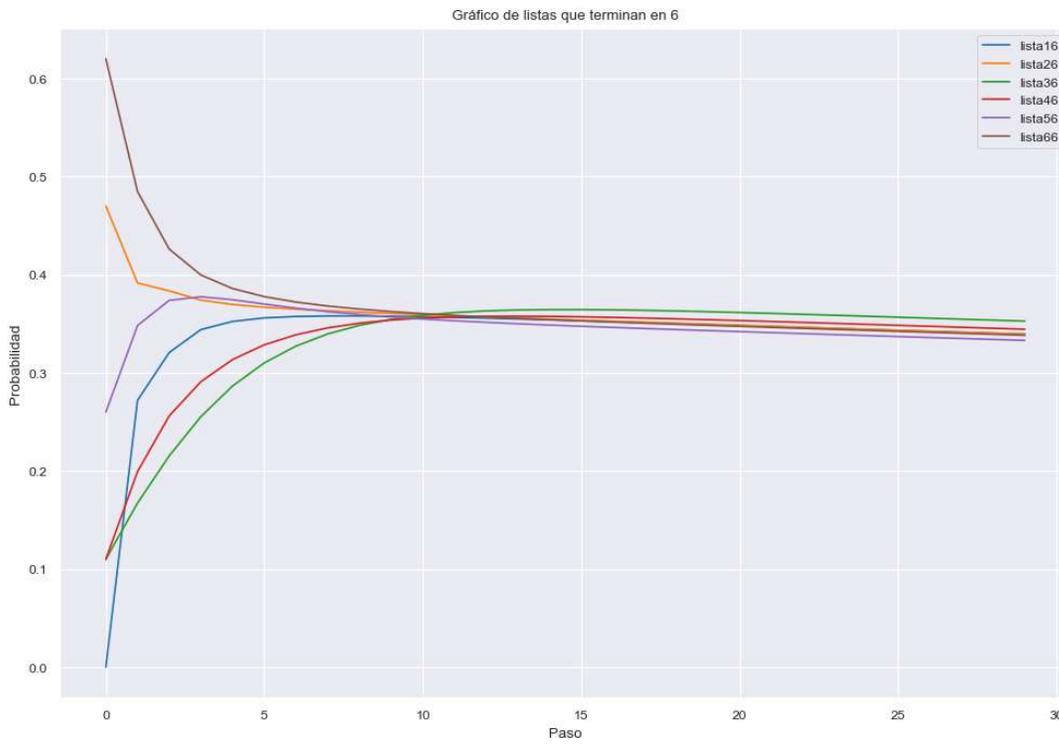
Anexo I: evolución de probabilidades de comenzar en el estado i y converger al estado 3.



Anexo J: evolución de probabilidades de comenzar en el estado i y converger al estado 4.



Anexo K: evolución de probabilidades de comenzar en el estado i y converger al estado 5.



Anexo L: evolución de probabilidades de comenzar en el estado i y converger al estado 6.