



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE INFORMÁTICA Y ELECTRÓNICA

CARRERA SOFTWARE

UNA APLICACIÓN WEB PARA EL APOYO DE TOMA DE DECISIONES EN EL PROCESO DE TITULACIÓN DE LA CARRERA DE SOFTWARE - ESPOCH

Trabajo de Integración Curricular

Tipo: Proyecto Técnico

Presentado para optar el grado académico de:

INGENIERO DE SOFTWARE

AUTOR:

HOMERO VINICIO BARRAGÁN SILVA

Riobamba – Ecuador

2023



ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE INFORMÁTICA Y ELECTRÓNICA

CARRERA SOFTWARE

**UNA APLICACIÓN WEB PARA EL APOYO DE TOMA DE
DECISIONES EN EL PROCESO DE TITULACIÓN DE LA
CARRERA DE SOFTWARE - ESPOCH**

Trabajo de Integración Curricular

Tipo: Proyecto Técnico

Presentado para optar el grado académico de:

INGENIERO DE SOFTWARE

AUTOR: HOMERO VINICIO BARRAGÁN SILVA

DIRECTORA: ING. IVONNE ELIZABETH RODRÍGUEZ FLORES

Riobamba – Ecuador

2023

© 2023, **Barragán Silva Homero Vinicio**

Se autoriza la reproducción total o parcial, con fines académicos, por cualquier medio o procedimiento, incluyendo cita bibliográfica del documento, siempre y cuando se reconozca el Derecho del Autor.

Yo, Homero Vinicio Barragán Silva, declaro que el presente Trabajo de Integración Curricular es de mi autoría y los resultados del mismo son auténticos. Los textos en el documento que provienen de otras fuentes están debidamente citados y referenciados.

Como autor asumo la responsabilidad legal y académica de los contenidos de este Trabajo de Integración Curricular; el patrimonio intelectual pertenece a la Escuela Superior Politécnica de Chimborazo.

Riobamba, 4 de diciembre de 2023



Homero Vinicio Barragán Silva


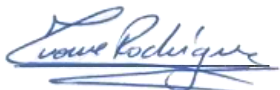

060628484-2

ESCUELA SUPERIOR POLITÉCNICA DE CHIMBORAZO

FACULTAD DE INFORMÁTICA Y ELECTRÓNICA

CARRERA SOFTWARE

El Tribunal del Trabajo de Integración Curricular certifica que: El Trabajo de Integración Curricular; Tipo Proyecto Técnico, **UNA APLICACIÓN WEB PARA EL APOYO DE TOMA DE DECISIONES EN EL PROCESO DE TITULACIÓN DE LA CARRERA DE SOFTWARE - ESPOCH**, realizado por el señor: **HOMERO VINICIO BARRAGÁN SILVA**, ha sido minuciosamente revisado por los Miembros del Trabajo de Integración Curricular, el mismo que cumple con los requisitos científicos, técnicos, legales, en tal virtud el Tribunal Autoriza su presentación.

	FIRMA	FECHA
Ing. Oscar Danilo Gavilánez Alvares PRESIDENTE DEL TRIBUNAL		2023-12-04
Ing. Ivonne Elizabeth Rodríguez Flores PhD. DIRECTORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR		2023-12-04
Ing. Gisel Katerine Bastidas Guacho ASESORA DEL TRABAJO DE INTEGRACIÓN CURRICULAR		2023-12-04

DEDICATORIA

Dedico este trabajo a mi padre, Vinicio, quien me otorgo el invaluable regalo de la educación; aunque hoy ya no esté conmigo, sus valores y enseñanzas me acompañarán el resto de mis días. A mi madre, Mariana, ejemplo de resiliencia, amor y alegría, que es la más grande figura a la que puedo aspirar. A mis hermanos, Claudia y Sebastián; a mis abuelitos, tíos, primos, familiares en general, por ser el pilar de mayor soporte en el transcurso de mi carrera. A mis amigos, quienes han sido la razón de mi alegría y han hecho de esta etapa la mejor de mi vida. A todos ellos, dedico este trabajo, que es el fruto del esfuerzo motivado por la certeza de siempre contar con su apoyo.

Homero

AGRADECIMIENTO

Mi más sincero agradecimiento a la Ing. Ivonne Rodríguez, por el apoyo y conocimientos brindados en el desarrollo del presente trabajo; su dedicación y compromiso en alcanzar la excelencia, fueron el motor que impulsó la culminación de este estudio. A la Ing. Gisel Bastidas, quien aportó significativamente en aspectos teóricos y prácticos del proyecto, pero sobre todo, por su colaboración paciente y constante orientación académica. A la Ing. Gloria Arcos e Ing. Diego Ávila, por compartir los consejos que permitieron mejorar continuamente la estructura y contenido del trabajo. A la Escuela Superior Politécnica de Chimborazo y sus docentes, por facilitar los espacios necesarios para el crecimiento personal e intelectual y aportar significativamente en mi educación profesional.

Homero

ÍNDICE DE CONTENIDO

ÍNDICE DE TABLAS.....	xi
ÍNDICE DE ILUSTRACIONES.....	xii
ÍNDICE DE ANEXOS.....	xiii
RESUMEN.....	xiv
SUMMARY.....	xv
INTRODUCCIÓN.....	1

CAPÍTULO I

1. DIAGNÓSTICO DEL PROBLEMA.....	2
1.1. Antecedentes del problema.....	2
1.2. Planteamiento del problema.....	3
1.2.1. <i>Formulación del problema</i>	3
1.2.2. <i>Sistematización del problema</i>	3
1.3. Justificación.....	3
1.3.1. <i>Justificación teórica</i>	3
1.3.2. <i>Justificación aplicada</i>	4
1.4. Objetivos.....	5
1.4.1. <i>Objetivo general</i>	5
1.4.2. <i>Objetivos específicos</i>	5

CAPÍTULO II

2. MARCO TEÓRICO.....	6
2.1. Toma de decisiones.....	6
2.1.1. <i>Toma de decisiones basada en datos</i>	6
2.1.2. <i>La toma de decisiones basada en datos en el contexto educativo</i>	7
2.2. Sistemas de apoyo a la toma de decisiones.....	7
2.2.1. <i>Características</i>	8
2.2.2. <i>Aplicaciones web de apoyo a la toma de decisiones</i>	8
2.3. Analítica de datos.....	9
2.3.1. <i>Minería de datos</i>	9
2.3.2. <i>Tipos de analítica de datos</i>	10
2.3.2.1. <i>Analítica descriptiva y de diagnóstico</i>	10

2.3.2.2.	<i>Analítica predictiva</i>	10
2.3.2.3.	<i>Analítica prescriptiva</i>	11
2.3.3.	<i>Analítica de datos en el ámbito educativo</i>	12
2.3.4.	<i>Métodos y técnicas de analítica de datos y su aplicación el ámbito educativo</i>	13
2.4.	Metodología ágil Scrum	16
2.4.1.	<i>Elementos de Scrum</i>	17
2.4.1.1.	<i>Equipo y roles</i>	17
2.4.1.2.	<i>Artefactos</i>	18
2.4.1.3.	<i>Eventos</i>	18
2.5.	Metodología CRISP-DM	19
2.5.1.	<i>Fases del proceso</i>	19
2.5.1.1.	<i>Comprensión del negocio</i>	20
2.5.1.2.	<i>Comprensión de los datos</i>	20
2.5.1.3.	<i>Preparación de los datos</i>	20
2.5.1.4.	<i>Modelado</i>	20
2.5.1.5.	<i>Evaluación</i>	20
2.5.1.6.	<i>Despliegue</i>	21
2.6.	Herramientas y lenguajes de desarrollo	21
2.6.1.	<i>Python</i>	21
2.6.2.	<i>Streamlit</i>	22
2.7.	Éxito de los sistemas de información	22
2.7.1.	<i>Modelo de éxito de los sistemas de información</i>	22
2.8.	Trabajos relacionados	24

CAPÍTULO III

3.	MARCO METODOLÓGICO	26
3.1.	Diseño de estudio	26
3.1.1.	<i>Tipo de estudio</i>	26
3.1.2.	<i>Métodos, técnicas y fuentes de estudio</i>	26
3.1.3.	<i>Proceso para la evaluación de la satisfacción de usuario de la aplicación web</i>	28
3.1.4.	<i>Operacionalización de variables</i>	28
3.1.5.	<i>Instrumento de medición</i>	30
3.1.6.	<i>Planteamiento de hipótesis</i>	30
3.1.7.	<i>Población y muestra</i>	31
3.2.	Desarrollo de la solución utilizando Scrum	31

3.2.1.	Análisis preliminar.....	31
3.2.1.1.	Requisitos de la aplicación.....	31
3.2.1.2.	Módulos de la aplicación.....	33
3.2.1.3.	Diagrama de caso de uso.....	34
3.2.1.4.	Equipo y roles.....	35
3.2.2.	Planificación.....	35
3.2.2.1.	Estudio de factibilidad técnica.....	35
3.2.2.2.	Análisis de riesgos.....	36
3.2.2.3.	Product Backlog.....	38
3.2.2.4.	Sprint Backlog.....	40
3.2.3.	Desarrollo de modelos de analítica utilizando CRISP-DM.....	41
3.2.3.1.	Comprensión del negocio.....	41
3.2.3.2.	Comprensión de los datos.....	43
3.2.3.3.	Preparación de los datos.....	45
3.2.3.4.	Modelado.....	47
3.2.3.5.	Evaluación.....	49
3.2.3.6.	Despliegue.....	51
3.2.4.	Desarrollo de la aplicación.....	53
3.2.4.1.	Arquitectura de la aplicación.....	53
3.2.4.2.	Estilo de codificación.....	55
3.2.4.3.	Diseño de la base de datos.....	55
3.2.4.4.	Diccionario de datos.....	56
3.2.4.5.	Diseño de las interfaces.....	57
3.2.5.	Cierre.....	57
3.2.5.1.	Burndown chart.....	58
3.2.5.2.	Implementación.....	58

CAPÍTULO IV

4.	ANÁLISIS E INTERPRETACIÓN DE RESULTADOS.....	60
4.1.	Evaluación de las variables según el modelo de DeLone y McLean.....	60
4.2.	Análisis de resultados.....	60
4.2.1.	Prueba del modelo de medidas.....	61
4.2.2.	Prueba del modelo estructural.....	63

CAPÍTULO V

5.	CONCLUSIONES Y RECOMENDACIONES	66
5.1.	Conclusiones	66
5.2.	Recomendaciones	67

GLOSARIO

BIBLIOGRAFÍA

ANEXOS

ÍNDICE DE TABLAS

Tabla 2-1: Técnicas de analítica aplicadas en el contexto de la educación.....	15
Tabla 3-1: Matriz de consistencia	26
Tabla 3-2: Operacionalización conceptual.....	28
Tabla 3-3: Operacionalización metodológica	29
Tabla 3-4: Indicadores clave	32
Tabla 3-5: Historias de usuario	32
Tabla 3-6: Equipo Scrum para el desarrollo del proyecto.....	35
Tabla 3-7: Ficha de gestión para el riesgo RG1	37
Tabla 3-8: Equivalencias para <i>T-Shirt Sizing</i>	38
Tabla 3-9: <i>Product backlog</i>	39
Tabla 3-10: Historia de usuario HU-02.....	40
Tabla 3-11: <i>Sprint backlog</i>	40
Tabla 3-12: Estructura de la Comisión de Titulación	42
Tabla 3-13: Vista general del plan del proyecto	43
Tabla 3-14: Descripción de las tablas de interés para el análisis	43
Tabla 3-15: Descripción de las columnas de interés para el análisis	44
Tabla 3-16: Convenciones de programación	55
Tabla 3-17: Diccionario de datos para la tabla "indicador"	56
Tabla 4-1: Pruebas de cargas externas	61
Tabla 4-2: Fiabilidad y validez convergente al eliminar cada ítem	61
Tabla 4-3: Resultados de la prueba de fiabilidad	62
Tabla 4-4: Prueba de validez convergente (AVE)	62
Tabla 4-5: Prueba de validez discriminante (HTMT)	63
Tabla 4-6: Coeficientes de regresión obtenidos para cada relación	64
Tabla 4-7: Resultados obtenidos al ejecutar el algoritmo de <i>bootstrapping</i>	64

ÍNDICE DE ILUSTRACIONES

Ilustración 2-1: Tipos de analítica de datos.....	11
Ilustración 2-2: El marco de trabajo Scrum.....	17
Ilustración 2-3: Fases del modelo referencial CRISP-DM.....	19
Ilustración 2-4: Modelo de éxito de los sistemas de información	23
Ilustración 3-1: Hipótesis por validar en el modelo	30
Ilustración 3-2: Módulos de la aplicación web “ <i>IMAK</i> ”	33
Ilustración 3-3: Contexto de uso del módulo de gráficos e indicadores.....	34
Ilustración 3-4: Estructura de descomposición de riesgos	36
Ilustración 3-5: Matriz de gravedad del riesgo.....	37
Ilustración 3-6: Reporte generado con <i>pandas-profiling</i>	44
Ilustración 3-7: Proceso ETL para la carga de datos en la tabla "estudiante_carrera"	46
Ilustración 3-8: Proceso de limpieza de los trabajos de titulación	47
Ilustración 3-9: Obtención del diccionario de palabras y corpus	47
Ilustración 3-10: Carga del modelo pre-entrenado para etiquetado gramatical.....	48
Ilustración 3-11: Función para extraer las palabras clave	49
Ilustración 3-12: Entrenamiento del modelo LDA.....	49
Ilustración 3-13: Precisión obtenida del modelo de etiquetado gramatical.....	50
Ilustración 3-14: Coherencia obtenida del modelo LDA	50
Ilustración 3-15: Resultados del ajuste de hiperparámetros	51
Ilustración 3-16: Nube de palabras con los términos recurrentes.....	51
Ilustración 3-17: Visualización de los tópicos y términos más relevantes	52
Ilustración 3-18: Distribución de la duración en meses de los proyectos no finalizados	53
Ilustración 3-19: Diagrama de contexto de la aplicación	54
Ilustración 3-20: Diagrama de contenedor de la aplicación	54
Ilustración 3-21: Diseño de la base de datos	56
Ilustración 3-22: <i>Wireframe</i> de la pantalla “Panel”.....	57
Ilustración 3-23: <i>Burndown chart</i>	58
Ilustración 3-24: Diagrama de despliegue.....	59
Ilustración 4-1: Coeficientes de regresión obtenidos	64

ÍNDICE DE ANEXOS

ANEXO A: CUESTIONARIO PARA EVALUACIÓN DE SATISFACCIÓN DE USUARIO

ANEXO B: MANUAL TÉCNICO - SCRUM

ANEXO C: MANUAL TÉCNICO – CRISP-DM

ANEXO D: MANUAL DE USUARIO

RESUMEN

La Comisión de Titulación de la carrera de Software - ESPOCH no contaba con la información suficiente que facilitara la toma de decisiones, ya que carecían de una herramienta tecnológica que apoyara la ejecución de sus labores; como consecuencia, a sus miembros se les dificultaba realizar el seguimiento al proceso de titulación. En el presente trabajo se implementó una aplicación web denominada “*IMAK*”, diseñada para apoyar la toma de decisiones de la Comisión de Titulación. El enfoque metodológico del estudio fue de tipo aplicativo, ya que se seleccionaron y emplearon técnicas de analítica de datos para generar información de valor, utilizando para ello la metodología CRISP-DM; además, se desarrolló la aplicación web empleando Scrum en un total de ocho *sprints* o iteraciones. Las variables de satisfacción de usuario, calidad del sistema, calidad de la información y uso fueron consideradas en la evaluación de la aplicación web, las cuales se estimaron con un modelo de ecuaciones estructurales mediante un análisis de mínimos cuadrados parciales (PLS-SEM). Los resultados de las pruebas demostraron la validez y fiabilidad de los indicadores seleccionados para medir las variables. Se determinó que la calidad del sistema y la calidad de la información tuvieron influencia positiva sobre la satisfacción de usuario, en contraposición al uso, que tuvo influencia negativa. Sin embargo, con un valor de $\alpha = 0.05$, ninguna relación resultó ser significativa y el poder de las variables dependientes para predecir la variable dependiente es débil, según el valor obtenido del coeficiente de determinación R^2 .

Palabras clave: <INGENIERÍA DE SOFTWARE>, <APLICACIÓN WEB PARA EL APOYO A LA TOMA DE DECISIONES (DSS)>, <ANALÍTICA DE DATOS>, <SATISFACCIÓN DE USUARIO>, <DELONE & MCLEAN>.



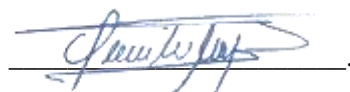
20-12-2023

2272-DBRA-UPT-2023

SUMMARY

The ESPOCH-Software Degree Commission did not have sufficient information to facilitate decision-making, since they lacked a technological tool to support the execution of their tasks; As a consequence, it was difficult for its members to follow up on the graduation process. In this work, a web application called "IMAK" was implemented, designed to support the decision-making of the Degree Commission. The methodological approach of the study was of an application type, since data analysis techniques were selected and used to generate valuable information, using the CRISP-DM methodology; In addition, the web application was developed using Scrum in a total of eight sprints or iterations. The variables of user satisfaction, system quality, information quality and use were considered in the evaluation of the web application, which were estimated with a structural equation model through a partial least squares analysis (PLS-SEM). The test results demonstrated the validity and reliability of the indicators selected to measure the variables. It was determined that the quality of the system and the quality of the information had a positive influence on user satisfaction, as opposed to use, which had a negative influence. However, with a value of $\alpha = 0.05$, no relationship was found to be significant and the power of the dependent variables to predict the dependent variable is weak, according to the value obtained from the coefficient of determination R^2 .

Keywords: <SOFTWARE ENGINEERING>, <WEB APPLICATION DECISION SUPPORT SYSTEM (DSS)>, <DATA ANALYTICS>, <USER SATISFACTION>, <DELONE & MCLEAN>.



Lic. Nelly Padilla P. Mgs.

0603818717

DOCENTE FIE

INTRODUCCIÓN

La necesidad de contar con aplicaciones que apoyen a la toma de decisiones en las instituciones educativas es fundamental en la actualidad, debido a la cantidad de datos que generan como resultado de los procesos de enseñanza-aprendizaje. Los datos, por sí solos, no son una fuente de la cual se pueda derivar conclusiones en el ámbito organizacional, sino que deben ser contextualizados, con el objetivo de extraer la información relevante para estudiantes, docentes y administradores.

Los sistemas de apoyo a la toma de decisiones son herramientas que proveen información útil a los individuos en una organización, de tal manera que puedan tomar decisiones informadas en el contexto de las actividades que desempeñan. En el ámbito educativo, la adopción de sistemas de apoyo a la toma de decisiones se convierte en una necesidad central a nivel administrativo, ya que estos brindan a los líderes educativos una visión completa y basada en datos sobre aspectos críticos, como la asignación de recursos, la planificación curricular y la evaluación institucional. Al recolectar, analizar y presentar de manera accesible datos relevantes, estos sistemas ayudan a los administradores a tomar decisiones estratégicas sólidamente fundamentadas, optimizando la gestión de la institución y dirigiendo los esfuerzos hacia la mejora continua y el logro de los objetivos educativos preestablecidos.

El objetivo del presente trabajo es implementar una aplicación web que apoye a la toma de decisiones de la Comisión de Titulación de la carrera de Software de la ESPOCH, quienes han almacenado datos referentes al proceso de titulación pero no cuentan con información útil que les permita tomar decisiones estratégicas para solventar problemas en el ámbito de sus labores.

El trabajo se estructura en cinco capítulos:

CAPÍTULO I: Se identifica el problema a abordar, así como sus antecedentes, justificación y los objetivos a cumplir en el estudio.

CAPÍTULO II: Describe los aspectos teóricos que fundamentan el desarrollo de la solución, permitiendo identificar los componentes conceptuales de relevancia.

CAPÍTULO III: Presenta el diseño de estudio y el desarrollo de la aplicación web para el apoyo a la toma de decisiones, así como la construcción de analíticas de interés para la Comisión de Titulación.

CAPÍTULO IV: Aborda los resultados obtenidos en la evaluación de la aplicación web.

CAPÍTULO V: Se describen las conclusiones y recomendaciones a las que se llegaron tras el desarrollo del proyecto.

CAPÍTULO I

1. DIAGNÓSTICO DEL PROBLEMA

En este capítulo se plantea el problema abordado, sus antecedentes, justificación y los objetivos que se cumplieron en el desarrollo del proyecto.

1.1. Antecedentes del problema

Como mencionan Webber y Zheng (2020), al igual que en el mundo de los negocios, la educación superior ha experimentado la necesidad de sumarse al movimiento de la analítica. Las universidades están interesadas en utilizar los datos que generan, que por sí solos no mejorarán el proceso de toma de decisiones. Para ello, es necesario emplear información contextualizada y técnicas de análisis que se ajusten al caso específico de cada institución.

Los sistemas de soporte a decisiones están diseñados para apoyar y asistir el proceso de toma de decisiones (Alahmadi & Jamjoom, 2022, p.114). Este tipo de aplicaciones buscan ser una herramienta que amplie las habilidades de las personas al brindar facilidad de uso, adaptabilidad y flexibilidad, soporte en todos los niveles administrativos, facilidad de acceso a los datos, entre otras características (Rashidi et al., 2018).

La Comisión de Unidad de Integración Curricular¹ es un comité conformado por el Coordinador de la Carrera, un profesor de la Unidad Profesional y un profesor de la Unidad de Integración Curricular, quienes ejercen sus funciones por un mínimo de dos periodos académicos ordinarios. Sus atribuciones incluyen la planificación y estructuración de actividades, además de la evaluación de las solicitudes de los estudiantes para acceder a las diferentes modalidades de titulación². La Comisión de Titulación (CT) de la carrera de Software almacena datos referentes al proceso de titulación de los estudiantes, sin embargo, no cuentan con la información suficiente para facilitar la toma de decisiones que conduzcan a solucionar problemas dentro del contexto de sus labores, esto debido a que carecen de una herramienta tecnológica que permita a sus miembros disponer de información útil sobre los trabajos de integración curricular que se realizan en la carrera. Como consecuencia, a los miembros de la comisión se les dificulta tomar decisiones adecuadas e identificar falencias en el proceso de titulación.

Dentro del ámbito de la educación superior se han desarrollado varias aplicaciones de apoyo a la toma de decisiones. Ramalho y Segundo (2020, pp.1-8) desarrollaron un sistema web utilizando R-

¹ En el Artículo 1, Política No. 9 de la resolución 002.CP.2023 del Consejo Politécnico, se señala que la Comisión de Unidad de Integración Curricular pasa a denominarse “Comisión de Titulación” (CT), por lo que se utilizará este término en lo que resta del trabajo.

² Reforma al Reglamento del Régimen Académico. Resolución 259.CP.2022, Consejo Politécnico, ESPOCH (19 de abril de 2022).

Shiny para la visualización de datos del “*Brazilian Digital Library of Theses and Dissertations*”, utilizando además métodos estadísticos. Esquivel y Esquivel (2021, pp.50-54) describen en su trabajo la construcción de un sistema de apoyo a la toma de decisiones que aplica un modelo de regresión logística para predecir la admisión de nuevos estudiantes en una universidad filipina. Dentro del país, Mendoza et al. (2017) desarrollaron un sistema de información web para apoyar la toma de decisiones respecto a factores socioeconómico del alumnado de la Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López.

Como una forma de solventar el problema anteriormente expuesto, se plantea el desarrollo de una aplicación de apoyo a decisiones que pueda ser utilizada por los miembros de la Comisión de Titulación de la carrera de Software de la ESPOCH. Los datos con los que se trabajarán corresponden a registros proporcionados por la Comisión.

1.2. Planteamiento del problema

1.2.1. Formulación del problema

¿Cómo una aplicación web para el apoyo a la toma de decisiones influye en la satisfacción del usuario?

1.2.2. Sistematización del problema

- ¿Qué técnicas de analítica de datos pueden emplearse en la toma de decisiones en el ámbito educativo?
- ¿Cuál es la información requerida por los miembros de la Comisión de Titulación para la toma de decisiones?
- ¿Cuál es el grado de satisfacción con la aplicación web?

1.3. Justificación

1.3.1. Justificación teórica

La toma de decisiones es fundamental en cualquier tipo de organización, incluyendo las universidades. Las decisiones que se toman dentro de este tipo de instituciones tienen un impacto tanto en sus miembros como en aspectos económicos, políticos y sociales del entorno en el que se establecen. Como resultado, administradores y directores de instituciones de educación superior se han apoyado en herramientas informativas para facilitar este proceso (Nieto et al., 2019, pp.4145-4153). Los avances tecnológicos, como la capacidad mejorada para almacenar datos, son

un desafío y oportunidad en la toma de decisiones dentro de las universidades. Gracias a la analítica, se ha facilitado el cumplimiento de metas y ayudado a obtener una mejor eficiencia en las universidades (Webber y Zheng, 2020). Hoy en día, las instituciones almacenan una gran cantidad de datos como resultado de los procesos que realizan, lo que ha hecho factible la toma de decisiones informadas. Murray (2014) menciona que, a pesar de la complejidad del ciclo de enseñanza y aprendizaje, el uso apropiado de los datos puede informar y mejorar el proceso de toma de decisiones. Al igual que sucede en otras áreas como la medicina, la ingeniería y la fabricación de automóviles, los datos son la base para la evaluación del progreso y toma de decisiones de calidad. La analítica de datos provee un nuevo modelo que permite a los líderes educativos mejorar la eficiencia organizacional, y como consecuencia, convertirse en la piedra angular para un cambio sistemático (Siemens y Long, 2011, pp.31-40). No obstante, contar con una gran cantidad de información derivada de la analítica no conducirá por sí sola a ver estos beneficios reflejados en las instituciones de educación superior. En cambio, el proceso de utilizar datos para la toma de decisiones en el ámbito educativo debe involucrar la generación de información de valor que pueda guiar las actividades pedagógicas de los docentes (Wise y Jung, 2019, pp.53-69).

1.3.2. Justificación aplicativa

El desarrollo de una aplicación web para el apoyo a la toma de decisiones ayuda a que la Comisión de Titulación pueda realizar análisis informados de los datos que recolectan. La aplicación cuenta con elementos para visualizar indicadores de interés, apoyados en técnicas que facilitan la analítica de datos.

Los módulos con los que cuenta la aplicación son los siguientes:

Módulo de gráficos e indicadores

- Visualizar indicadores clave.
- Crear indicadores personalizados.
- Cargar datos externos.

Módulo de analíticas

- Visualizar analíticas de datos

Módulo de administración

- Actualizar vistas materializadas.
- Generar plantillas en formato .xlsx.

Módulo de autenticación

- Autenticar usuario en la aplicación.

El presente trabajo coincide con la línea de investigación transversal de la ESPOCH: Tecnología de la Información y la Comunicación; programa: Ingeniería de Software. Así mismo, a la línea de investigación: Administración y Economía Popular; programa: Gestión de los Sistemas de Información. Se articula también al Plan Nacional de Desarrollo, específicamente al Eje Social, objetivo 7: Potenciar las capacidades de la ciudadanía y promover una educación innovadora, inclusiva y de calidad en todos los niveles.

1.4. Objetivos

1.4.1. Objetivo general

Implementar una aplicación web que apoye a la toma de decisiones de la Comisión de Titulación de la carrera de Software de la ESPOCH.

1.4.2. Objetivos específicos

- Seleccionar técnicas de analítica de datos que apoyen la toma de decisiones de la Comisión de Titulación.
- Aplicar técnicas de analítica de datos para generar información que apoye la toma de decisiones.
- Desarrollar la aplicación web para la toma de decisiones con base en la información requerida por la Comisión de Titulación.
- Evaluar el nivel de satisfacción de los usuarios respecto al uso de la aplicación web para el apoyo a la toma de decisiones.

CAPÍTULO II

2. MARCO TEÓRICO

El presente capítulo fundamenta conceptos teóricos que apoyan el desarrollo del proyecto técnico. Se exploran las definiciones de toma de decisiones, analítica y su aplicación en el ámbito educativo, metodologías para el desarrollo de software y minería de datos, herramientas, lenguajes de programación y calidad de software. Finalmente, se exponen los trabajos relacionados a este estudio.

2.1. Toma de decisiones

Eichhorn et al. (2018, pp.263-312) definen a la toma de decisiones como una de las etapas en el proceso administrativo de las empresas. Para Chiavenato (2009) es “un proceso que consiste en hacer un análisis y elegir entre varias opciones un curso de acción”. Se trata de una actividad en la que las personas encargadas se enfrentan a múltiples factores que modelan el contexto en el que las decisiones son seleccionadas (Munier et al., 2019). No consiste simplemente en escoger cualquiera de las opciones entre todas las alternativas, sino la mejor, o por lo menos aquella que más se ajuste a las necesidades de la empresa. El proceso de toma de decisiones depende en gran medida de quienes participan en él. De hecho, es una acción “humano-dependiente”, pues involucra el razonamiento, la lógica e inclusive la experiencia de quien toma las decisiones. Pero el factor humano no es el único elemento que influye en este proceso. Chiavenato (2009) afirma que también implica la participación de seis elementos, los cuales son, además de la persona, los objetivos, las preferencias, la estrategia, la situación y el resultado.

2.1.1. Toma de decisiones basada en datos

La toma de decisiones basada en datos puede definirse como la práctica de basar las decisiones en el análisis de datos en lugar de emplear únicamente la intuición (Diván, 2017, pp.50-56). Por otro lado, Donhost et al. (2015, pp.56-63) la definen como un proceso continuo, cíclico y que involucra trabajo en equipo y colaboración. No es un evento que ocurre una sola vez o que se completa al disponer de herramientas para el apoyo a la toma de decisiones. Aunque esta técnica pueda mejorar el proceso, no implica que las decisiones deban guiarse únicamente por datos. Como se mencionaba con anterioridad, el proceso de toma de decisiones es humano – dependiente, y estriba en gran medida en las capacidades de las personas para discernir entre los distintos cursos de acción que se hayan identificado. La información puede tener significados distintos, pues cada

uno la interpreta de forma diferente basado en su propia experiencia e intuición; razón por la cual, tomar decisiones no solamente se apoya en datos (Datnow et al., 2017, pp.354-360).

2.1.2. La toma de decisiones basada en datos en el contexto educativo

Al igual que cualquier organización, las instituciones educativas también toman decisiones. Directores, docentes, personal administrativo, incluso los propios estudiantes, deben seleccionar la mejor alternativa entre un catálogo de opciones de manera frecuente. Al mismo tiempo, las instituciones educativas recolectan datos de distintas fuentes que podrían convertirse en medios potenciales para el apoyo a la toma de decisiones informadas. Schildkamp (2019, pp.257-273) afirma que los datos juegan un rol importante en la mejora de procesos en el ámbito educativo, haciendo énfasis además en que estos por sí solos no son capaces de solucionar los problemas de las instituciones educativas, sino que deben contextualizarse y ponerse a disposición de los *decision makers* para ser analizados. Por otro lado, Donhost et al. (2015, pp.56-63) destacan el problema referente a la gran cantidad de datos disponibles en el contexto educativo y que no se procesa adecuadamente, asegurando que lejos de mejorar el panorama, podría llegar a confundir a los involucrados. Consecuentemente los datos deben ser organizados y filtrados de tal forma que aporten valor a las organizaciones. Caso contrario, los *decision makers* no serán capaces de encontrar el potencial real de la información y el proceso de toma de decisiones no tendrá relevancia para las instituciones. Para Cramer et al. (2014, pp.389-400), incorporar un proceso de toma de decisiones basada en datos común en las instituciones de educación superior es fundamental. Esto debido a que permite a los docentes mantener una estructura estándar de objetivos y competencias en múltiples niveles, facilitando la comprensión de los sistemas educativos en toda su complejidad. Los datos reflejan el trabajo de los pedagogos, el esfuerzo y el resultado de los procesos que implementan, información que permite diseñar estrategias para mejorar de manera conjunta el proceso de enseñanza – aprendizaje (Cramer et al., 2014, pp.389-400).

2.2. Sistemas de apoyo a la toma de decisiones

En el proceso de toma de decisiones es importante contar con herramientas que apoyen a los *decision makers*. Los datos son un soporte útil en este proceso, siempre y cuando se organicen y contextualicen adecuadamente. La implementación de soluciones tecnológicas puede ayudar en la administración y presentación de la información requerida a través de gráficas e indicadores que son fácilmente entendibles.

Los sistemas de apoyo a la toma de decisiones (en inglés *Decision Support System* o DSS) se definen como soluciones tecnológicas que pueden ser utilizadas en el apoyo a la toma de decisiones complejas. Un DSS ofrece funcionalidades de recolección y análisis de datos,

comunicación de información de interés a través de gráficos y la integración de distintos tipos de datos (Zabeo et al., 2022, pp.2101-2113). Los DSS pueden apoyar distintas actividades del proceso de toma de decisiones y su importancia radica en el área de investigación bastante amplia que maneja, combinando disciplinas como la medicina, la estadística, la economía, entre otras (Belciug y Gorunescu, 2019, pp.57-70). Los sistemas de apoyo a la toma de decisiones pueden dividirse en cinco categorías según su enfoque, sean estos basados en comunicación, datos, documentos, conocimiento o modelos (Bandyopadhyay, 2023).

2.2.1. Características

Rashidi et al. (2018) identifican algunas características de los sistemas de apoyo a la toma de decisiones que los distinguen de otros tipos de sistemas:

- Asisten la toma de decisiones respecto a problemas poco o nada estructurados;
- pueden estar destinados a distintos niveles administrativos, así como a individuos y grupos;
- abarcan todas las fases del proceso de toma de decisiones;
- cubren una o varias herramientas de análisis de decisión;
- son flexibles, adaptativos, amigables con el usuario y cuentan con interfaces gráficas robustas;
- su enfoque es mejorar la efectividad de la toma de decisiones, mas no su eficiencia;
- sirven de soporte para los *decision makers*, pero no pretenden reemplazarlos;
- generalmente emplean modelos para el análisis de problemas, lo que permite la experimentación con distintas configuraciones y estrategias;
- brindan acceso a diversas fuentes y formatos de datos; y
- pueden acoplarse a otros sistemas, así como ser distribuidos a través de la red.

2.2.2. Aplicaciones web de apoyo a la toma de decisiones

Una aplicación web de apoyo a la toma de decisiones consiste en una herramienta computarizada que provee al usuario de información o recursos que apoyen el proceso de toma de decisiones, todo esto a través de un navegador web (Dastres y Soori, 2021, pp.1-15). La expansión y popularización del Internet a finales del siglo pasado facilitó y promovió el uso de este tipo de sistemas que actualmente son de interés dentro de áreas como la salud, el gobierno, la educación, el sector privado, entre otros. Han dejado de ser una herramienta añadida para convertirse en una parte esencial del proceso del negocio (Bhargava et al., 2007, pp.1083-1095). La principal ventaja de este tipo de DSS es la misma que la mayoría de las aplicaciones web que han reemplazado sistemas de

escritorio: estar al alcance de los usuarios desde prácticamente cualquier dispositivo. Pero su aparición también ha implicado nuevos desafíos para quienes los implementan, como el transportar grandes cantidades de datos a través de redes sin estado, retrasos en el envío, necesidad de nuevos modelos de pago para acceder a la información y el apareamiento de nuevos tipos de usuarios sin experiencia previa (Bhargava et al., 2007, pp.1083-1095).

Conforme a las características señaladas previamente, el presente trabajo se enfoca en el desarrollo de una aplicación web para el apoyo a la toma de decisiones.

2.3. Analítica de datos

La analítica de datos consiste en el análisis de grandes cantidades de datos utilizando sistemas informáticos, de tal manera que faciliten el proceso de toma de decisiones. Este campo involucra disciplinas como la estadística, el reconocimiento de patrones, la inteligencia artificial, el aprendizaje de máquina, entre otras (Runkler, 2020). La analítica también implica la recolección de datos de diversas fuentes, procesarlos para encontrar en ellos patrones de interés y distribuir los resultados a los *stakeholders* adecuados. Su campo de estudio ha conducido la búsqueda de métodos efectivos para la recolección, procesamiento, comprensión, extracción y presentación del conocimiento incrustado en los datos (Soltanpoor y Sellis, 2016, pp.245-256).

2.3.1. Minería de datos

Minería de datos es una de las etapas en el proceso *Knowledge Discovery in Databases* o KDD, introducido por primera vez en 1989, aunque en el mundo de las bases de datos ya se había utilizado este término con anterioridad (Piatetsky-Shapiro, 2000, pp.59-61). Autores como Fayyad et al. (1996, p.37), Calders y Custers (2013, pp.27-42) la definen como el proceso en el que, tras aplicar una serie de técnicas, se obtiene como resultado un conjunto de modelos o patrones. Estas técnicas pueden incluir análisis de datos, aprendizaje de máquina, estadística o reconocimiento de patrones y se aplican a largos conjuntos de datos provenientes de fuentes como *data warehouses* o la propia Web. Minería de datos, análisis y analítica de datos son términos utilizados frecuentemente de forma indistinta (Calders y Custers, 2013, pp.27-42). Aunque no se ha llegado a un acuerdo unánime sobre su uso, “analítica de datos” podría considerarse como una modernización del término “minería de datos” (Rodríguez y Bernal, 2019).

2.3.2. Tipos de analítica de datos

Según Delen (2015), la analítica de datos puede dividirse en tres categorías dependiendo el nivel de madurez de la organización o el grado de sofisticación que se desea alcanzar, estas son: descriptiva, predictiva y prescriptiva.

2.3.2.1. Analítica descriptiva y de diagnóstico

En esta categoría se describe un fenómeno a través de diferentes medidas o indicadores que muestran sus dimensiones relevantes (Banerjee et al., 2013, pp.1-12). Es el nivel más básico de analítica, ya que en este se emplean métodos estadísticos simples para describir los datos. La analítica descriptiva es conocida también como la fase de reportería, resumen de los datos o de inteligencia del negocio, ya que se realiza una presentación de lo que ocurrió en el pasado de la organización, respondiendo a la pregunta ¿qué es lo que pasó? Por otro lado, la analítica de diagnóstico es una extensión de la analítica descriptiva enfocada en responder a la pregunta ¿por qué pasó? (Delen, 2015; Banerjee et al., 2013, pp.1-12; Soltanpoor y Sellis, 2016, pp.245-256). Al ser el nivel más básico de analítica, no se puede obviar esta etapa si lo que se busca es aplicar cualquiera de los niveles superiores. La analítica descriptiva y de diagnóstico constituyen un paso inicial para entender los datos, capturando sus atributos en medidas que sean entendibles por los *decision makers*.

2.3.2.2. Analítica predictiva

Es el siguiente nivel en la categoría y va un paso más allá que la analítica descriptiva, ya que en ella se predicen sucesos futuros a través de modelos que se consiguen aplicando técnicas como el aprendizaje de máquina, la minería de datos³ o incluso simulaciones; en la analítica predictiva se responde a la pregunta ¿qué pasará? (Soltanpoor y Sellis, 2016, pp.245-256). Para Delen (2015), el hecho de que una organización realice analítica predictiva implica un nivel de madurez superior que aquellas empresas que solamente realizan analítica descriptiva. En este sentido, el nivel de madurez puede reflejar ciertos aspectos de las organizaciones, por ejemplo, la capacidad aumentada de las empresas para recolectar y procesar datos. Las organizaciones que recolectan más datos tienen la oportunidad de construir mejores modelos y, por ende, obtener resultados más precisos (Soltanpoor y Sellis, 2016, pp.245-256).

³ En este contexto se considera a la minería de datos como un conjunto de técnicas, mas no como un sinónimo de analítica de datos.

2.3.2.3. Analítica prescriptiva

La última categoría corresponde a la analítica prescriptiva. En este nivel se pretende encontrar, mediante complejos modelos matemáticos, el mejor curso de acción entre una serie de alternativas identificadas a través de la analítica descriptiva y/o predictiva de los datos (Delen, 2015). La analítica prescriptiva corresponde a la categoría más sofisticada entre las descritas con anterioridad, siendo la que aporta un mayor valor a la organización. Utilizando técnicas como la simulación y la optimización, responde a la pregunta ¿qué debería hacerse? (Lepenioti et al., 2020, pp.57-70; Šikšnys y Pedersen, 2018, pp.2792-2793).

Las principales características descritas en los apartados anteriores se resumen en la **Ilustración 2-1**.

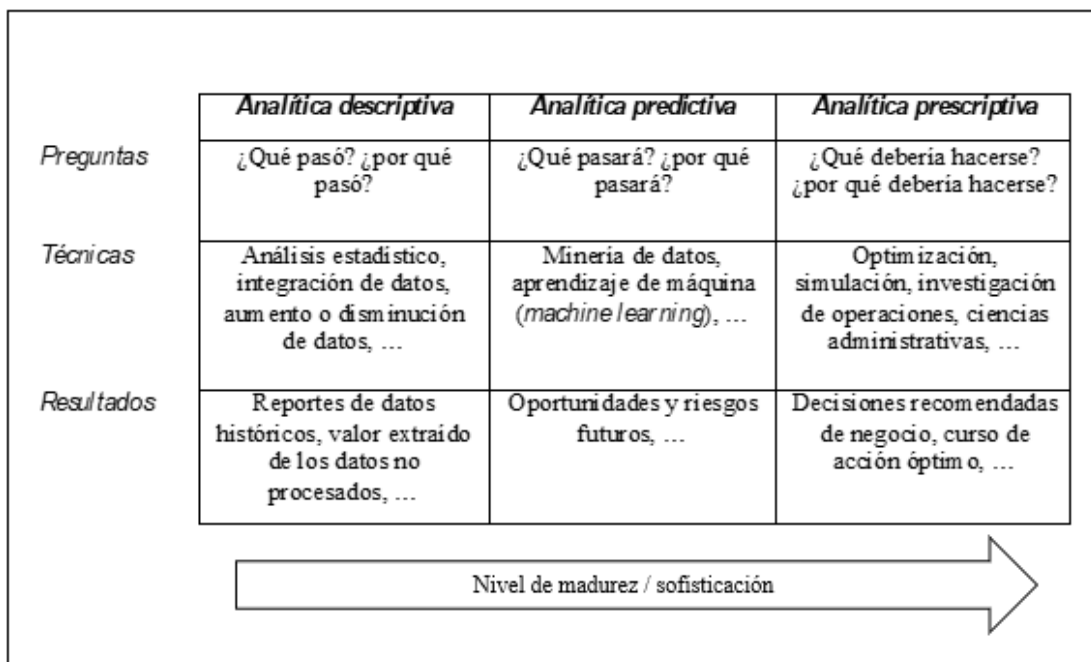


Ilustración 2-1: Tipos de analítica de datos

Fuente: Denle, 2015; Soltanpoor & Sellis, 2016, pp.245-256

Realizado por: Barragán, H., 2023

A pesar de que el conocimiento en los campos de analítica descriptiva y predictiva ha evolucionado durante los últimos años, la analítica prescriptiva todavía sigue siendo objeto de investigación y desarrollo (Lepenioti et al., 2020, pp.57-70). Los sistemas de apoyo a la toma de decisiones no incorporan necesariamente todos los niveles de analítica; esto dependerá del propósito del sistema y el alcance que tendrá en funcionamiento.

2.3.3. *Analítica de datos en el ámbito educativo*

Debido a la pandemia de Covid-19, el uso de sistemas gestores de aprendizaje o *learning management systems* se popularizó, provocando que muchas instituciones trasladaran sus cursos a plataformas como Blackboard o Moodle (Guzmán-Valenzuela et al., 2021, p.23). Esto dio paso a que las instituciones educativas reconozcan la necesidad de implementar plataformas tecnológicas que permitan una transformación de los modelos de enseñanza y aprendizaje. Sin embargo, la gran cantidad de datos que generan hacen necesario su análisis para encontrar verdadero valor en la información (Paz y Cazella, 2019, pp.31710-31716). La analítica de datos en el ámbito educativo se desarrolla en los campos de *learning analytics* y *academic analytics*.

Según Siemens y Long (2011, pp.31-40), los conceptos de *learning analytics* (analíticas del aprendizaje) y *academic analytics* (analíticas académicas) hacen referencia al enfoque actual de análisis de datos en las instituciones educativas. Aunque estos términos han sido utilizados indistintamente, lo cierto es que cuentan con enfoques diferentes. Mientras que *learning analytics* está centrado en la aplicación de técnicas analíticas para mejorar el desempeño en el aprendizaje de los estudiantes, *academic analytics* está dirigido a niveles administrativos y tiene un alcance institucional (Clow, 2013, pp.683-695; Siemens y Long, 2011, pp.31-40).

Learning analytics es un campo cuyo interés principal corresponde a incrementar el éxito en el aprendizaje y la mejora del proceso de enseñanza a través de un análisis personalizado de los datos académicos de los estudiantes (Leitner et al., 2017, pp.1-23). Un término relacionado es *educational data mining* (minería de datos educativa), frecuentemente abreviado EDM. Este concepto, aunque similar a *learning analytics*, se diferencia en que está enfocado en generar modelos para el desarrollo de sistemas de tutorías inteligentes, mientras que LA tiende a apoyar a docentes y administradores en la de toma de decisiones relativas al proceso de aprendizaje (Guzmán-Valenzuela et al., 2021, p.23). Las aplicaciones de *learning analytics* están relacionadas con los modelos predictivos y son bastante amplias. Sin embargo, su uso más importante es la estimación de la probabilidad de que un estudiante complete un curso determinado, permitiendo a docentes y administrativos mejorar las tasas de graduación (Clow, 2013, pp.683-695).

En cuando a *academic analytics* el enfoque es distinto; mientras LA trabaja con datos obtenidos del proceso de aprendizaje de los estudiantes, *academic analytics* podría considerarse como la aplicación de inteligencia del negocio al ámbito educativo, enfatizando el análisis a nivel institucional, regional e inclusive internacional (Siemens y Long, 2011, pp.31-40). Para Santos et al. (2021, pp.997-1003), el concepto de *academic analytics* comprende un conjunto de acciones que facilitan la toma de decisiones en todos los niveles, permitiendo a las instituciones educativas construir su propio conocimiento e inclusive mejorar su reputación. Esto puede ser expresado a través de indicadores como la tasa de retención, aceptación, eliminación de las diferencias de género, inserción en la comunidad, utilización eficiente del presupuesto, entre otros.

Todos estos términos convergen en un único punto: la aplicación de técnicas analíticas en diferentes niveles institucionales. Sin embargo, no se ha llegado a un acuerdo claro sobre su uso, especialmente porque se trata de conceptos relativamente nuevos, dando origen a confusiones (Siemens y Long, 2011, pp.31-40; Paz y Cazella, 2019, pp.31710-31716).

2.3.4. Métodos y técnicas de analítica de datos y su aplicación el ámbito educativo

En el campo de analítica de datos, se han desarrollado diversos métodos y técnicas que permiten encontrar valor en la información, las cuales se aplican en varios sectores, como por ejemplo, la educación. A continuación, se describen los más importantes.

Predicción

Consiste en crear un modelo para inferir una característica de los datos a partir de otras variables conocidas como predictoras. La predicción puede dividirse en clasificación, regresión y estimación de densidad, dependiendo de si la variable a identificar es categórica, continua o una función de densidad de probabilidad, respectivamente. En el campo de EDM, las técnicas populares para clasificación incluyen los árboles de decisión, regresión logística y máquinas de vectores de soporte; para regresión existen la regresión lineal, redes neuronales y regresión de máquina de vectores de soporte; en la estimación de densidad se utilizan, por ejemplo, las funciones Gaussianas (Baker, 2010, pp.112-118).

Descubrimiento de estructuras

Los algoritmos dentro de esta categoría se utilizan para encontrar estructuras en los datos sin una idea previa de lo que se desea obtener. Dos técnicas populares dentro de EDM y LA para el descubrimiento de estructuras son el *clustering*, donde se busca agrupar puntos de datos comunes dentro de *clusters*, y el análisis de redes sociales, para identificar patrones entre interacciones de individuos (Baker y Siemens, 2014, pp.253-272). Algunos algoritmos para *clustering* son *k-means*, esperanza-maximización (EM) y DBSCAN (Mittal et al., 2019, pp.1-14). En el análisis de redes sociales se emplean técnicas como el estudio de grafos, detección de comunidades, perfilado de usuarios, entre otros (Camacho et al., 2020, pp.88-120).

Detección de valores atípicos

La detección de valores atípicos o anomalías consiste en la identificación de puntos de datos que resultan significativamente distintos a los demás. Sus aplicaciones más importantes van desde la limpieza de *datasets* hasta la identificación de fraudes bancarios. Esta técnica puede ejecutarse con diferentes enfoques y algoritmos, como *k-nearest*, *clustering*, regresión, GMM, aprendizaje profundo, entre otros (Wang et al., 2019, pp.107964-108000).

Minería de relaciones

El objetivo de esta técnica es descubrir relaciones entre variables en un conjunto de datos. De forma general, puede dividirse en minería de reglas de asociación, correlacional, de patrones secuenciales y de datos causal. Algunos algoritmos utilizados dentro de la minería de relaciones son ECLAT, FP-Growth, GSP, A priori, SPADE, LCD, entre otros (Aleem y Gore, 2020, pp.182-188).

Minería de procesos

En el contexto de la analítica de datos educativa, se refiere al conjunto de técnicas y métodos enfocados en la extracción del conocimiento de los datos generados por procesos en los sistemas de tipo gestores del aprendizaje (LMS) (Juhaňák et al., 2019, pp.496-506). Se realiza a través de algoritmos como *heuristic miner* y *fuzzy miner*, incluso aplicando técnicas como el *clustering* (dos Santos Garcia et al., 2019, pp.260-295).

Minería de texto y procesamiento del lenguaje natural

Esta técnica permite derivar información de alta calidad proveniente de textos. Algunas tareas comunes incluyen la categorización, *clustering*, extracción de entidades/conceptos, producción de taxonomías granulares, análisis de sentimientos, resumen de documentos y modelado de relaciones de entidades (Romero y Ventura, 2013, pp.12-27). Para ello, se utilizan técnicas como los clasificadores bayesianos, árboles de decisión, máquinas de vectores de soporte, visualización, entre otros (Gaikwad et al., 2014).

Destilación de datos

Es un enfoque para la presentación de los datos, de tal manera que las personas puedan realizar inferencias sobre ellos. Sus dos objetivos principales son la identificación de la información que facilite el reconocimiento de patrones en los datos y la clasificación de estos para la construcción posterior de modelos analíticos (Baker, 2010, pp.112-118).

Aprendizaje de máquina

El *machine learning* es la ciencia de programar computadores para que puedan aprender de los datos. Un enfoque común para dividir el aprendizaje de máquina es de acuerdo con la intervención humana en el entrenamiento de los modelos, pudiendo ser de aprendizaje supervisado o no supervisado. *K-nearest*, regresión lineal, regresión logística, máquinas de vectores de soporte, árboles de decisión y redes neuronales son algunos de los algoritmos utilizados en el aprendizaje supervisado, mientras que otros como *k-means*, DBSCAN, *isolation forest*, PCA y Apriori se utilizan en aplicaciones de aprendizaje no supervisado (Géron, 2019).

Descubrimiento con modelos

En este enfoque, se desarrolla un modelo mediante técnicas como predicción o *clustering*, para ser utilizado como componente en un nuevo análisis. La minería de relaciones es un ejemplo de las actividades que pueden realizarse empleando esta técnica; se puede utilizar para encontrar la conexión entre las predicciones de un modelo existente y otras variables adicionales en un *dataset* (Baker, 2010, pp.112-118).

Estadística

Es un campo que provee métodos para derivar valor de los datos, facilitando la toma de decisiones informadas. Estos se agrupan en dos grandes ramas: la estadística descriptiva, que incluye formas de organizar y resumir la información; y la estadística inferencial, que involucra la generalización de una característica de la población a partir de una muestra (Peck et al., 2019). En la estadística descriptiva se utilizan técnicas como el análisis de distribuciones de frecuencia, medidas de tendencia central y medidas de dispersión; la estadística inferencial emplea estimaciones y pruebas de hipótesis (Salvatore y Reagle, 2011).

La **Tabla 2-1** muestra las aplicaciones de las técnicas descritas en el ámbito educativo.

Tabla 2-1: Técnicas de analítica aplicadas en el contexto de la educación

<i>Métodos/Técnica</i>	<i>Aplicación</i>
Predicción	Predicción del rendimiento y detección del comportamiento de los estudiantes
<i>Clustering</i>	Agrupamiento de estudiantes similares con relación a sus patrones de aprendizaje e interacción
Detección de valores atípicos	Detección de estudiantes con dificultades o procesos de aprendizaje irregular
Minería de relaciones	Identificación de relaciones en el comportamiento de los estudiantes y diagnóstico de dificultades en el aprendizaje
Análisis de redes sociales	Interpretación de relaciones en actividades colaborativas e interacción con herramientas de comunicación
Minería de procesos	Representación de comportamiento de los estudiantes a través de sus notas en un curso y momentos determinados
Minería de texto y procesamiento del lenguaje natural	Análisis del contenido de foros, chats, páginas web y documentos
Destilación de datos	Apoyo para que los docentes puedan visualizar y analizar las actividades en curso y el uso de la información

<i>Métodos/Técnica</i>	<i>Aplicación</i>
Aprendizaje de máquina	Encontrar valor escondido en los datos de forma automática mediante modelos
Descubrimiento con modelos	Identificación de relaciones entre el comportamiento de los estudiantes y variables contextuales
Estadística	Análisis e interpretación de datos cuantitativos para la toma de decisiones

Fuente: Leitner et al., 2017, pp.1-23

Realizado por: Barragán, H., 2023

Aunque Leitner et al. (2017, pp.1-23) se refiere a estos enfoques como “técnicas”, autores como Romero et al. (2020) y Khalil et al. (2015, pp.1326-1336) los denominan “métodos”, pues en la literatura no hay una distinción clara entre ambos términos.

2.4. Metodología ágil Scrum

Scrum es un marco de trabajo que apoya a las personas, equipos y organizaciones en la generación de valor mediante soluciones sencillas a problemas complejos. Se caracteriza por ser ligero, simple, iterativo e incremental (Schwaber y Sutherland, 2020). Scrum, por sí solo, no brinda las pautas técnicas para el desarrollo de software ágil, sino que su enfoque está en la administración del proceso. Esta característica permite que los equipos de trabajo puedan elegir libremente las prácticas de programación a utilizar, haciendo del desarrollo un proceso flexible y adaptativo (Sommerville, 2011). El marco de trabajo Scrum está incompleto de forma deliberada, ya que se busca que los equipos lo “construyan” conforme avance el trabajo; lejos de brindar instrucciones detalladas sobre cómo llevar el proceso, pretende guiar las relaciones e interacciones entre las personas (Schwaber y Sutherland, 2020). Según Sommerville (2011), el proceso de Scrum consta de tres fases: la planificación del bosquejo, los ciclos *sprint* y el cierre del proyecto. La **Ilustración 2-2** muestra un vistazo general al proceso de desarrollo y elementos de Scrum, que serán detallados más adelante.

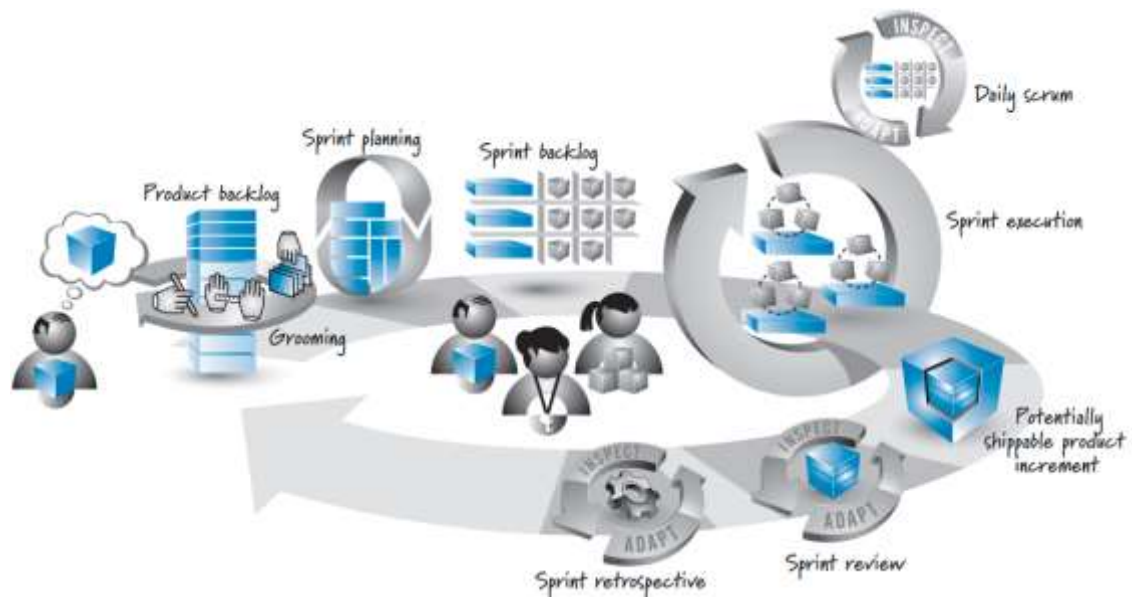


Ilustración 2-2: El marco de trabajo Scrum

Fuente: Rubin, 2012

Scrum está enfocado en el desarrollo de proyectos de pequeña y mediana escala, interviniendo para ello grupos con pocas personas (usualmente diez o menos) que trabajan independientemente. Se ha demostrado que los equipos organizados de esta forma tienden a trabajar de manera más efectiva (Rubin, 2012; Schwaber y Sutherland, 2020; Rising y Janoff, 2000, pp.26-32).

2.4.1. Elementos de Scrum

En el marco de trabajo Scrum, un equipo conformado por personas que desempeñan distintos roles, generan artefactos a lo largo de varios eventos. Todos estos factores que intervienen en el proceso de desarrollo de un proyecto son conocidos como los elementos de Scrum. Es importante identificar correctamente los roles de cada miembro del equipo, los artefactos que generarán y los eventos que se ejecutarán.

2.4.1.1. Equipo y roles

Un grupo pequeño de personas conocidas como *Scrum team* son los encargados de llevar a cabo el proyecto; lo conforman un *Scrum Master*, un *Product Owner* y los desarrolladores (Schwaber y Sutherland, 2020). El *Scrum Master* es el encargado de ayudar al resto de personas en el equipo a entender y apropiarse de los valores, prácticas y principios de Scrum. Sus responsabilidades incluyen el entrenar a los otros miembros respecto a prácticas ágiles, tener autoridad en el proceso, proteger al equipo de distracciones, remover obstáculos, solventar problemas, entre otras (Rubin, 2012). El *Product Owner* se encarga de maximizar el valor del producto resultante del trabajo del

Scrum team (Schwaber y Sutherland, 2020). Constituye el punto de conexión entre los *stakeholders* y el *Scrum team*, ya que sus responsabilidades son, en primer lugar, comprender las necesidades y prioridades de los *stakeholders*, para posteriormente comunicar al equipo qué debe desarrollar y en qué orden (Rubin, 2012). Los equipos de desarrolladores pueden ser multidisciplinarios, es decir, incluir varios roles como programadores, *testers*, administradores de bases de datos, arquitectos de software, administradores de bases de datos, y demás; lo importante es contar con las personas capaces de conseguir el producto final deseado (Rubin, 2012).

2.4.1.2. Artefactos

Scrum cuenta con tres artefactos: *product backlog*, *sprint backlog* y los incrementos regulares (Sutherland y Coplien, 2019). El *product backlog* es un listado de las actividades necesarias a ejecutarse para mejorar el producto, es decir, los requerimientos. Cuando dichas actividades son seleccionadas para ser parte de un *sprint*, forman el *sprint backlog* junto con las metas del *sprint* y el plan de ejecución. Los resultados del desarrollo de software en una iteración se traducen en incrementos, que son aportes concretos y funcionales al producto final (Schwaber y Sutherland, 2020).

2.4.1.3. Eventos

El evento más importante y que constituye el corazón de Scrum es el *sprint*. Un *sprint* es un evento, ciclo o iteración de longitud fija, usualmente entre una semana y un mes, en la que el *Scrum team* transforma las ideas en valor, entregando como resultado un potencial incremento al producto final (Schwaber y Sutherland, 2020; Rubin, 2012). El punto de partida de un *sprint* es la revisión del *product backlog*, que debe ser analizado para asignar prioridades y riesgos a las actividades; este evento es conocido como *sprint planning* (Sommerville, 2011). Los miembros del equipo seleccionan las actividades a ejecutar, establecen un tiempo fijo y determinan las metas a cumplir, dando como resultado el *sprint backlog* de la iteración (Rubin, 2012). Posteriormente inicia la ejecución del *sprint*, que consiste en el desarrollo del software en sí. Se realizan reuniones diarias o *daily scrum*, cuya duración no supera los quince minutos y tienen como objetivo involucrar a todo el equipo en el aporte de ideas que permitan mejorar el proceso para alcanzar las metas de la iteración (Rubin, 2012; Schwaber y Sutherland, 2020; Sommerville, 2011). Al finalizar cada *sprint*, se llevan a cabo dos eventos más; un *sprint review*, donde el equipo Scrum se reúne con los *stakeholders* y determinan los ajustes necesarios en futuros *sprints* y el paso siguiente a ejecutar; y un *sprint retrospective*, donde se planifican nuevas formas de incrementar la calidad y efectividad del equipo (Rising y Janoff, 2000, pp.26-32; Schwaber y Sutherland, 2020). Finalmente, el proceso se repite, dando paso al siguiente *sprint*.

2.5. Metodología CRISP-DM

CRISP-DM, siglas para *Cross-Industry Standard Process for Data Mining*, es un modelo de proceso independiente de la industria, diseñado originalmente para la minería de datos; consta de seis fases iterativas, que van desde la comprensión del negocio hasta el despliegue de la solución final (Schröer et al., 2021, pp.526-534). El modelo de referencia CRISP-DM brinda “una visión general del ciclo de vida de un proyecto de minería de datos, contiene las fases con sus objetivos, las tareas y las relaciones entre éstas, y las instrucciones paso a paso que se deben llevar a cabo” (Cobos et al., 2010, pp.45-56).

2.5.1. Fases del proceso

La **Ilustración 2-3** muestra las fases del proceso en el modelo de referencia CRISP-DM.

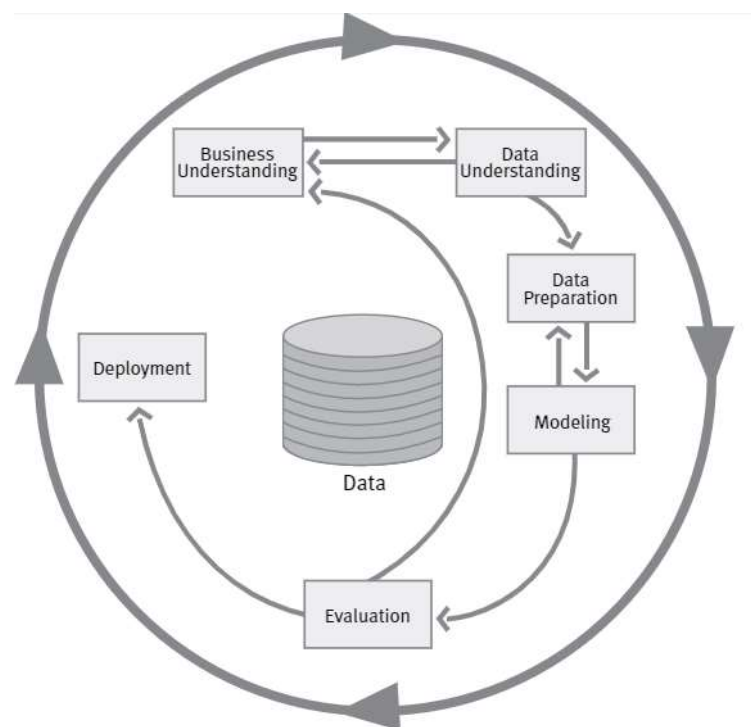


Ilustración 2-3: Fases del modelo referencial CRISP-DM

Fuente: Chapman et al., 2000

Una característica fundamental señala por Chapman et al. (2000) es que la secuencia del proceso no es rígida, sino que necesariamente se tendrá que mover entre fases de acuerdo con los resultados obtenidos en cada una; de ahí que la ilustración muestre la sucesión de pasos de forma cíclica.

2.5.1.1. Comprensión del negocio

La primera fase consiste en la comprensión del problema desde una perspectiva del negocio, para posteriormente convertirlo en una solución en términos de minería de datos y elaborar un plan inicial para alcanzar las metas deseadas. Las tareas en esta fase son: determinar los objetivos del negocio, evaluar la situación, determinar los objetivos de la minería de datos y producir el plan del proyecto (Chapman et al., 2000).

2.5.1.2. Comprensión de los datos

La fase de comprensión del negocio consiste en determinar los datos disponibles y que puedan ser utilizados en el proceso de minería, realizando además una exploración para entender la información e identificar posibles problemas (Putler y Krider, 2012, pp.17-29). Las tareas genéricas de esta fase son: recolectar, describir, explorar y verificar la calidad de los datos (Chapman et al., 2000).

2.5.1.3. Preparación de los datos

La siguiente fase comprende la construcción del conjunto de datos final, lo que implica seleccionar, limpiar e incluso transformar los datos crudos que vayan a ser utilizados en el análisis. La selección de los datos es un paso fundamental en esta fase, pues se establecen qué variables pueden ser explotadas en los modelos de minería de datos, así como la naturaleza de los registros para el análisis (Putler y Krider, 2012, pp.17-29).

2.5.1.4. Modelado

Esta fase consiste en la selección de varias técnicas de modelado, que son ajustadas para cada problema y posteriormente aplicadas. Las actividades por realizar son la selección de técnicas de modelado, generación de diseño de pruebas, construcción y evaluación del modelo (Chapman et al., 2000).

2.5.1.5. Evaluación

En esta fase se evalúa no solo el rendimiento del modelo, sino también otros aspectos como requerimientos no abordados o posibles efectos adversos asociados a su uso. Está conformado por la evaluación de los resultados, la revisión del proceso y la determinación del siguiente paso a ejecutar (Putler y Krider, 2012, pp.17-29).

2.5.1.6. Despliegue

La fase final es el despliegue de los resultados obtenidos. Como mencionan Chapman et al. (2000), incluso cuando el propósito del modelo es incrementar el conocimiento obtenido de los datos, este conocimiento debe ser presentado de alguna manera al cliente de tal forma que pueda utilizarlo. La naturaleza de esta fase variará dependiendo del tipo de proyecto, sin embargo, CRISP-DM establece cuatro actividades genéricas a ejecutar: planificación del despliegue, planificación del monitoreo y mantenimiento, creación de un reporte final y revisión del proyecto (Putler y Krider, 2012, pp.17-29).

2.6. Herramientas y lenguajes de desarrollo

El desarrollo de una aplicación puede ser acelerado al utilizar las herramientas adecuadas a cada contexto. A continuación, se presenta el lenguaje de programación y *framework* a emplear en el desarrollo de los modelos de analítica de datos y en la aplicación web.

2.6.1. Python

Se trata de un lenguaje de programación moderno y *open source*, que soporta los paradigmas orientado a objetos, funcional e imperativo. Su facilidad de uso y legibilidad lo han popularizado entre principiantes, pues es posible escribir programas con menos líneas de código en comparación a otros lenguajes como Java o C++ (Kelly, 2016, pp.3-5). Python es uno de los lenguajes de programación que mejor se ajustan a tareas de manipulación de datos, especialmente en el contexto de la ingeniería (Salihoun, 2020, pp.58-76). Las aplicaciones de Python en el campo de la computación científica y numérica son bastante amplias. Como se detalla en su sitio web oficial⁴, puede utilizarse en aplicaciones dentro de estos campos gracias a que soporta potentes librerías como SciPy, que cuenta con algoritmos y estructuras de datos de uso recurrente, y Pandas, una herramienta para el análisis y la manipulación de datos. Además, muchas de estas librerías para ciencia de datos están construidas sobre capas subyacentes de Fortran y C, lo que acelera los cálculos con vectores y *arrays* multidimensionales, mejorando el rendimiento (Raschka et al., 2022).

⁴ www.python.org

2.6.2. *Streamlit*

Streamlit es un *framework* de código abierto para Python, que permite desarrollar y desplegar aplicaciones web para ciencia de datos y modelos de aprendizaje de máquina (McDonald, 2022). En su sitio web oficial⁵ describen sus características principales, las cuales son:

- Compatibilidad con una gran cantidad de librerías como TensorFlow, Pandas, PyTorch, OpenCV, NumPy, Scikit-learn, entre otras;
- capacidad para crear aplicaciones web con pocas líneas de código y sin la necesidad de emplear otros lenguajes como HTML o JavaScript;
- despliegue inmediato y gratuito de aplicaciones; y
- variedad de plantillas para crear aplicaciones de ciencia de datos de forma inmediata.

2.7. **Éxito de los sistemas de información**

La medición del éxito de los sistemas de información resulta crítica en la comprensión del valor que tiene su gestión e inversión (DeLone y McLean, 2003, pp.9-30). Sin embargo, como afirman Nguyen et al. (2015, pp.242-256), no existe una única definición para “éxito” en este contexto, pues varía desde la perspectiva de cada persona: lo que para un desarrollador puede considerarse exitoso (el producto funcionando correctamente o terminado dentro de lo planificado), para el cliente o usuario puede ser algo distinto (un sistema que mejore su satisfacción o rendimiento). Así mismo, lejos de existir una única medida para determinar el éxito de un sistema, hay tantas como estudios relacionados, principalmente por la naturaleza de la información, que puede ser cuantificada en diferentes niveles (DeLone y McLean, 1992, pp.60-95).

2.7.1. *Modelo de éxito de los sistemas de información*

Para DeLone y McLean (1992, pp.60-95), el éxito de un sistema de información no se mide a través de una única variable, sino como un constructo multidimensional que interrelaciona siete dimensiones, como se muestra en la **Ilustración 2-4**. Su estudio se apoya en el trabajo de Shannon y Weaver (1949), quienes aseguraban que la información podía ser medida en tres niveles: técnico, semántico y efectivo. En este sentido, la información fluye a través de estas etapas, desde su producción hasta que los usuarios la consumen, lo que sugiere que las medidas de éxito de un sistema deberían separarse por cada nivel (DeLone y McLean, 1992, pp.60-95).

⁵ www.streamlit.io

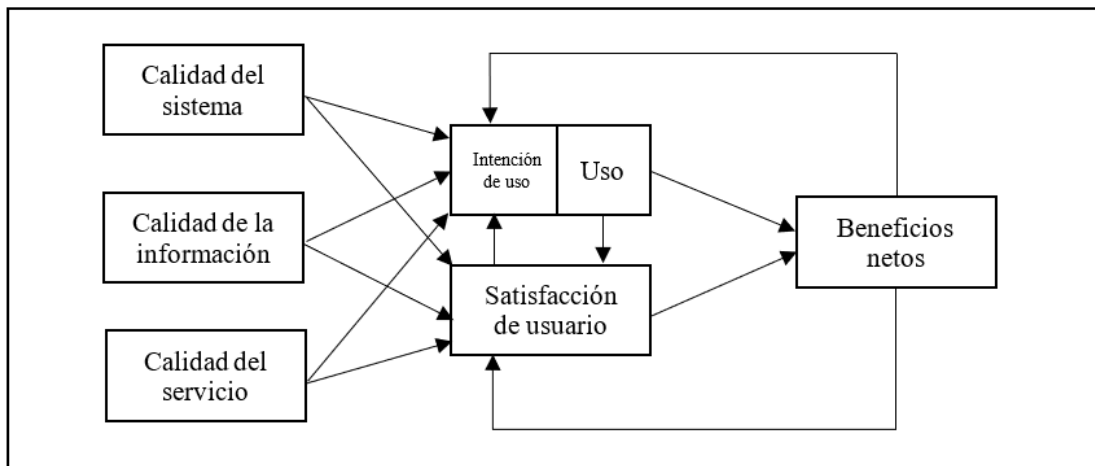


Ilustración 2-4: Modelo de éxito de los sistemas de información

Fuente: DeLone & McLean, 2003, pp.9-30

Realizado por: Barragán, H., 2023

Los autores del modelo resaltan el impacto de la calidad del sistema, de la información que provee, y del servicio que prestan, en la intención de uso y la satisfacción del usuario. El uso del sistema influye en la satisfacción del usuario, que a su vez, impacta en la intención de su uso. Como consecuencia, se generan beneficios netos que también influenciarán en la intención de uso y la satisfacción del usuario (DeLone y McLean, 2003, pp.9-30).

A continuación, se presenta la definición de cada una de las dimensiones del modelo propuesto por DeLone y McLean.

Calidad del sistema

Corresponde a las características deseadas en un sistema de información y que incluyen, por ejemplo, la facilidad de uso, flexibilidad, fiabilidad, facilidad de aprendizaje, sofisticación y tiempos de respuesta (Petter et al., 2008, pp.236-263). La calidad del sistema empata con el nivel técnico propuesto por Shannon y Weaver, pues aquí se mide la precisión y eficiencia del sistema que produce la información (DeLone y McLean, 1992, pp.60-95).

Calidad de la información

Esta dimensión se enfoca en la calidad de las salidas que arroja un sistema, es decir, de la información que genera, comúnmente presentada en forma de reportes (DeLone y McLean, 1992, pp.60-95). Algunas de las características deseadas en la información son la relevancia, comprensibilidad, precisión, consistencia, completitud, pertinencia en el tiempo y usabilidad (Petter et al., 2008, pp.236-263).

Calidad del servicio

La incorporación de esta dimensión en el modelo actualizado corresponde al papel que toman las empresas como proveedores de servicio y no únicamente como desarrolladores de productos (DeLone y McLean, 2003, pp.9-30). En esta se mide la calidad del soporte que reciben los clientes por parte de la empresa, considerando la capacidad de respuesta, precisión, fiabilidad, competencia técnica y empatía del personal (Petter et al., 2008, pp.236-263).

Intención de uso / uso

El grado en el que los usuarios utilizan las funcionalidades del sistema de información (Petter et al., 2008, pp.236-263). Originalmente, solo se consideraba el uso, sin embargo, debido a la malinterpretación de esta variable, se optó por añadir la “intención de uso” como una alternativa en determinados contextos (DeLone y McLean, 2003, pp.9-30). Algunas medidas para esta dimensión son la cantidad, frecuencia, naturaleza, idoneidad, alcance y propósito del uso (Petter et al., 2008, pp.236-263).

Satisfacción de usuario

La satisfacción de usuario es un criterio importante para medir el éxito de un sistema de información (Zviran y Erlich, 2003, p.5). Se define como el nivel de satisfacción del usuario con los reportes, sitios web y servicios de soporte (Petter et al., 2008, pp.236-263). Sin embargo, esta definición puede ampliarse para incluir cualquier funcionalidad de un sistema, como el concepto brindado por la norma ISO/IEC 25010 (2011), que la define como el “grado en que se satisfacen las necesidades del usuario cuando un producto o sistema se utiliza en un contexto de uso específico”. La satisfacción de usuario puede medirse con uno o varios ítems, de forma global o específica (satisfacción en la toma de decisiones, respecto al software, respecto a la información, etc.) (DeLone y McLean, 1992, pp.60-95).

Beneficios netos

Esta dimensión extiende los beneficios individuales y organizacionales a cualquier elemento al que pueda influenciar el éxito del sistema de información, como grupos, industrias, incluso naciones; la mejora en la toma de decisiones, productividad, ventas, ingresos, reducción de costos, eficiencia del mercado, creación de empleo, desarrollo económico, entre otros, son medidas para cuantificar los beneficios netos (Petter et al., 2008, pp.236-263).

2.8. Trabajos relacionados

Algunos trabajos que han servido como apoyo para el estudio incluyen el desarrollado por Moscoso-Zea et al. (2019, pp.4-13), quienes determinaron qué métodos existen en el campo de la

minería de datos educativa para estimar la tasa de graduación de estudiantes. Posteriormente, construyeron un modelo de DM (*data mining*) para encontrar patrones y tendencias en los datos respecto a la tasa de graduación. Para la investigación, los autores contaban con un conjunto de datos provenientes del *Data Warehouse* de una institución de educación superior y correspondientes a los estudiantes de la carrera de Ciencias de la Computación entre los años 2002 y 2015. Los datos incluían información académica y personal, destacando el campo “Graduación”, que sirvió como la clase de análisis. El método utilizado para el proceso de descubrimiento de conocimiento fue “*Knowledge Discovery in Databases*” (KDD), mientras que CRISP-DM se empleó para la experimentación con minería de datos.

Dewantoro et al. (2020, pp.105-109) aplicaron técnicas de minería de datos educativa para construir modelos predictivos que apoyen la toma de decisiones respecto a la admisión de nuevos estudiantes en una universidad. Se consideró la duración en la carrera y el GPA (*grade point average*) como indicadores para determinar la probabilidad de que un alumno complete sus estudios. Estos valores se estimaron utilizando técnicas de minería de datos, considerando el promedio final, la especialidad obtenida y la localización del colegio en el que se graduó el postulante como variables de análisis. Entre los modelos evaluados, las redes neuronales artificiales resultaron ser el mejor predictor.

Otro trabajo similar es el de Esquivel & Esquivel (2021, pp.50-54), quienes desarrollaron un modelo predictivo utilizando regresión logística para estimar las opciones de matriculación de los estudiantes. Con los resultados obtenidos se planteó la creación de un sistema de apoyo a la toma de decisiones que sería empleado por los *stakeholders* de la universidad, el cual fue desarrollado utilizando R y R-Shiny.

Kim et al. (2012, pp.35-43) evaluaron la satisfacción de usuario respecto a un sistema de apoyo a la toma de decisiones clínico que ayudaba a evitar errores a la hora de dispensar medicamentos. Se utilizó una adaptación del modelo de DeLone et al. (1992, pp.60-95), tomando en cuenta tres características como medida de éxito: calidad del sistema, calidad de la información y factor de apoyo. Posteriormente, se determinó estadísticamente cómo influyen estos componentes en la satisfacción de usuario de acuerdo con dos dimensiones: satisfacción del sistema y satisfacción de la información. Además, los autores compararon los resultados obtenidos entre los grupos de profesionales encuestados, que incluían doctores, enfermeros y farmacéuticos.

CAPÍTULO III

3. MARCO METODOLÓGICO

El tercer capítulo abarca el diseño de estudio y el desarrollo de la solución empleando la metodología ágil Scrum, que contempla la creación de modelos de analítica de datos utilizando CRISP-DM.

3.1. Diseño de estudio

Se definen los elementos que sirven de apoyo para la ejecución del proyecto técnico: tipo de estudio, métodos, técnicas, fuentes, operacionalización de las variables, planteamiento de hipótesis, población y muestra.

3.1.1. Tipo de estudio

El presente trabajo es un estudio de tipo aplicativo, ya que se implementó una solución que apoya el proceso de toma de decisiones de la Comisión de Titulación; la información que provee la aplicación se obtuvo a partir de técnicas analíticas empleadas en el ámbito educativo. También, es un estudio correlacional, pues se evaluaron las dimensiones de éxito del modelo de DeLone y McLean: calidad del sistema, calidad de la información y uso, para determinar su influencia sobre la satisfacción de usuario.

3.1.2. Métodos, técnicas y fuentes de estudio

Los métodos, técnicas y fuentes empleadas, en relación con los objetivos del presente estudio, se resumen en la **Tabla 3-1** mostrada a continuación.

Tabla 3-1: Matriz de consistencia

Objetivos	Métodos	Técnicas	Fuentes
Seleccionar técnicas de analítica de datos que apoyen a la toma de decisiones de la Comisión de Titulación	Analítico y sintético	Revisión de documentación	Revistas, libros y tesis

Objetivos	Métodos	Técnicas	Fuentes
Aplicar técnicas de analítica de datos para generar información que apoye la toma de decisiones	CRISP-DM	Limpieza de datos, análisis estadístico y procesos ETL	Knime y Python
Desarrollar la aplicación web para la toma de decisiones con base a la información requerida por la Comisión de Titulación	Scrum	Revisión de documentación, historias técnicas, historias de usuario, <i>sprint backlog</i>	Libros, equipo de desarrollo, miembros de la Unidad de Titulación
Evaluar el nivel de satisfacción de los usuarios respecto al uso de la aplicación web para el apoyo a la toma de decisiones	Inductivo y estadístico	Encuesta, PLS-SEM	Cuestionarios, miembros de la Unidad de Titulación, docentes, estudiantes y SmartPLS

Realizado por: Barragán, H., 2023

Método analítico.- Se empleó en el estudio de la pertinencia de las técnicas de analítica de datos que se identificaron mediante revisión de documentación, filtrando aquellas de interés para el trabajo. La búsqueda en fuentes de información permitió descubrir tendencias en los campos de EDM y LA, determinando posibles aplicaciones en el contexto de los requerimientos de la Comisión de Titulación y el desarrollo de la solución.

Método sintético.- Se utilizó este método para recopilar y resumir las técnicas de analítica de datos identificadas y sus aplicaciones en el contexto educativo. El resultado de la revisión teórica permitió disponer de información que debió ser sintetizada con el fin de establecer, de forma concreta, las posibles técnicas a utilizar en el desarrollo de modelos analíticos que fueron implementados en la solución.

CRISP-DM.- Esta metodología se aplicó para generar información de utilidad para la Comisión de Titulación, partiendo de los datos recopilados por sus miembros. Se buscó construir modelos aplicando las técnicas de analítica de datos identificadas, a través de los seis pasos que componen el proceso.

Scrum.- La metodología ágil que orientó el desarrollo de la aplicación web para el apoyo a la toma de decisiones y la implementación de modelos analíticos. El uso de Scrum implicó la identificación de equipos y roles, así como la ejecución de eventos, en los que se llevaron a cabo iteraciones de desarrollo y se generaron los entregables del proyecto.

Método inductivo.- Se empleó para determinar el nivel de satisfacción de usuario, uso, calidad del sistema y de la información de la aplicación web, así como las relaciones entre ellos, partiendo de las evaluaciones individuales de cada miembro de la Comisión de Titulación.

Método estadístico.- Fue utilizado en las pruebas de hipótesis para la validación de las relaciones planteadas en el modelo de DeLone y McLean. Se empleó el modelo de ecuaciones estructurales por el método de mínimos cuadrados parciales (PLS-SEM), que permitió determinar la influencia entre constructos y la validez y fiabilidad de sus indicadores.

3.1.3. *Proceso para la evaluación de la satisfacción de usuario de la aplicación web*

Para el cumplimiento del objetivo relacionado con la evaluación de la satisfacción de usuario respecto al uso de la aplicación, se utilizó el modelo de éxito de los sistemas de información propuesto por DeLone y McLean (2003, pp.9-30). Sin embargo, no se consideró únicamente la satisfacción de usuario, ya que, como mencionan los autores, se deberían combinar sistemáticamente varias métricas de las diferentes categorías del modelo para crear un instrumento de medición comprensible (DeLone y McLean, 1992, pp.60-95). Así mismo, al determinar cómo la aplicación web para el apoyo a la toma de decisiones influye en la satisfacción de usuario, no solo se deberían tomar en cuenta los aspectos técnicos del sistema, sino también los relacionados a la información que este genera.

3.1.4. *Operacionalización de variables*

La **Tabla 3-2** muestra la operacionalización conceptual de las variables que son objeto de estudio, de acuerdo con la pregunta de investigación planteada en la formulación del problema y el modelo a validar.

Tabla 3-2: Operacionalización conceptual

Formulación del problema	Variable	Tipo	Concepto
¿Cómo una aplicación web para el apoyo a la toma de decisiones influye en la satisfacción del usuario?	Satisfacción de usuario	Cualitativa	“Grado en que se satisfacen las necesidades del usuario cuando un producto o sistema se utiliza en un contexto de uso específico” (ISO/IEC 25010, 2011).

Formulación del problema	Variable	Tipo	Concepto
	Uso	Cualitativa	Grado en el que los usuarios utilizan las funcionalidades del sistema de información (Petter et al., 2008, pp.236-263).
	Calidad del sistema	Cualitativa	Las características deseables de un sistema de información y sus medidas (Urbach y Müller, 2012, pp.1-18).
	Calidad de la información	Cualitativa	Las características deseables de las salidas (información) de un sistema (Petter et al., 2008, pp.236-263).

Realizado por: Barragán, H., 2023

Las variables, indicadores, técnicas y fuentes, conformes a la formulación del problema, se presentan en la **Tabla 3-3**.

Tabla 3-3: Operacionalización metodológica

Formulación del problema	Variable	Indicadores	Técnicas	Fuentes
¿Cómo una aplicación web para el apoyo a la toma de decisiones influye en la satisfacción del usuario?	Satisfacción de usuario	Satisfacción general con el sistema	Encuesta	(Iivari, 2005, pp.8-27)
		Satisfacción general con la información		
	Uso	Resolución de problemas		(Doll y Torkzadeh, 1998, pp.171-185)
		Racionalización de decisiones		
	Calidad del sistema	Flexibilidad		(Iivari, 2005, pp.8-27)
		Tiempo de respuesta		
	Calidad de la información	Compleitud		(Lee et al., 2002, pp.133-146)
		Interpretabilidad		
		Relevancia		
		Pertinencia en el tiempo		

Realizado por: Barragán, H., 2023

3.1.5. Instrumento de medición

Se diseñó un cuestionario para la medición de las variables y validación del modelo que considera los indicadores descritos en la operacionalización metodológica. Se adaptaron tres instrumentos: AIMQ, una metodología para la evaluación de la calidad de la información a través de un cuestionario propuesto por Lee et al. (2002, pp.133-146); las medidas multidimensionales para cuantificar el uso de un sistema en contextos organizacionales desarrolladas por Doll y Torkzadeh (1998, pp.171-185); y las métricas de calidad del sistema y satisfacción de usuario empleadas en la validación empírica ejecutada por Iivari (2005, pp.8-27). El cuestionario completo se encuentra documentado en el **ANEXO A**.

3.1.6. Planteamiento de hipótesis

De acuerdo con las dimensiones escogidas en la validación del modelo, se plantearon las siguientes hipótesis:

H_1 = El uso influye positivamente en la satisfacción de usuario

H_2 = La calidad del sistema influye positivamente en la satisfacción de usuario

H_3 = La calidad de la información influye positivamente en la satisfacción de usuario

Las relaciones del modelo por validar se muestran a continuación en la **Ilustración 3-1**.

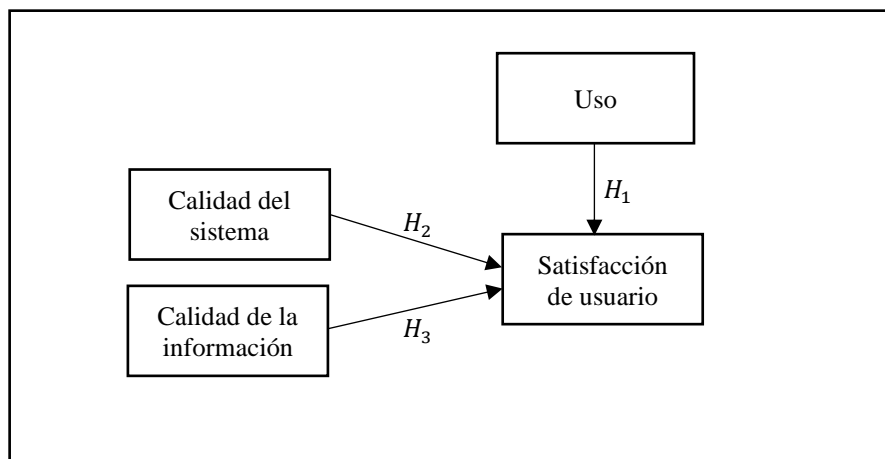


Ilustración 3-1: Hipótesis por validar en el modelo

Realizado por: Barragán, H., 2023

Finalmente, se detallan las hipótesis nulas establecidas para el estudio:

H_{1-0} = El uso influye de forma neutra o negativa en la satisfacción de usuario

H_{2-0} = La calidad del sistema influye de forma neutra o negativa en la satisfacción de usuario.

H_{3-0} = La calidad de la información influye de forma neutra o negativa en la satisfacción de usuario.

3.1.7. Población y muestra

Se ha tomado como población los miembros de la Comisión de Titulación de la carrera de Software, ESPOCH, siendo un total de $N = 3$ usuarios. Sin embargo, debido a que el método estadístico seleccionado para la validación del modelo requiere un mínimo de muestras, se optó por incluir a un total de $N = 16$ usuarios (estudiantes y docentes de la carrera, incluyendo los miembros de la CT). Al ser una población pequeña, se trabajó con toda la población y no únicamente con una muestra.

3.2. Desarrollo de la solución utilizando Scrum

El marco de trabajo Scrum permitió gestionar el desarrollo de la aplicación web “*IMAK*” para el apoyo a la toma de decisiones, así como los modelos de analítica de datos implementados, que se construyeron de forma paralela utilizando CRISP-DM. La descripción del proceso y entregables completos se encuentran documentados en el “Manual técnico – Scrum”, disponible en el **ANEXO B**.

3.2.1. Análisis preliminar

A través de entrevistas a los miembros de la Comisión de Titulación y el análisis de documentos, se pudo establecer los equipos y roles necesarios, así como los requerimientos iniciales del proyecto.

3.2.1.1. Requisitos de la aplicación

Tres fuentes principales fueron consultadas en el proceso de elicitación de requisitos: los miembros de la Comisión, los informes semestrales de los proyectos de integración curricular y los datos recolectados en cada periodo referentes al proceso de titulación. Se realizaron entrevistas a cada uno de sus integrantes, con el fin de conocer su perspectiva respecto a los requerimientos que podrían constituir parte del desarrollo de la solución. Así mismo, utilizando la técnica de revisión de documentación, se analizaron los informes y datos recopilados durante varios

periodos. De esta forma, se pudo establecer la información a incluir en la aplicación web, que se presenta en la **Tabla 3-4** en forma de indicadores clave (KPI).

Tabla 3-4: Indicadores clave

Código	Descripción
KPI-1	Número de trámites/actividades atendidas por la CT.
KPI-2	Cantidad de proyectos de Integración Curricular aprobados por periodo académico.
KPI-3	Número de estudiantes con relación al número de proyectos aprobados por periodo académico.
KPI-4	Duración en meses para el desarrollo de los proyectos
KPI-5	Cantidad de proyectos de Integración Curricular por docentes, según su función y estado actual.
KPI-6	Cantidad de proyectos aprobados, según la modalidad para los Trabajos de Integración Curricular
KPI-7	Número de proyectos concluidos, entre pre-defensas y defensas públicas
KPI-8	Número de proyectos concluidos por periodo académico
KPI-9	Cantidad de proyectos aprobados, por línea de investigación y periodo académico

Realizado por: Barragán, H., 2023

Se identificaron un total de 13 requisitos, que fueron documentados empleado Historias de Usuario y que se resumen en la **Tabla 3-5**.

Tabla 3-5: Historias de usuario

Código	Descripción
HU-01	Desarrollar analíticas de datos
HU-02	Autenticar usuario
HU-03	Visualizar analíticas de datos
HU-04	Visualizar indicadores clave
HU-05	Cargar datos externos
HU-06	Crear indicadores personalizados
HU-07	Modificar tipo de gráfico
HU-08	Modificar título del gráfico
HU-09	Modificar estilo del gráfico
HU-10	Descargar gráficos en formato .png
HU-11	Descargar tablas en formato .xlsx

Código	Descripción
HU-12	Actualizar vistas materializadas
HU-13	Generar plantillas en formato .xlsx

Realizado por: Barragán, H., 2023

Los indicadores clave e historias de usuario detalladas se encuentran documentados al final de este documento en el **ANEXO B**.

3.2.1.2. Módulos de la aplicación

Las funcionalidades de la aplicación web se distribuyen en cuatro módulos, tal y como se muestra en la **Ilustración 3-2**.

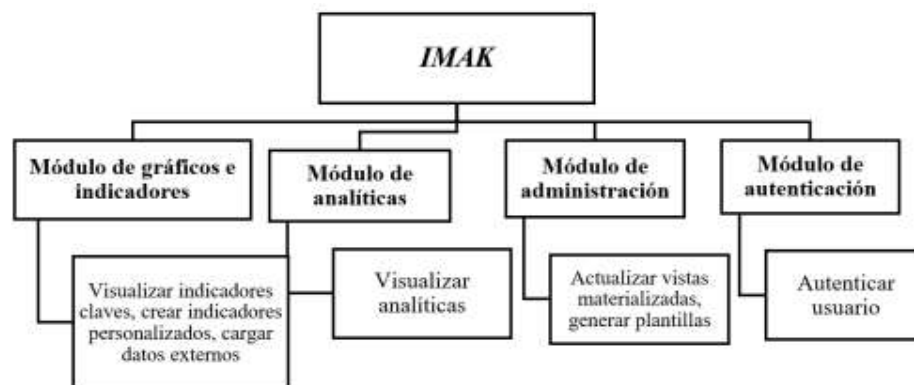


Ilustración 3-2: Módulos de la aplicación web “*IMAK*”

Realizado por: Barragán, H., 2023

Módulo de gráficos e indicadores

El módulo de gráficos e indicadores agrupa las funcionalidades de la aplicación relacionadas con la visualización de indicadores clave en forma de gráficos y tablas, así como la construcción de indicadores personalizados.

Módulo de analíticas

El módulo de analíticas incluye los modelos desarrollados con CRISP-DM y las estadísticas descriptivas que se han propuesto de acuerdo con los requerimientos de Comisión y los datos que esta dispone.

Módulo de administración

El módulo de administración permite actualizar las vistas materializadas que almacenan los datos utilizados en la aplicación. También, incluye la opción de generar plantillas en formato .xlsx para cargar datos externos.

Módulo de autenticación

El módulo de autenticación permite a los usuarios iniciar sesión con las credenciales registradas en el sistema de gestión de trabajos de titulación de la carrera.

3.2.1.3. Diagrama de caso de uso

Se emplearon diagramas de caso de uso para representar el contexto en el que se utiliza la aplicación web “*IMAK*”. La **Ilustración 3-3** presenta el diagrama correspondiente al módulo de gráficos e indicadores.

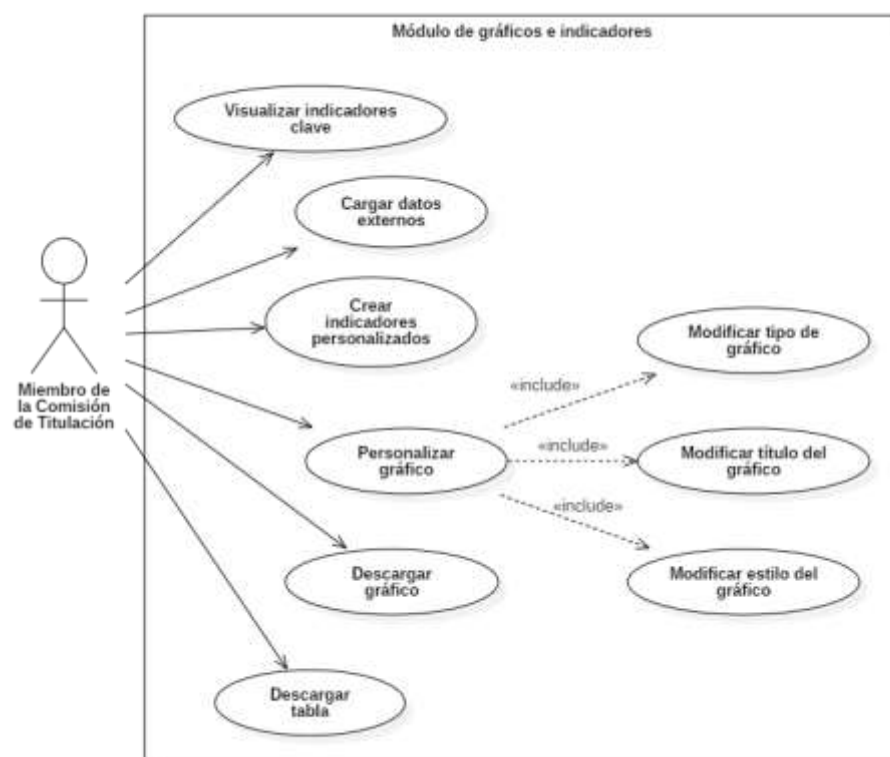


Ilustración 3-3: Contexto de uso del módulo de gráficos e indicadores

Realizado por: Barragán, H., 2023

Los diagramas de caso de uso correspondientes al resto de módulos pueden encontrarse en el **ANEXO B**.

3.2.1.4. Equipo y roles

En la construcción de la aplicación web intervino un grupo de actores conocidos como *Scrum team*, conformados por un *Product owner*, un *Scrum master* y el equipo de desarrollo. La **Tabla 3-6** resume la información de las personas y los roles que desempeñaron dentro del equipo.

Tabla 3-6: Equipo Scrum para el desarrollo del proyecto

Persona	Contacto	Rol	Descripción
Ing. Ivonne Rodríguez	ivonne.rodriguez@epoch.edu.ec	<i>Scrum Master</i>	Encargada de liderar el equipo y promover los valores y principios que guían el proceso de <i>Scrum</i> entre el resto de los integrantes.
Homero Barragán	homero.barragan@epoch.edu.ec	<i>Product Owner</i>	Su papel es servir de punto de conexión entre los intereses de los miembros de la Comisión de Titulación y el <i>Scrum team</i> , de tal forma que pueda canalizar las prioridades y necesidades de ambas partes.
Homero Barragán	homero.barragan@epoch.edu.ec	Equipo de desarrollo	Cumple la tarea de desarrollar la aplicación web de apoyo a la toma de decisiones correctamente.

Realizado por: Barragán, H., 2023

3.2.2. Planificación

En esta sección se describen las actividades ejecutadas respecto a la planificación del proyecto, incluyendo el estudio de factibilidad, análisis de riesgos, estimaciones y organización del desarrollo de las funcionalidades de la solución.

3.2.2.1. Estudio de factibilidad técnica

El estudio de factibilidad técnica establece la posibilidad de ejecución del proyecto al comparar los componentes *software* y *hardware* necesarios con aquellos ya disponibles. El **ANEXO B** muestra el estudio de factibilidad técnica completo para el presente trabajo.

3.2.2.2. Análisis de riesgos

El análisis de riesgos es una actividad que permite identificar y evaluar potenciales situaciones inciertas que tienen un impacto en la ejecución del proyecto. Estas tareas facilitan la creación de planes para prevenir o bien reducir el impacto de estos sucesos fortuito. Se identificaron ocho riesgos utilizando la estructura de descomposición de riesgos (EDR) que se muestra en la **Ilustración 3-4**.

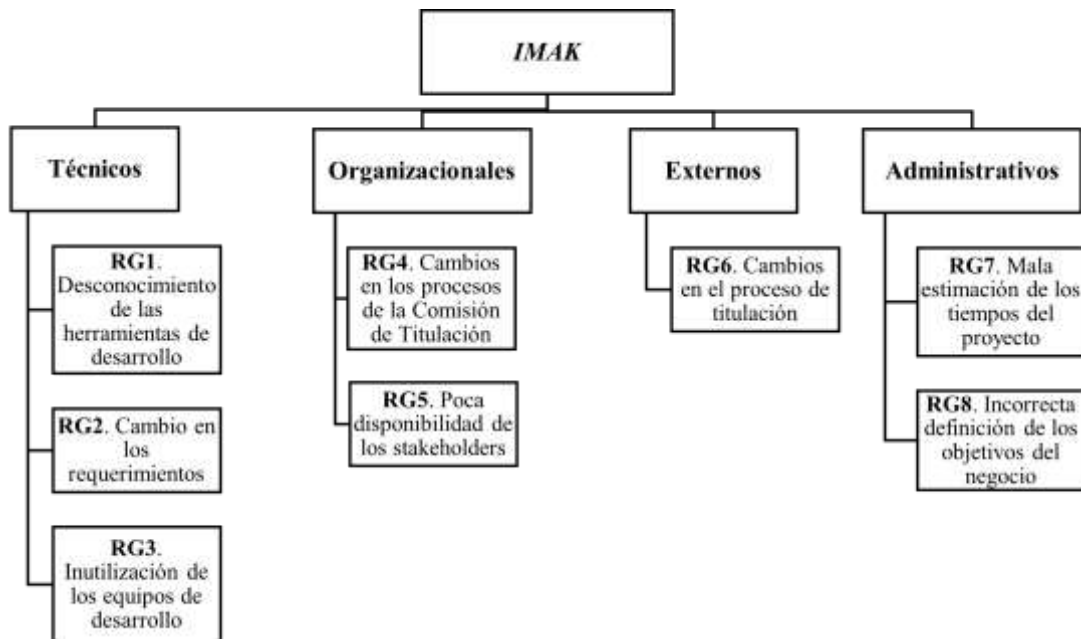


Ilustración 3-4: Estructura de descomposición de riesgos

Realizado por: Barragán, H., 2023

Para visualizar la probabilidad de ocurrencia e impacto de cada riesgo se utilizó una matriz de gravedad del riesgo que se muestra en la **Ilustración 3-5**.

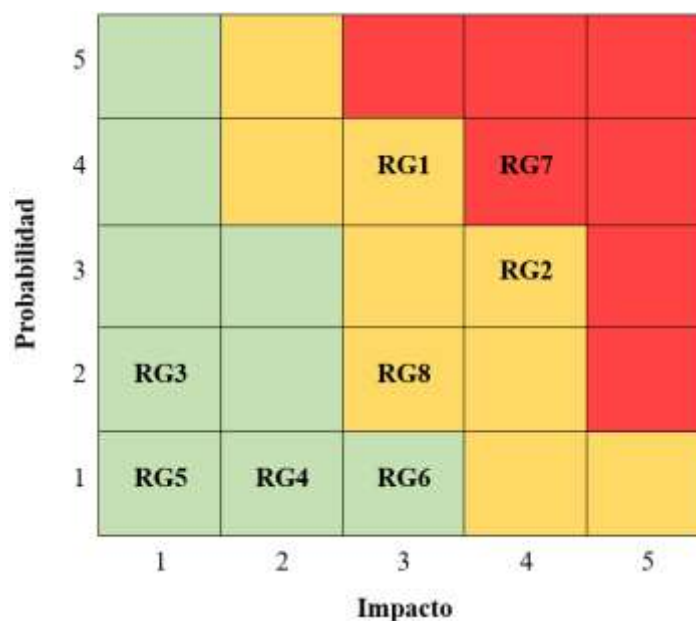


Ilustración 3-5: Matriz de gravedad del riesgo

Realizado por: Barragán, H., 2023

Finalmente, se crearon fichas para identificar causas, consecuencias, estrategias de prevención y/o mitigación, acciones de supervisión y de gestión de los riesgos. La **Tabla 3-7** muestra la ficha de gestión para el riesgo RG1.

Tabla 3-7: Ficha de gestión para el riesgo RG1

Ficha de gestión de riesgos		
ID: RG1	Fecha: 07/05/2023	
Descripción: Desconocimiento de las herramientas de desarrollo		
Probabilidad de ocurrencia: Alta	Nivel de impacto: Moderado	Nivel de gravedad: Moderado
Refinamiento		
Causas:		
<ul style="list-style-type: none"> • Poca familiarización con el <i>framework</i>, librerías y lenguaje de programación. • Falta de práctica de las herramientas de desarrollo. 		
Consecuencias:		
<ul style="list-style-type: none"> • Incremento en los tiempos del proyecto. • Mayor posibilidad de encontrar errores en el código. 		
Acciones de prevención y/o mitigación:		
<ul style="list-style-type: none"> • Capacitación autónoma previo al inicio del proyecto y durante los fines de semana. • Recurrir a foros y ayuda de expertos para solventar dudas respecto a las herramientas. • Búsqueda y selección de herramientas alternas. 		

Ficha de gestión de riesgos		
Supervisión:		
<ul style="list-style-type: none"> Evaluación continua de las habilidades respecto a las herramientas necesarias para el desarrollo del proyecto. 		
Acciones de gestión del riesgo:		
<ul style="list-style-type: none"> Incluir en las estimaciones los tiempos empleados en capacitación y corrección de errores. 		
Estado actual	Fase de reducción iniciada	
	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables:		
Homero Barragán		

Realizado por: Barragán, H., 2023

El **ANEXO B** recoge el análisis de riesgos completo que se ha generado para el desarrollo del proyecto.

3.2.2.3. Product Backlog

El *producto backlog* es la lista priorizada de las funcionalidades deseadas del producto, proporcionando una guía comprensible de lo que se debe construir y en qué orden (Rubin, 2012). Para estimar las actividades a realizar en cada uno de los ítems del *product backlog*, se utilizó la técnica de *T-Shirt Sizing*, cuyas equivalencias con los puntos estimados y las horas de trabajo se han establecido acorde a lo presentado en la **Tabla 3-8**.

Tabla 3-8: Equivalencias para *T-Shirt Sizing*

Talla	Iteración	Puntos estimados	Horas de trabajo
XS	1/10 iteración	1	5
S	1/5 iteración	2	10
M	1/2 iteración	5	25
L	1 iteración	10	50
XL	2 iteraciones	20	100

Realizado por: Barragán, H., 2023

Las estimaciones se realizaron para una sola persona, tomando en cuenta que una jornada laboral representa 5 horas de trabajo y una iteración equivale a dos semanas o 10 días laborales. Con estas consideraciones, se efectuaron los cálculos correspondientes para cada talla.

Posteriormente, se construyó el *producto backlog*, recopilando las estimaciones y priorizando cada historia técnica y de usuario, como se muestra en la **Tabla 3-9**.

Tabla 3-9: *Product backlog*

Código	Descripción	Talla	Puntos	Horas	Prioridad
HT-01	Diseñar la arquitectura de la aplicación	S	2	10	Alta
HT-02	Diseñar la base de datos	XS	1	5	Alta
HT-03	Diseñar las vistas materializadas	S	2	10	Alta
HT-04	Cargar los datos	M	5	25	Alta
HT-05	Establecer el estándar de codificación	XS	1	5	Baja
HT-06	Desplegar la aplicación web.	M	5	25	Media
HU-01	Desarrollar analíticas de datos	XL	20	100	Alta
HU-02	Autenticar usuario	S	2	10	Baja
HU-03	Visualizar analíticas de datos	M	5	25	Media
HU-04	Visualizar indicadores clave	M	5	25	Alta
HU-05	Cargar datos externos	L	10	50	Media
HU-06	Crear indicadores personalizados	M	5	25	Media
HU-07	Modificar tipo de gráfico	XS	1	5	Baja
HU-08	Modificar título del gráfico	XS	1	5	Baja
HU-09	Modificar estilo del gráfico	S	2	10	Baja
HU-10	Descargar gráficos en formato .png	S	2	10	Media
HU-11	Descargar tablas en formato .xlsx	S	2	10	Media
HU-12	Actualizar vistas materializadas	S	2	10	Baja
HU-13	Generar plantillas en formato .xlsx	S	2	10	Baja

Realizado por: Barragán, H., 2023

Los ítems se priorizaron de acuerdo con las necesidades de la Comisión, cuya solicitud principal era poder visualizar los indicadores clave descritos con anterioridad.

Se llenaron las fichas de historias de usuario con las estimaciones realizadas. La **Tabla 3-10** muestra la historia de usuario HU-02, correspondiente al requerimiento de autenticación de usuario.

Tabla 3-10: Historia de usuario HU-02

ID	HU-02	Usuario	Miembro de la Comisión de Titulación
Nombre	Autenticar usuario		
Descripción	Iniciar sesión en la aplicación utilizando su correo y contraseña almacenados en la base de datos del sistema de gestión de trabajos de titulación.		
Validación El miembro del comité puede iniciar sesión en la aplicación con las credenciales que se encuentran guardadas en la base de datos del sistema de gestión de trabajos de titulación.	Prioridad	Baja	
	Puntos	2	
	Estimación (Horas)	10	

Realizado por: Barragán, H., 2023

Las historias de usuario para todos los requerimientos se encuentran documentadas al final de este documento en el **ANEXO B**.

3.2.2.4. *Sprint Backlog*

Las actividades planificadas se ejecutaron 8 *sprints* de dos semanas, a excepción del último, cuyos puntos estimados totales fueron 5, correspondientes a una semana (media iteración). La **Tabla 3-11** muestra el *sprint backlog* para el desarrollo de las actividades descritas en las historias técnicas y de usuario.

Tabla 3-11: *Sprint backlog*

N°	Código	Descripción	Fecha Ini.	Fecha Fin	Pts. estim.	Pts. totales
1	HT-01	Diseñar la arquitectura de la aplicación	24/04/23	25/04/23	2	10
	HT-02	Diseñar la base de datos	26/04/23	26/04/23	1	
	HT-03	Diseñar las vistas materializadas	27/04/23	28/04/23	2	
	HT-04	Cargar los datos	01/05/23	05/05/23	5	
2	HU-05	Cargar datos externos	08/05/23	19/05/23	10	10
3	HU-01	Desarrollar analíticas de datos	22/05/23	02/06/23	10	10
4	HU-01	Desarrollar analíticas de datos	05/06/23	16/06/23	10	10
5	HT-05	Establecer el estándar de codificación	19/06/23	19/06/23	1	10

N°	Código	Descripción	Fecha Ini.	Fecha Fin	Pts. estim.	Pts. totales
	HU-02	Autenticar usuario	20/06/23	21/06/23	2	
	HU-04	Visualizar indicadores clave	22/06/23	28/06/23	5	
	HU-10	Descargar gráficos en formato .png	29/06/23	30/06/23	2	
6	HU-06	Crear indicadores personalizados	03/07/23	07/07/23	5	10
	HU-08	Modificar título del gráfico	10/07/23	10/07/23	1	
	HU-09	Modificar estilo del gráfico	11/07/23	12/07/23	2	
	HU-11	Descargar tablas en formato .xlsx	13/07/23	14/07/23	2	
7	HU-12	Actualizar vistas materializadas	17/07/23	18/07/23	2	10
	HU-13	Generar plantillas en formato .xlsx	19/07/23	20/07/23	2	
	HU-07	Modificar tipo de gráfico	21/07/23	21/07/23	1	
	HU-03	Visualizar analíticas de datos	24/07/23	28/07/23	5	
8	HT-06	Desplegar la aplicación web.	31/07/23	04/08/23	5	5

Realizado por: Barragán, H., 2023

3.2.3. Desarrollo de modelos de analítica utilizando CRISP-DM

El desarrollo de analíticas de datos para la aplicación web “IMAK” se realizó empleado el modelo estándar CRISP-DM. La información presentada a continuación representa los entregables más significativos de cada etapa. El proceso completo se encuentra documentado en el **ANEXO C: MANUAL TÉCNICO – CRISP-DM**.

3.2.3.1. Comprensión del negocio

La Comisión de Titulación (CT) de la carrera de Software, cuyos miembros se presentan en la **Tabla 3-12**, ha almacenado datos referentes al proceso de titulación de los estudiantes. Sin embargo, al no contar con la información suficiente para facilitar la toma de decisiones que conduzcan a solucionar problemas dentro de este contexto, buscaban extraer valor de los datos recolectados de manera automatizada.

Tabla 3-12: Estructura de la Comisión de Titulación

Miembro	Función
Patricio Moreno	Presidente CT
Danilo Pástor	Miembro CT
Ivonne Rodríguez	Miembro CT

Realizado por: Barragán, H., 2023

De acuerdo con los datos que se almacenan y las funciones que sus miembros desempeñan en el proceso de titulación, la Comisión manifestó una serie de necesidades que se convirtieron en objetivos de minería de datos:

Necesidades

- Disponer de información útil que les permita tomar mejores decisiones respecto a los trabajos de titulación.
- Darles un valor agregado a los indicadores claves que manejan, al incluir un análisis más sofisticado de los datos.
- Comunicar de manera oportuna y precisa la información que se modele, de forma tal que individuos externos a la Comisión puedan comprender su valor.
- Migrar los datos almacenados de forma provisional en SQL Server a la base de datos definitiva en PostgreSQL del sistema de gestión de los trabajos de titulación de la carrera.

Objetivos

- **DM1:** Utilizar los títulos de los trabajos de titulación para extraer las palabras clave en un rango de periodos académicos (procesamiento del lenguaje natural para etiquetado gramatical).
- **DM2:** Identificar los temas y términos recurrentes de los trabajos de titulación (procesamiento del lenguaje natural para modelado de temas).
- **DM3:** Visualizar los proyectos no finalizados y el tiempo en meses desde su aprobación (estadística descriptiva).

La ejecución de las fases del proceso de minería de datos se planificó en concordancia con lo descrito en el *sprint backlog*. La **Tabla 3-13** muestra la vista general del plan del proyecto.

Tabla 3-13: Vista general del plan del proyecto

Fases CRISP-DM	Tiempo	Recursos	Riesgos
Comprensión del negocio	2 días	Analista, Miembros de la Comisión	Cambio en los requerimientos.
Comprensión de los datos	3 días	Analista	Cambio en los requerimientos, desconocimiento de las herramientas.
Preparación de los datos	5 días	Analista	Desconocimiento de las herramientas.
Modelado	5 días	Analista	Desconocimiento de las herramientas.
Evaluación	2 día	Analista, Miembros de la Comisión	Desconocimiento de las herramientas.
Despliegue	3 días	Analista/Desarrollador	Desconocimiento de las herramientas.

Realizado por: Barragán, H., 2023

3.2.3.2. *Comprensión de los datos*

Se realizó un análisis de la base de datos provisional “DB_SeguimientoUIC” (SQL Server), con el objetivo de determinar los campos de interés para el análisis y los posibles inconvenientes que deban ser tratados. Dicha base de datos ha sido construida durante varios periodos al recopilar información referente al proceso de titulación, específicamente el seguimiento histórico de los proyectos y estudiantes. Cuenta con 12 tablas, de las cuales 5 son relevantes para la construcción de las analíticas y se presentan en la **Tabla 3-14**.

Tabla 3-14: Descripción de las tablas de interés para el análisis

Tabla	Descripción	Número de columnas
Proyectos aprobados (dsa.proy_aprobado)	Almacena los datos de los proyectos aprobados por la Comisión de Titulación	7
Docente por proyecto (dsa.proy_docente)	Almacena la información del docente por proyecto según su función.	8
Estudiante por proyecto (dsa.proy_estud)	Almacena la información del estudiante en relación con su proyecto	12
Catálogo de procesos (dsa.catProcesos)	Mantiene un catálogo de los procesos por los que atraviesa un proyecto durante su ejecución	5

Tabla	Descripción	Número de columnas
Seguimiento de proyectos (dsa.segProy)	Registra el seguimiento de los proyectos a través de los procesos establecidos por la Comisión.	7

Realizado por: Barragán, H., 2023

Se realizó un perfilado de las columnas en cada tabla para describir, explorar y verificar la calidad de los datos. El módulo *pandas-profiling* de Python fue utilizado para generar reportes de los datos contenidos en cada columna. La **Ilustración 3-6** muestra el reporte generado para la tabla *dsa.proy_aprobado* y la columna *COD_PROYECTO_UIC*.



Ilustración 3-6: Reporte generado con *pandas-profiling*

Realizado por: Barragán, H., 2023

Tras el análisis de los resultados obtenidos en los reportes, se estableció las características de las columnas de interés para el desarrollo de las analíticas, que se describen en la **Tabla 3-15**.

Tabla 3-15: Descripción de las columnas de interés para el análisis

Tabla	Columna	Descripción	Inconvenientes
dsa.proy_aprobado	COD_PERIODO	Código del periodo académico	-
	PROYECTO	Título del proyecto	Caracteres especiales (~,) no deseados, no hay normalización entre mayúsculas y minúsculas
	COD_PROYECTO_UIC	Código asignado al proyecto en la base de datos de la CT	-

Tabla	Columna	Descripción	Inconvenientes
	VIGENCIA	Vigencia del proyecto	-
dsa.proy_docente	COD_PROYECTO_UIC	Código asignado al proyecto en la base de datos de la CT	-
	DOCENTE	Nombre del docente	Nombres mal escritos y repetidos
	FUNCION	Función que desempeña el docente en el proyecto	-
	VIGENCIA	Vigencia del docente en el proyecto	-
dsa.proy_estud	COD_PROYECTO_UIC	Código asignado al proyecto en la base de datos de la CT	-
	APELLIDOS	Apellidos del estudiante	Caracteres especiales (~, ^) no deseados, no hay normalización entre mayúsculas y minúsculas
	NOMBRE	Nombre del estudiante	Caracteres especiales (~, ^) no deseados, no hay normalización entre mayúsculas y minúsculas
dsa.catProcesos	ID_PROCESO	ID del proceso	-
	ACTIVIDAD	Nombre del proceso/actividad	-
dsa.segProy	ID_PROCESO	ID del proceso	-
	COD_PROYECTO_UIC	Código asignado al proyecto en la base de datos de la CT	-
	FECHA_UIC	Fecha de registro del seguimiento en la base de datos de la CT	-

Realizado por: Barragán, H., 2023

Para el desarrollo de la analítica DM2, se utilizaron 13 trabajos de titulación descargados del repositorio DSPACE ESPOCH, que corresponden a proyectos elaborados entre los años 2019 y 2022. Además, se adicionaron 10 proyectos brindados por la CT y que fueron aprobados en el periodo septiembre 2022 - marzo 2023. De esta manera, se contó con un total de 23 proyectos de titulación para el análisis.

3.2.3.3. Preparación de los datos

En esta etapa, se migró la base de datos provisional “DB_SeguimientoUIC” a la base de datos definitiva en PostgreSQL, sobre la que se trabajaría el modelado de las analíticas. Para ello, se ejecutaron procesos ETL para la selección, limpieza, construcción, integración, formateo y carga

de los datos. La herramienta Knime Analytics Platform en su versión 4.7.3 fue utilizada en este proceso. La **Ilustración 3-7** muestra el flujo diseñado en dicho software para la tabla “estudiante_carrera”.

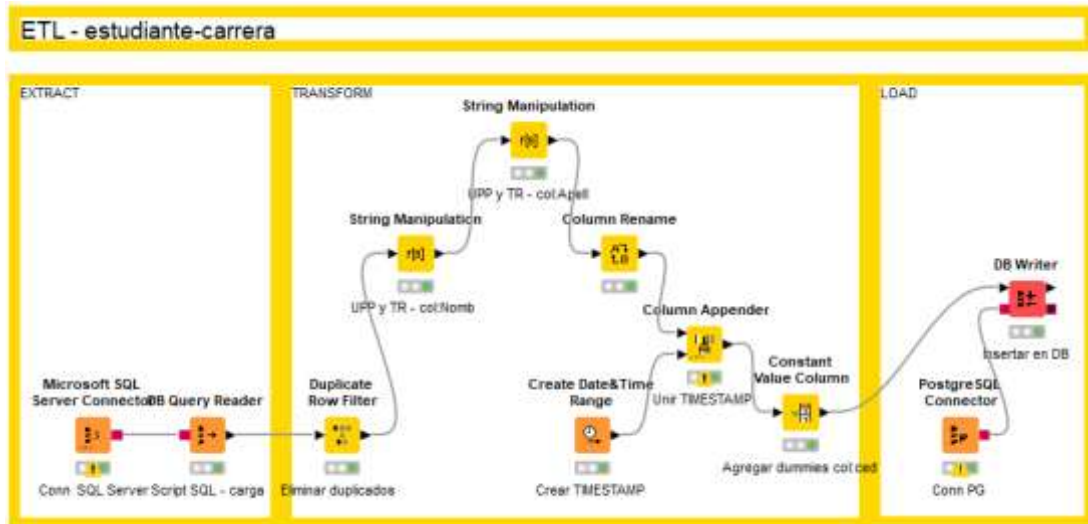


Ilustración 3-7: Proceso ETL para la carga de datos en la tabla "estudiante_carrera"

Realizado por: Barragán, H., 2023

El diseño de los procesos ETL se realizó considerando los inconvenientes detectados en la etapa anterior. De esta manera, se consiguió que los datos cargados en la base de datos PostgreSQL estén preparados para el análisis y su utilización en los indicadores presentados por la aplicación web.

Para la analítica DM2, se desarrollaron varias funciones en Python para la carga y limpieza de los trabajos de titulación. Se utilizó la librería Gensim (Řehuřek y Sojka, 2010, pp.45-50), que proporciona las herramientas necesarias para el modelado de temas. Como parte del proceso de limpieza, se cargaron los archivos y se removieron caracteres especiales; posteriormente, se extrajeron los lemas de las palabras; se removieron los nombres de personas para evitar que aparezcan como términos recurrentes; se dividieron las oraciones en palabras y se eliminaron las palabras vacías (artículos, preposiciones, conectores, etc.). La **Ilustración 3-8** muestra el llamado a las funciones de limpieza.

```
#Clean text
lemmatized = lemmatization(pdfes)
no_names = remove_names(lemmatized)
words_sent = list(sent_to_words(no_names))
data_words = remove_stopwords(words_sent)
✓ 1m 29.8s
```

Ilustración 3-8: Proceso de limpieza de los trabajos de titulación

Realizado por: Barragán, H., 2023

Con el módulo “corpora” de Gensim, se construyó un diccionario con las palabras obtenidas y un ID entero asignado a cada una. Finalmente, se transformó el diccionario a un formato de “bolsa de palabras” (*bag-of-words*), donde aparecen las distintas palabras y su frecuencia (**Ilustración 3-9**).

```
id2word = corpora.Dictionary(data_words)
texts = data_words
corpus = [id2word.doc2bow(text) for text in texts]
✓ 0.0s
```

Ilustración 3-9: Obtención del diccionario de palabras y corpus

Realizado por: Barragán, H., 2023

3.2.3.4. Modelado

En esta etapa, se definió el diseño de evaluación y se construyeron los modelos de analítica de datos necesarios para cumplir con los objetivos de minería planteados en el proyecto. De las tres analíticas planteadas, solamente DM1 y DM2 requieren ser modeladas y evaluadas, ya que DM3 emplea únicamente estadística descriptiva.

Diseño de evaluación

En la evaluación del etiquetado gramatical (DM1) se calculó la exactitud (*Accuracy*). Dalianis (2018, pp.45-53) la define como:

$$Exactitud = \frac{tp + tn}{tp + tn + fp + fn}$$

Donde:

tp: verdaderos positivos;

tn: verdaderos negativos;

fp: falsos positivos;

fn: falsos negativos.

La fórmula puede ser interpretada como:

$$\text{Exactitud} = \frac{\text{Número de palabras clasificadas correctamente}}{\text{Número total de palabras clasificadas}}$$

Ya que los títulos de los trabajos de titulación no se encuentran etiquetados, se evaluó el corpus AnCora, que consta de 500 000 palabras en español anotadas a diferentes niveles lingüísticos (Taulé et al., 2008).

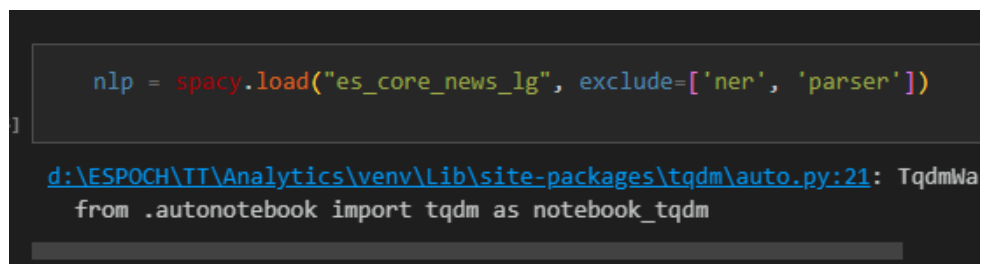
Para la evaluación del modelado de temas, se empleó la medida de coherencia C_{UMass} , que cuantifica el grado de similitud semántica de las palabras más importantes en un tema o tópico (Zoya et al., 2021, pp.127531-127547). Según Mimno et al. (2011, pp.262-272), se calcula:

$$C_{UMass} = \sum_{m=2}^M \sum_{l=1}^{m-1} \log \frac{D(v_m^{(t)}, v_l^{(t)}) + 1}{D(v_l^{(t)})}$$

Gensim cuenta con un módulo para determinar esta medida de forma automática.

Construcción de los modelos

En el etiquetado gramatical de los títulos de los trabajos de titulación, se empleó la librería spaCy, que cuenta con modelos pre entrenados en más de 60 idiomas, entre ellos español; a diferencia de otras librerías para procesamiento del lenguaje natural como NLTK, CoreNLP o ZPar, spaCy está enfocado en código de producción y puede llegar a ser más eficiente (Altinok, 2021). En la **Ilustración 3-10** se muestra la importación del modelo en Python para su uso.



```
nlp = spacy.load("es_core_news_lg", exclude=['ner', 'parser'])  
]  
d:\ESPOCH\TT\Analytics\venv\Lib\site-packages\tqdm\auto.py:21: TqdmWa  
from .autonotebook import tqdm as notebook_tqdm
```

Ilustración 3-10: Carga del modelo pre-entrenado para etiquetado gramatical

Se diseñó una función para extraer las palabras claves de cada título, filtrando únicamente los sustantivos, nombres propios y adjetivos. La función se presenta en la **Ilustración 3-11**.

```
def extract_keywords(title: str):
    doc = nlp(title)
    keywords = []
    for word in doc:
        if not word.is_stop and (word.pos_ in ["NOUN", "PROPN", "ADJ"]):
            keywords.append(word)
    return keywords
```

Ilustración 3-11: Función para extraer las palabras clave

Realizado por: Barragán, H., 2023

El modelo *Latent Dirichlet Allocation* (LDA), disponible en la librería Gensim, fue utilizado en el desarrollo de la analítica DM2. Se seleccionó LDA debido a que es la técnica de modelado de temas más utilizada (Kim et al., 2022, pp.263-278). La **Ilustración 3-12** muestra los parámetros escogidos para el entrenamiento del modelo.

```
lda_model = gensim.models.LdaModel(corpus=corpus,
    id2word=id2word,
    num_topics=10,
    random_state=100,
    chunksize=100,
    passes=10,
    per_word_topics=True
)
```

Ilustración 3-12: Entrenamiento del modelo LDA

Realizado por: Barragán, H., 2023

3.2.3.5. Evaluación

El resultado obtenido en el cálculo de la exactitud de la analítica DM1 se muestra en la **Ilustración 3-13**.

```
accuracy = results['pos_acc']
print(f'Exactitud del modelo "es_core_news_lg": {accuracy:.2f}')
✓ 0.0s
Exactitud del modelo "es_core_news_lg": 0.97
```

Ilustración 3-13: Precisión obtenida del modelo de etiquetado gramatical

Realizado por: Barragán, H., 2023

La exactitud obtenida es de 0.97, lo que indica que 97% del corpus de prueba pudo ser etiquetado correctamente. Según Barkved (2022), la interpretación de la exactitud de un modelo es subjetiva; aun así, un puntaje superior al 70% indica un buen desempeño.

Al calcular la coherencia del modelo de la analítica DM2, obtenemos los resultados presentados en la **Ilustración 3-14**.

```
coherence_model_lda = CoherenceModel(model=lda_model,
                                     texts=data_words,
                                     dictionary=id2word,
                                     coherence='u_mass')
coherence_lda = coherence_model_lda.get_coherence()
print(f'Puntaje de coherencia: {coherence_lda:.2f}')
✓ 0.0s
Puntaje de coherencia: -4.21
```

Ilustración 3-14: Coherencia obtenida del modelo LDA

Realizado por: Barragán, H., 2023

Se realizó un ajuste de los hiper parámetros α (densidad de tópicos en un documento), β (densidad de palabras en un tópico) y K (número de tópicos) del modelo para obtener distintos puntajes de coherencia y determinar la mejor configuración. Los valores de α y β se iteraron de 0.1 a 1 con un incremento de 0.3, mientras que K se incrementó de 2 a 30. Adicionalmente, se evaluaron los resultados para el 75% del corpus original, con el objetivo de determinar si existía una mejora en contraste al corpus completo. Los resultados del ajuste de hiper parámetros se muestra en la **Ilustración 3-15**, presentada a continuación.

La segunda analítica (DM2) se implementó utilizando la librería pyLDAvis para Python, que permite visualizar los tópicos y términos más importantes generados por el modelo de forma interactiva. La **Ilustración 3-17** presenta el reporte obtenido.

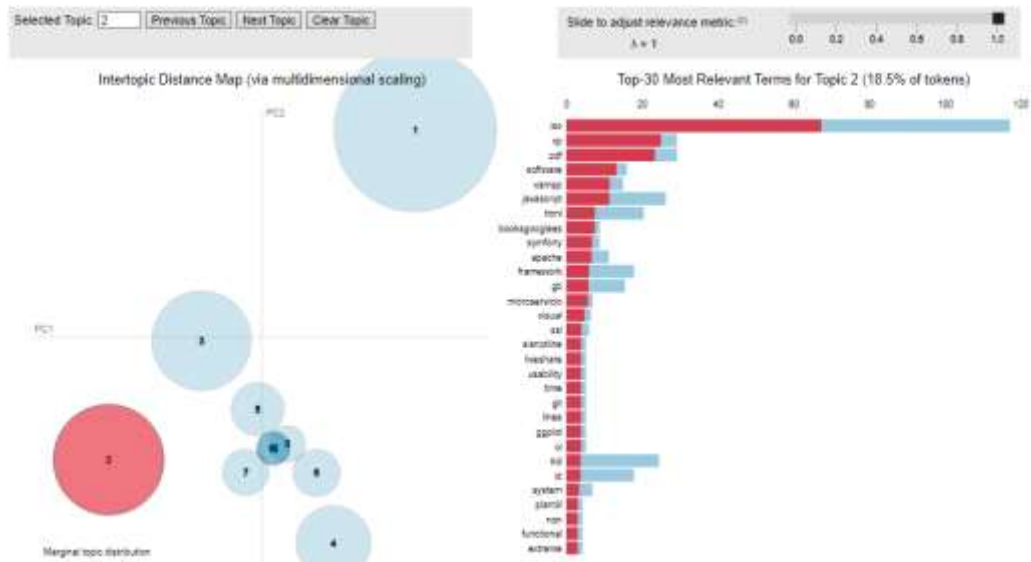


Ilustración 3-17: Visualización de los tópicos y términos más relevantes

Realizado por: Barragán, H., 2023

La tercera analítica se representó utilizando un diagrama de violín (**Ilustración 3-18**), mostrando la distribución de la duración en meses de los trabajos de titulación. Además, se agregó una tabla con la información detallada de cada proyecto y la opción de filtrar las columnas según los requerimientos del usuario.

DISTRIBUCIÓN DE LA DURACIÓN EN MESES DE LOS PROYECTOS NO FINALIZADOS

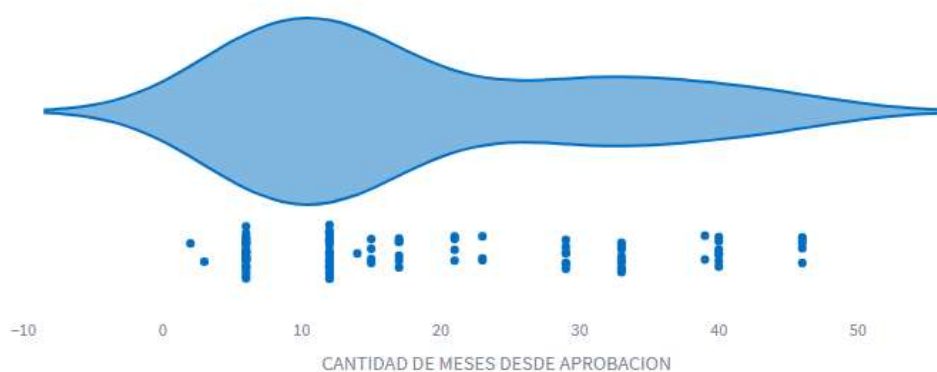


Ilustración 3-18: Distribución de la duración en meses de los proyectos no finalizados

Realizado por: Barragán, H., 2023

3.2.4. Desarrollo de la aplicación

La aplicación web para el apoyo a la toma de decisiones “*IMAK*” se desarrolló mediante la ejecución de tareas técnicas que incluyen la definición de la arquitectura, el establecimiento de estilos de codificación, el diseño de la base de datos y de las interfaces de usuario.

3.2.4.1. Arquitectura de la aplicación

Para documentar la arquitectura de la aplicación, se ha optado por emplear el modelo C4, que define cuatro vistas o niveles, cada una centrado en una perspectiva diferente del sistema. Los diagramas que propone este modelo son de contexto, contenedor, componente y código, y su sintaxis es simple y flexible (Vázquez-Ingelmo et al., 2020, pp.919-924). Se emplearon únicamente las dos primeras vistas propuestas en el modelo.

El primer diagrama presentado en la **Ilustración 3-19**, y que corresponde al contexto de la aplicación, evidencia la interacción de “*IMAK*” con el sistema de gestión de trabajos de titulación de la carrera en una arquitectura cliente - servidor.



Ilustración 3-19: Diagrama de contexto de la aplicación

Realizado por: Barragán, H., 2023

En la segunda vista, representada a través de un diagrama de contenedor (Ilustración 3-20), se observa que la aplicación web “IMAK” obtiene los datos de las vistas almacenadas en el esquema “kpi” de la base de datos del sistema de gestión de trabajos de titulación de la carrera a través de una arquitectura cliente-servidor.

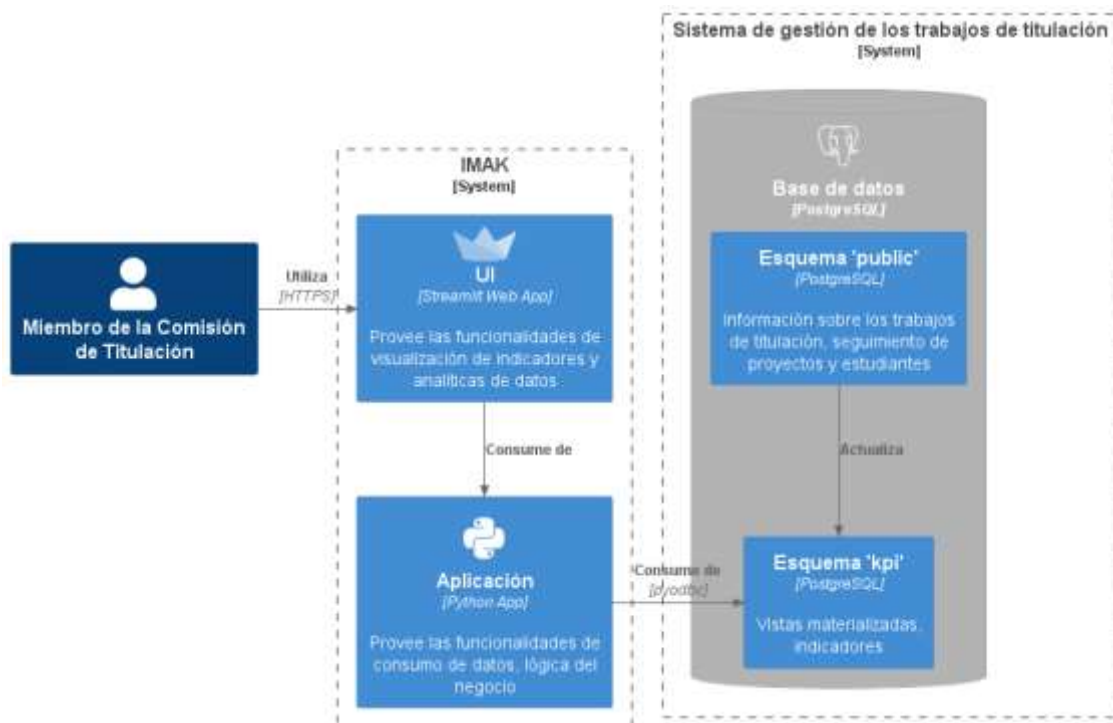


Ilustración 3-20: Diagrama de contenedor de la aplicación

Realizado por: Barragán, H., 2023

3.2.4.2. Estilo de codificación

Se utilizó la guía de estilo PEP 8, que provee la orientación necesaria para mejorar la lectura del código en Python y su consistencia. La **Tabla 3-16** resume algunas de las pautas establecidas para las características más importantes del código. Las convenciones completas pueden encontrarse en la guía oficial disponible en el sitio web⁶ de PEPS (*Python Enhancement Proposals*).

Tabla 3-16: Convenciones de programación

Característica	Convención
Disposición del código	4 espacios por nivel de sangrado (<i>indentation</i>); líneas de máximo 79 caracteres; funciones de alto nivel y clases rodeadas por dos líneas en blanco.
Comentarios	Priorizar la actualización de comentarios cuando el código cambie; usar oraciones completas; primera letra en mayúscula; comentarios en inglés.
Denominaciones	No utilizar los caracteres ‘l’, ‘O’, ‘I’ como nombres de variables de una única letra; los nombres de clase deben utilizar la convención <i>CapWords (PascalCase)</i> ; las funciones y variables se deben nombrar con minúsculas, separando las palabras con guion bajo (<i>snake_case</i>); las constantes se escriben en mayúsculas, utilizando guion bajo entre palabras.

Fuente: van Rossum et al., 2001

Realizado por: Barragán, H., 2023

3.2.4.3. Diseño de la base de datos

La aplicación web “*IMAK*” utiliza una base de datos propia en el esquema “*kpi*” para obtener la información referente a los indicadores clave requeridos por la Comisión de Titulación, que son almacenados en vistas materializadas y posteriormente consultadas para presentarlos en forma de gráficos. La **Ilustración 3-21** presenta el diseño de la base de datos.

⁶ <https://peps.python.org/pep-0008/>

gestion_columnas		
id_columna	INT2	<pk>
columna	VARCHAR(150)	
tabla	VARCHAR(150)	
alias	VARCHAR(150)	

indicador		
id_indicador	INT2	<pk>
id_dataset	INT2	<fk>
nombre_indicador	VARCHAR(110)	
descripcion_indicador	VARCHAR(250)	
fecha_creacion	DATE	
vigencia	BOOL	
datos_externos	BOOL	

dataset		
id_dataset	INT2	<pk>
descripcion_dataset	VARCHAR(250)	
vista_materializada	VARCHAR(250)	

Ilustración 3-21: Diseño de la base de datos

Realizado por: Barragán, H., 2023

3.2.4.4. Diccionario de datos

El diccionario de datos documenta y define la información principal de los campos de cada tabla de una base de datos, incluyendo el nombre de la columna, descripción, tipo, si acepta nulos y los valores que pueden ser ingresados. La **Tabla 3-17** muestra el diccionario de datos de la tabla “indicador”.

Tabla 3-17: Diccionario de datos para la tabla “indicador”

Nombre de la tabla: indicador				
Descripción: Indicador manejado por la Comisión de Titulación				
Columna	Descripción	Tipo	Nulos	Valor permitido
id_indicador (PK)	Identificador único del indicador	smallint	No	*0 a 32767*
id_dataset (FK)	Identificador del dataset que almacena los datos del indicador	smallint	No	*0 a 32767*
nombre_indicador	Nombre corto asignado al indicador para identificarlo	varchar(110)	No	{ [A-Z a-z] }
descripcion_indicador	Nombre largo (descripción) asignado al indicador	varchar(250)	Si	{ [A-Z a-z] }
fecha_creacion	Fecha de creación del indicador en la BD de la CT	date	No	* formato: aaaa – mm – dd *
vigencia	Vigencia del indicador en la CT	boolean	No	[true false]
datos_externos	Indica si el indicador requiere la carga de datos externos	boolean	No	[true false]

Realizado por: Barragán, H., 2023

El **ANEXO B** documenta el diccionario de datos para el resto de las tablas en el esquema “kpi”.

3.2.4.5. Diseño de las interfaces

Se diseñaron *wireframes* de los módulos de la aplicación web utilizando el software Balsamiq Wireframes. Esto ayudó a visualizar su estructura y los elementos que la constituyen, además de clarificar detalles relacionados al diseño y requisitos de la aplicación previo a su desarrollo. La **Ilustración 3-22** muestra el *wireframe* construido para la pantalla de inicio de sesión.

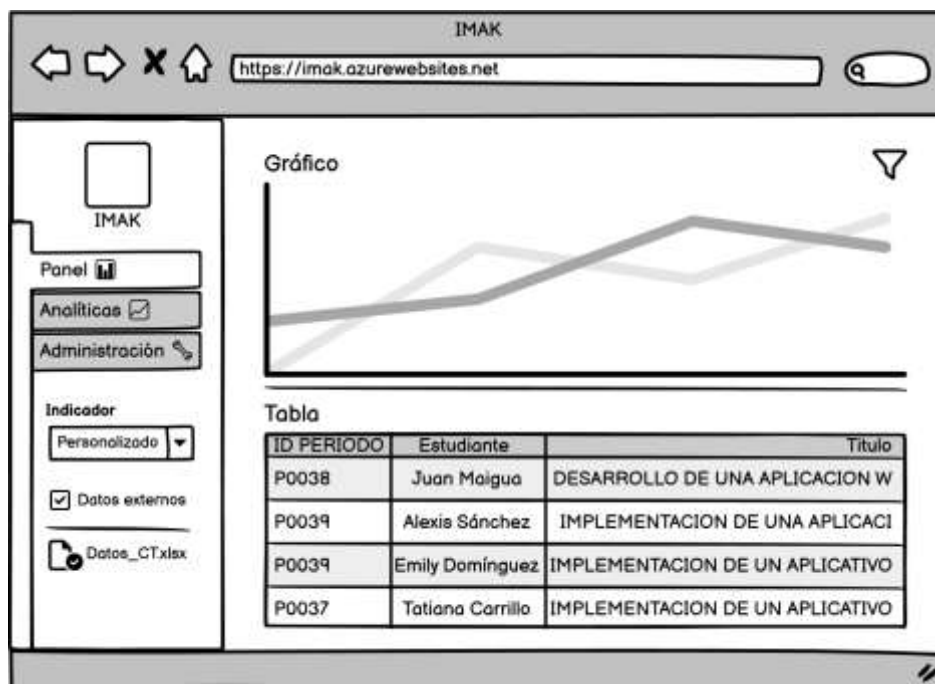


Ilustración 3-22: Wireframe de la pantalla “Panel”

Realizado por: Barragán, H., 2023

El diseño de las pantallas correspondientes a los demás módulos se ha recopilado en el **ANEXO B**.

3.2.5. Cierre

Se describen las actividades realizadas que marcaron la finalización de la gestión y el desarrollo del proyecto.

3.2.5.1. Burndown chart

Para visualizar la diferencia entre los puntos de historia estimados y los puntos reales completados en cada *sprint*, se empleó el *burndown chart*, que se muestra en la **Ilustración 3-23**.

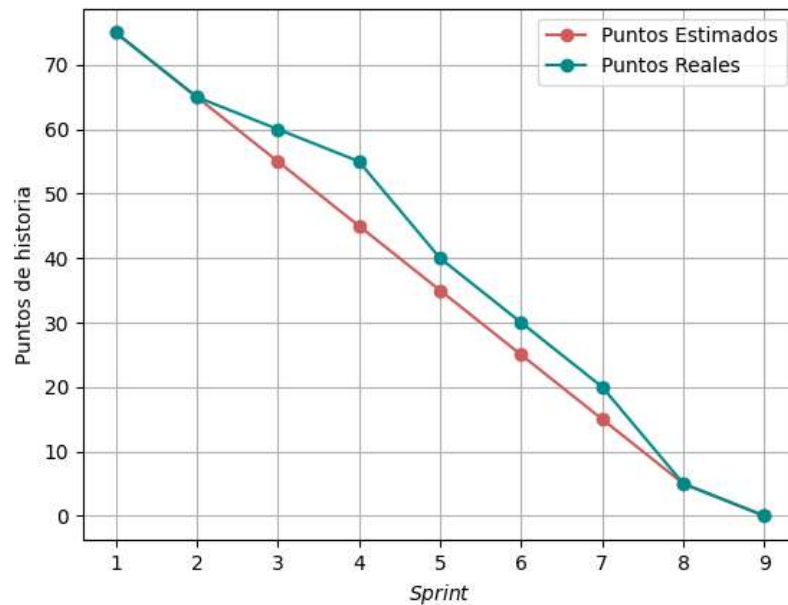


Ilustración 3-23: *Burndown chart*

Realizado por: Barragán, H., 2023

En general, los puntos reales fueron mayores a los estimados, especialmente en el *sprint* cuatro durante el desarrollo de las analíticas; esto se debe a que la Comisión de Titulación incorporó nuevos datos del último periodo académico, lo que requirió un tiempo adicional en los procesos ETL. Sin embargo, al finalizar el *sprint* ocho se pudieron cumplir todas las actividades planificadas.

3.2.5.2. Implementación

Luego de finalizar el desarrollo, se desplegó la aplicación en un servidor de la carrera de Software. La **Ilustración 3-24** muestra el diagrama de despliegue.

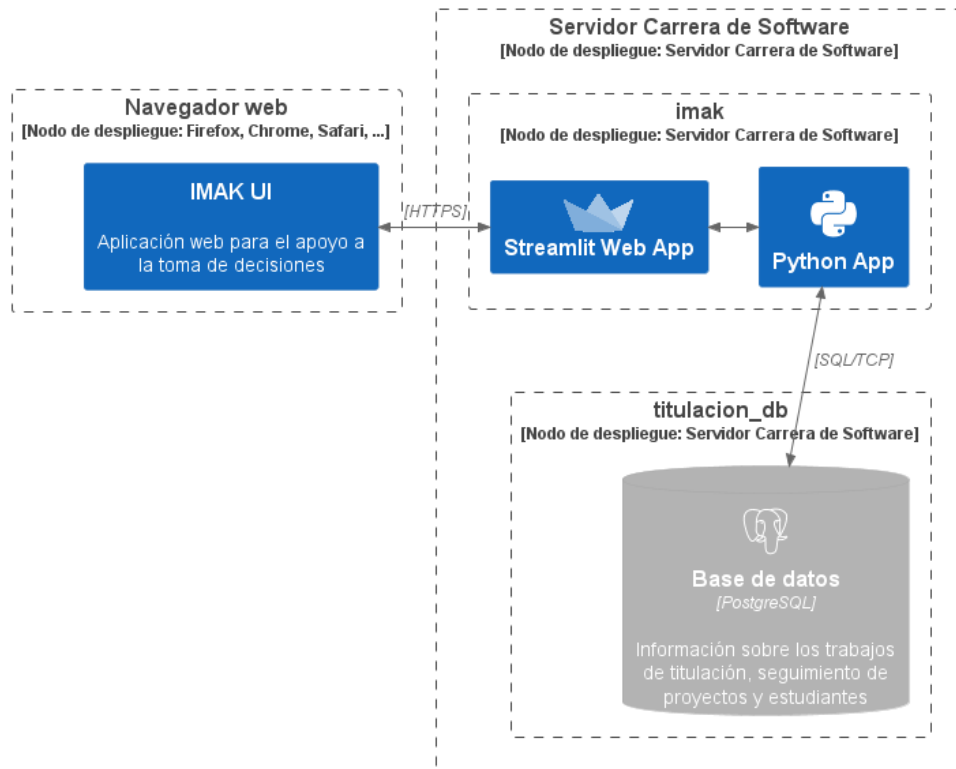


Ilustración 3-24: Diagrama de despliegue

Realizado por: Barragán, H., 2023

Se les comunicó a los miembros de la Comisión de Titulación el sitio web en el que podrían acceder a la aplicación. Además, se les brindó una breve explicación del uso básico y las funcionalidades disponibles.

CAPÍTULO IV

4. ANÁLISIS E INTERPRETACIÓN DE RESULTADOS

El cuarto capítulo detalla los resultados obtenidos en la evaluación de las dimensiones seleccionadas del modelo de éxito de los sistemas de información de DeLone y McLean. Además, se presenta la validación estadística del modelo de ecuaciones estructurales y comprobación de hipótesis por el método de mínimos cuadrados parciales (PLS-SEM).

4.1. Evaluación de las variables según el modelo de DeLone y McLean

Para la evaluación de las variables (también llamadas dimensiones o constructos) se utilizó un cuestionario (documentado en el **ANEXO A**) adaptado de tres instrumentos: AIMQ, una metodología para la evaluación de la calidad de la información propuesta por Lee et al. (2002, pp.133-146); las medidas multidimensionales para cuantificar el uso de un sistema en contextos organizacionales desarrolladas por Doll y Torkzadeh (1998, pp.171-185); y las métricas de calidad del sistema y satisfacción de usuario empleadas en la validación empírica ejecutada por Iivari (2005, pp.8-27). Dicho cuestionario se dividió en cuatro secciones, de acuerdo con las variables e indicadores escogidos (Ver **Tabla 3-3**). La principal adaptación respecto a los instrumentos originales consistió en utilizar una escala de Likert de 1 a 5 para todos los ítems (1: “Completamente en desacuerdo”; 2: “En desacuerdo”; 3: “Indiferente”; 4: “De acuerdo”; 5: “Completamente de acuerdo”), facilitando la tabulación y análisis de los resultados.

4.2. Análisis de resultados

El modelo de ecuaciones estructurales por el método de mínimos cuadrados parciales (PLS-SEM) fue utilizado en la comprobación de las hipótesis planteadas, con base a los datos recolectados en la encuesta aplicada mediante cuestionario. Su uso es ideal en esta situación, ya que PLS-SEM trabaja con muestras pequeñas en comparación con el método basado en covarianzas (CB-SEM), además de que permite la incorporación de indicadores reflectivos y formativos (Hair et al., 2011, pp.139-152). Anderson y Gerbing (1988, pp.411-423) proponen un método de dos pasos para la comprobación de hipótesis. La primera etapa consiste en la comprobación de la validez y fiabilidad de los ítems utilizados en el modelo de medidas (relación entre los indicadores y las variables latentes). En la segunda etapa se prueban las relaciones del modelo estructural, es decir, se validan las hipótesis planteadas entre los constructos. El software SmartPLS (Ringle et al., 2022) en su versión 4.0.9.5 fue utilizado para generar los indicadores del análisis.

4.2.1. Prueba del modelo de medidas

Se obtuvo un reporte de sus cargas externas en cada constructo para determinar la fiabilidad de los indicadores. Los resultados generados por el software SmartPLS se muestran en la **Tabla 4-1**.

Tabla 4-1: Pruebas de cargas externas

Indicador	Satisfacción de usuario (STF)	Uso (USO)	Calidad del sistema (SQ)	Calidad de la información (IQ)
STF1	0.870	-	-	-
STF2	0.870	-	-	-
USO1	-	0.850	-	-
USO2	-	0.695	-	-
USO3	-	0.756	-	-
USO4	-	0.757	-	-
USO5	-	0.775	-	-
USO6	-	0.748	-	-
SQ1	-	-	0.781	-
SQ2	-	-	0.781	-
IQ1	-	-	-	0.771
IQ2	-	-	-	0.477
IQ3	-	-	-	0.414
IQ4	-	-	-	0.064
IQ5	-	-	-	0.611
IQ6	-	-	-	0.485

Realizado por: Barragán, H., 2023

Se recomienda que las cargas de los indicadores estén por arriba de 0.708; las cargas entre 0.4 y 0.708 pueden eliminarse si esto conduce a incrementar la fiabilidad en la consistencia interna o la validez convergente, mientras que aquellas inferiores a 0.4 deberán ser siempre removidos (Hair et al., 2021). De acuerdo con los resultados obtenidos, IQ4 no cumple esta condición, por lo que fue eliminado del análisis. La **Tabla 4-2** muestra los valores obtenidos de la fiabilidad compuesta (Rho_c) y validez convergente (AVE) al incluir y remover cada indicador en el modelo.

Tabla 4-2: Fiabilidad y validez convergente al eliminar cada ítem

Indicador	Fiabilidad (Rho_c)	Fiabilidad (Rho_c) sin el indicador	Validez convergente (AVE)	Validez convergente (AVE) sin el indicador
STF1	0.862	-	0.757	-
STF2	0.862	-	0.757	-
USO1	0.894	0.868	0.585	0.569
USO2	0.894	0.894	0.585	0.628
USO3	0.894	0.884	0.585	0.604
USO4	0.894	0.885	0.585	0.607
USO5	0.894	0.881	0.585	0.599

Indicador	Fiabilidad (Rho_c)	Fiabilidad (Rho_c) sin el indicador	Validez convergente (AVE)	Validez convergente (AVE) sin el indicador
USO6	0.894	0.885	0.585	0.608
SQ1	0.758	-	0.610	-
SQ2	0.758	-	0.610	-
IQ1	0.734	0.635	0.375	0.314
IQ2	0.734	0.721	0.375	0.406
IQ3	0.734	0.799	0.375	0.502
IQ5	0.734	0.700	0.375	0.398
IQ6	0.734	0.733	0.375	0.418

Realizado por: Barragán, H., 2023

El único ítem que al ser eliminado mejoró la fiabilidad y la validez convergente de su constructo fue IQ3, por lo tanto, se descartó del modelo. En la **Tabla 4-3** se muestran los resultados finales del coeficiente Rho_c para cada variable latente.

Tabla 4-3: Resultados de la prueba de fiabilidad

Variable	Rho_c	Estado
Calidad de la información	0.799	Fiable
Calidad del sistema	0.758	Fiable
Satisfacción de usuario	0.862	Fiable
Uso	0.894	Fiable

Realizado por: Barragán, H., 2023

El coeficiente Rho_c mide la fiabilidad de la consistencia interna de un constructo considerando las cargas internas de los indicadores de una variable; valores entre 0.6 y 0.7 son aceptables en un estudio exploratorio, mientras que entre 0.7 y 0.9 son considerados satisfactorios (Hair et al., 2021). Todos los resultados son superiores a 0.7, concluyendo así que los constructos son fiables.

En el siguiente paso, se ejecutaron validaciones convergentes. Esta se determinó a través de la varianza promedio extraída (*average variance extracted* o AVE), que se calcula al sumar los cuadrados de las cargas y dividirlo para el número de indicadores; el AVE mínimo aceptable es de 0.5, cuyo significado es que el constructo explica al menos el 50% de la varianza de sus indicadores (Hair et al., 2021). Los resultados finales de la AVE para cada variable latente se muestran en la **Tabla 4-4**.

Tabla 4-4: Prueba de validez convergente (AVE)

Variables	AVE	Estado
Satisfacción de usuario	0.502	Válido
Uso	0.610	Válido
Calidad del sistema	0.757	Válido

Variables	AVE	Estado
Calidad de la información	0.585	Válido

Realizado por: Barragán, H., 2023

Como todas las variables del modelo cumplen con la condición de ser superior o iguales a 0.5, podemos asegurar su validez convergente.

Para evaluar la validez discriminante de los constructos, se empleó el promedio de las correlaciones de los indicadores entre variables, conocido como HTMT (*heterotrait-monotrait ratio*). En la **Tabla 4-5** se muestran los resultados obtenidos en SmartPLS para esta medida.

Tabla 4-5: Prueba de validez discriminante (HTMT)

Relaciones entre variables	Heterotrait-monotrait ratio (HTMT)	Estado
SQ - IQ	0.572	Válido
STF - IQ	0.655	Válido
STF - SQ	0.498	Válido
USO - IQ	0.630	Válido
USO - SQ	0.790	Válido
USO - STF	0.419	Válido

Realizado por: Barragán, H., 2023

Según Hair et al. (2021), el valor de HTMT no debe ser mayor a 0.9, ya que esto indicaría que los ítems de diferentes variables están midiendo el mismo constructo. Ningún valor superó el 0.8, por lo tanto, todas las variables en el modelo resultaron ser discriminantemente válidas.

4.2.2. Prueba del modelo estructural

Una vez realizadas las pruebas de validez y fiabilidad de los indicadores de cada variable latente, el siguiente paso consistió en realizar las pruebas a la parte interna o estructural del modelo, es decir, las relaciones entre constructos. La **Ilustración 4-1** muestra el diagrama generado en SmartPLS, donde se pueden observar los coeficientes de regresión estandarizados (beta) de cada relación.

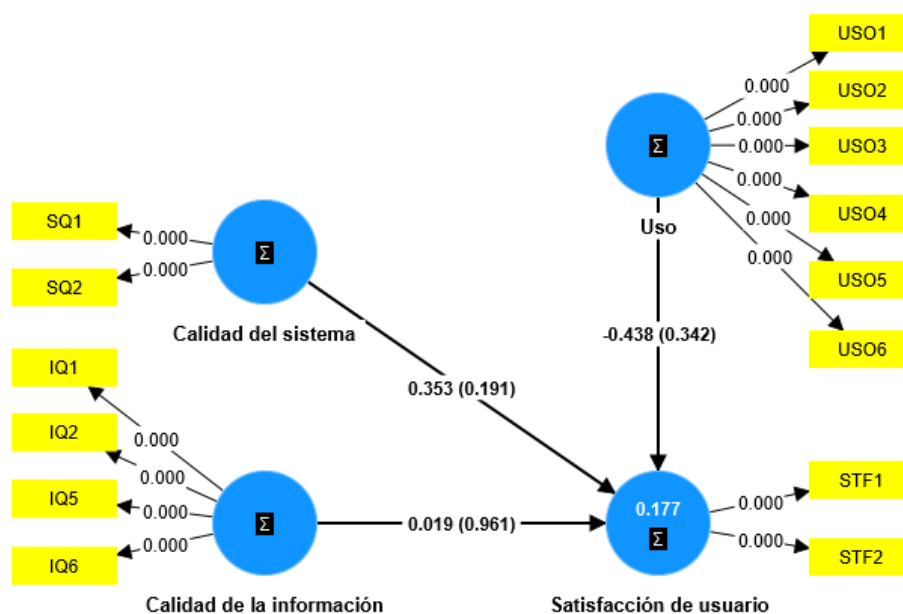


Ilustración 4-1: Coeficientes de regresión obtenidos

Realizado por: Barragán, H., 2023

La **Tabla 4-6** relaciona las hipótesis con los coeficientes de regresión obtenidos para las relaciones de interés.

Tabla 4-6: Coeficientes de regresión obtenidos para cada relación

Hipótesis	Relación	β	Estado
H_1	USO \rightarrow STF	-0.438	Rechazada
H_2	SQ \rightarrow STF	0.353	Aceptada
H_3	IQ \rightarrow STF	0.019	Aceptada

Realizado por: Barragán, H., 2023

Los coeficientes de regresión entre la satisfacción de usuario (STF) y la calidad del sistema (SQ) y de la información (IQ) son mayores a 0, indicando una influencia positiva entre dichas variables, aceptando las hipótesis H_2 y H_3 . Sin embargo, el coeficiente de regresión entre USO y STF es negativo, por lo que no podemos aceptar H_1 . SmartPLS permite ejecutar un algoritmo de *bootstrapping* para calcular el estadístico t y el valor p . Con un nivel de significancia $\alpha = 0.05$, se obtuvieron los resultados presentados en la **Tabla 4-7**.

Tabla 4-7: Resultados obtenidos al ejecutar el algoritmo de *bootstrapping*

Relación	Estadístico t	p -value ($\alpha = 0.05$)	Estado
USO \rightarrow STF	0.954	0.342	No significativa
SQ \rightarrow STF	1.315	0.191	No significativa
IQ \rightarrow STF	0.048	0.961	No significativa

Realizado por: Barragán, H., 2023

Ninguna relación resultó ser significativa para $\alpha = 0.05$. Otra consideración relevante en el modelo es el coeficiente de determinación R^2 , cuyo valor para la satisfacción de usuario es 0.177. Un resultado inferior a 0.25 implica que el poder de las variables independientes para explicar la variable dependiente es débil (Hair et al., 2021).

CAPÍTULO V

5. CONCLUSIONES Y RECOMENDACIONES

5.1. Conclusiones

De acuerdo con los resultados obtenidos en el presente estudio, se llegaron a las siguientes conclusiones:

- Se seleccionó y aplicó procesamiento del lenguaje natural para identificar las palabras clave en los títulos de los trabajos de integración curricular utilizando etiquetado gramatical; aplicando la misma técnica, se determinaron los tópicos y términos más recurrentes empleando LDA para el modelado de temas. Utilizando estadística descriptiva, se visualiza la distribución de la duración en meses de los trabajos no finalizados.
- Con el modelo pre-entrenado de la librería spaCy utilizado en la analítica para la extracción de palabras clave se obtuvo una precisión de 0.97, lo que implica que fue capaz de etiquetar el 97% del corpus de forma correcta. Así mismo, el puntaje de coherencia del modelo LDA de la analítica para identificar los temas y términos recurrentes, con un reajuste de los parámetros $K = 11$, $\alpha = 0.9099$, $\beta = 0.31$ y empleando únicamente el 75% del corpus, alcanzó un valor de -1.60.
- El desarrollo de la aplicación web para el apoyo a la toma de decisiones “*IMAK*” se ejecutó con la metodología Scrum, permitiendo la implementación de 13 historias de usuario y 6 historias técnicas en 8 iteraciones. La utilización del *framework* Streamlit facilitó la construcción del producto gracias a sus características enfocadas a la creación de aplicaciones orientadas a datos.
- La evaluación de la aplicación web “*IMAK*” para el apoyo a la toma de decisiones, se realizó empleando el modelo de éxito de los sistemas de información de DeLone y McLean, considerando las dimensiones, también llamadas variables: satisfacción de usuario, uso, calidad del sistema y calidad de la información. La medición de la relación entre estas variables se realizó mediante el modelo de ecuaciones estructurales con el análisis de mínimos cuadrados parciales (PLS-SEM).

- Se comprobó la fiabilidad y validez de las relaciones entre los indicadores y las variables. Al calcular los coeficientes de regresión β , se determinó que la calidad del sistema y la calidad de la información tuvieron influencia positiva sobre la satisfacción de usuario, mientras que el uso tuvo influencia negativa. Ninguna relación resultó ser significativa con un valor de $\alpha = 0.05$ y el poder de las variables independientes para predecir la variable dependiente al calcular el coeficiente de determinación R^2 es débil, lo que sugiere la necesidad de mayor estudio.

5.2. Recomendaciones

- Se recomienda considerar otras dimensiones del modelo de DeLone y McLean, como la Intención de Uso y el Beneficio Neto, en un análisis más extenso del éxito de la aplicación web desarrollada en el presente trabajo, ya que se podrían derivar interesantes conclusiones respecto a la influencia de factores como la calidad del sistema y de la información tras su uso prolongado.
- Ya que la aplicación web está enfocada al uso de modelos de analítica de datos, es recomendable que la solución se mantenga desplegada en una estructura adecuada que garantice su correcto rendimiento y disponibilidad.
- En un futuro, se recomienda volver a estudiar los datos que almacene la Comisión de Titulación, de tal forma que se pueda desarrollar analíticas de datos más sofisticadas, alcanzando niveles predictivos y prescriptivos de análisis de la información.
- Se sugiere, en próximos trabajos, la incorporación de un módulo de gestión de indicadores, que les permita a los miembros de la Comisión manejar *scripts* SQL para la creación de nuevas vistas materializadas de acuerdo con futuros requerimientos.

GLOSARIO

Decision maker: individuo o entidad encargada de tomar decisiones, especialmente en un contexto empresarial o de toma de decisiones estratégicas. Esta persona o grupo de personas evalúa información y opciones para elegir un curso de acción adecuado (Simon, 1997).

Framework: estructura reutilizable que proporciona una base para el desarrollo de aplicaciones. Ofrece un conjunto de herramientas, bibliotecas y patrones de diseño que facilitan la creación y el mantenimiento de software (Chacon y Straub, 2014).

KDD: Proceso no trivial de identificación de patrones válidos, novedosos, potencialmente útiles y, en última instancia, comprensibles en los datos, a menudo utilizando técnicas de minería para obtener información valiosa y conocimiento útil (Fayyad et al., 1996, p.37).

Minería de datos: se refiere a la práctica de explorar grandes conjuntos de datos para descubrir patrones, relaciones y tendencias ocultas. Utiliza técnicas y algoritmos computacionales para extraer información valiosa y conocimiento a partir de datos, lo que puede ayudar en la toma de decisiones y la predicción (Ha et al., 2011).

Módulo: unidad de código independiente y reutilizable que contiene funciones, variables y estructuras de datos relacionadas. Los módulos se utilizan para organizar y modularizar el código de un programa, lo que facilita su mantenimiento y la reutilización en diferentes partes del software (Lutz, 2011).

Open source: software cuyo código fuente es accesible y modificable por la comunidad, lo que fomenta la colaboración y la transparencia en el desarrollo. Esta filosofía ha llevado a la creación de una amplia variedad de proyectos de código abierto en diversas áreas (von Krogh y Spaeth, 2007, pp.236-253).

BIBLIOGRAFÍA

1. **ALAHMADI, Dimah Hussein; & JAMJOOM, Arwa A.** «Decision support system for handling control decisions and decision-maker related to supply chain». *Journal of Big Data* [en línea], 2022, 9(1), p.114. ISSN 2196-1115. DOI 10.1186/s40537-022-00653-9. Disponible en: <https://journalofbigdata.springeropen.com/articles/10.1186/s40537-022-00653-9>
2. **ALEEM, Abdul; & GORE, Manoj Madhava.** «Educational Data Mining Methods: A Survey». *2020 IEEE 9th International Conference on Communication Systems and Network Technologies (CSNT)* [en línea], 2020, (Gwalior, India), pp.182-188. DOI 10.1109/CSNT48778.2020.9115734. Disponible en: <https://ieeexplore.ieee.org/document/9115734>
3. **ALTINOK, Duygu.** *Mastering spaCy: An end-to-end practical guide to implementing NLP applications using the Python ecosystem*. Packt Publishing Ltd, 2021.
4. **ANDERSON, James C.; & GERBING, David W.** «Structural equation modeling in practice: A review and recommended two-step approach.» *Psychological Bulletin* [en línea], 1988, 103(3), pp.411-423. [Consulta: 10 julio 2023]. ISSN 1939-1455(Electronic),0033-2909(Print). DOI 10.1037/0033-2909.103.3.411. Disponible en: <https://psycnet.apa.org/doi/10.1037/0033-2909.103.3.411>
5. **BAKER, Ryan S. J. d.** «Data Mining». *International Encyclopedia of Education* [en línea]. 3ª ed. Oxford: Elsevier, 2010. ISBN 978-0-08-044894-7, pp.112-118. Disponible en: <https://www.sciencedirect.com/science/article/pii/B978008044894701318X>
6. **BAKER, Ryan; & SIEMENS, George.** «Educational Data Mining and Learning Analytics». *The Cambridge Handbook of the Learning Sciences* [en línea]. Cambridge University Press, 2014. pp.253-272. [Consulta: 26 abril 2023]. Disponible en: https://www.cambridge.org/core/product/identifier/9781139519526%23c03325-13-1/type/book_part
7. **BANDYOPADHYAY, Susmita.** *Decision support system: Tools and techniques* [en línea]. Boca Raton: CRC Press, 2023. ISBN 9781003307655. [Consulta: 16 abril 2023]. Disponible en: <https://doi.org/10.1201/9781003307655>
8. **BANERJEE, Arindam et al.** «Data Analytics: Hyped Up Aspirations or True Potential?» *Vikalpa* [en línea], 2013, 38(4), pp.1-12. DOI 10.1177/0256090920130401. Disponible en: <https://doi.org/10.1177/0256090920130401>
9. **BARKVED, Kirsten.** *How To Know if Your Machine Learning Model Has Good Performance - Obviously AI* [blog]. 9 de marzo de 2022. [Consulta: 12 agosto 2023]. Disponible en: <https://www.obviously.ai/post/machine-learning-model-performance>.
10. **BELCIUG, Smaranda; & GORUNESCU, Florin.** «A Brief History of Intelligent Decision Support Systems». *Intelligent Decision Support Systems—A Journey to Smarter Healthcare* [en línea]. Cham: Springer International Publishing, 2019. ISBN 978-3-030-14354-1, pp.57-70. [Consulta: 19 febrero 2023]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-030-14354-1_2
11. **BHARGAVA, Hemant et al.** «Progress in Web-based decision support technologies». *Decision Support Systems* [en línea], 2007, 43(4), pp.1083-1095. [Consulta: 21 enero 2023]. ISSN 0167-9236. DOI 10.1016/J.DSS.2005.07.002. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S0167923605001016>

12. **CALDERS, Toon; & CUSTERS, Bart.** «What Is Data Mining and How Does It Work?» *Discrimination and Privacy in the Information Society: Data Mining and Profiling in Large Databases* [en línea]. Berlin, Heidelberg: Springer Berlin Heidelberg, 2013. ISBN 978-3-642-30487-3, pp.27-42. Disponible en: https://doi.org/10.1007/978-3-642-30487-3_2
13. **CAMACHO, David et al.** «The four dimensions of social network analysis: An overview of research methods, applications, and software tools». *Information Fusion* [en línea], 2020, 63, pp.88-120. ISSN 1566-2535. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1566253520302906>
14. **CHACON, Scott; & STRAUB, Ben.** *Pro Git* [en línea]. 2ª ed. Apress, 2014. [Consulta: 7 noviembre 2023]. Disponible en: <https://doi.org/10.1007/978-1-4842-0076-6>
15. **CHAPMAN, Pete et al.** *CRISP-DM 1.0: Step-by-step data mining guide*. SPSS, 2000.
16. **CHIAVENATO, Idalberto.** *Comportamiento organizacional - La dinámica del éxito en las organizaciones*. 2ª ed. México: McGraw-Hill, 2009. ISBN 978-970-10-6876-2.
17. **CLOW, Doug.** «An overview of learning analytics». *Teaching in Higher Education* [en línea], 2013, 18(6), pp.683-695. [Consulta: 12 mayo 2023]. DOI 10.1080/13562517.2013.827653. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/13562517.2013.827653>
18. **COBOS, Carlos et al.** «CMIN - herramienta case basada en CRISP-DM para el soporte de proyectos de minería de datos». *Ingeniería e Investigación* [en línea], 2010, 30(2), pp.45-56. [Consulta: 12 mayo 2023]. Disponible en: <https://www.elibro.net/es/lc/epoch/titulos/96335>
19. **CRAMER, Elizabeth D. et al.** «Demystifying the Data-Based Decision-Making Process». *Action in Teacher Education* [en línea], 2014, 36(5-6), pp.389-400. [Consulta: 19 febrero 2023]. ISSN 21586098. DOI 10.1080/01626620.2014.977690. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/01626620.2014.977690>
20. **DALIANIS, Hercules.** «Evaluation Metrics and Evaluation». *Clinical Text Mining* [en línea]. Springer, Cham, 2018. pp.45-53. [Consulta: 8 agosto 2023]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-78503-5_6
21. **DASTRES, Roza; & SOORI, Mohsen.** «Advances in Web-Based Decision Support Systems». *International Journal of Engineering and Future Technology* [en línea], 2021, 19(1), pp.1-15. ISSN 2455-6432. Disponible en: <https://hal.science/hal-03367778>
22. **DATNOW, Amanda et al.** «Data use for equity: implications for teaching, leadership, and policy». *Journal of Educational Administration* [en línea], 2017, 55(4), pp.354-360. ISSN 0957-8234. DOI 10.1108/JEA-04-2017-0040. Disponible en: <https://doi.org/10.1108/JEA-04-2017-0040>
23. **DELEN, Dursun.** *Real-world data mining: applied business analytics and decision making*. Pearson Educacion, 2015. ISBN 978-0-13-355107-5.
24. **DELONE, William; & MCLEAN, Ephraim.** «Information Systems Success: The Quest for the Dependent Variable». *Information Systems Research* [en línea], 1992, 3(1), pp.60-95. [Consulta: 1 enero 2023]. ISSN 15265536. DOI 10.1287/ISRE.3.1.60. Disponible en: <https://pubsonline.informs.org/doi/abs/10.1287/isre.3.1.60>
25. **DELONE, William; & MCLEAN, Ephraim.** «The DeLone and McLean Model of Information Systems Success: A Ten-Year Update». *Journal of Management Information Systems* [en línea], 2003, 19(4), pp.9-30. ISSN 0742-1222. DOI 10.1080/07421222.2003.11045748. Disponible en: <https://doi.org/10.1080/07421222.2003.11045748>

26. **DEWANTORO, Gunawan; & ARDISA, Nathanael.** «A Decision Support System for Undergraduate Students Admissions using Educational Data Mining». *2020 7th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)* [en línea], 2020, pp.105-109. [Consulta: 11 mayo 2023]. DOI 10.1109/ICITACEE50144.2020.9239244. Disponible en: <https://ieeexplore.ieee.org/document/9239244>
27. **DIVÁN, M. J.** «Data-driven decision making». *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions) (ICTUS)* [en línea], 2017, pp.50-56. [Consulta: 11 mayo 2023]. DOI 10.1109/ICTUS.2017.8285973. Disponible en: <https://ieeexplore.ieee.org/document/8285973>
28. **DOLL, William J.; & TORKZADEH, Gholamreza.** «Developing a multidimensional measure of system-use in an organizational context». *Information & Management* [en línea], 1998, 33(4), pp.171-185. ISSN 0378-7206. DOI 10.1016/S0378-7206(98)00028-7. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0378720698000287>
29. **DONHOST, Michael J.; & ANFARA, Vincent A.** «Data-Driven Decision Making». *Middle School Journal* [en línea], 2015, 42(2), pp.56-63. [Consulta: 7 diciembre 2022]. ISSN 23276223. DOI 10.1080/00940771.2010.11461758. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/00940771.2010.11461758>
30. **DOS SANTOS GARCIA, Cleiton et al.** «Process mining techniques and applications – A systematic mapping study». *Expert Systems with Applications* [en línea], 2019, 133, pp.260-295. ISSN 0957-4174. DOI 10.1016/j.eswa.2019.05.003. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0957417419303161>
31. **EICHHORN, Peter; & TOWERS, Ian.** «Managerial Methods». *Principles of Management: Efficiency and Effectiveness in the Private and Public Sector* [en línea]. Cham: Springer International Publishing, 2018. ISBN 978-3-319-70902-4, pp.263-312. [Consulta: 13 mayo 2023]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-70902-4_9
32. **ESQUIVEL, Joseph A.; & ESQUIVEL, James A.** «A Machine Learning Based DSS in Predicting Undergraduate Freshmen Enrolment in a Philippine University». *International Journal of Computer Trends and Technology* [en línea], 2021, 69(5), pp.50-54. [Consulta: 6 noviembre 2022]. DOI 10.14445/22312803/IJCTT-V69I5P107. Disponible en: <https://ijcttjournal.org/archives/ijctt-v69i5p107>
33. **FAYYAD, Usama et al.** «From Data Mining to Knowledge Discovery in Databases». *AI Magazine* [en línea], 1996, 17(3), p.37. DOI 10.1609/aimag.v17i3.1230. Disponible en: <https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/1230>
34. **GAIKWAD, Sonali Vijay et al.** «Text mining methods and techniques». *International Journal of Computer Applications* [en línea], 2014, 85(17), [Consulta: 13 mayo 2023]. DOI 10.5120/14937-3507. Disponible en: <https://www.ijcaonline.org/archives/volume85/number17/14937-3507>
35. **GÉRON, Aurélien.** *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*. 2ª ed. O'Reilly Media, 2019. ISBN 9781492032649.
36. **GUZMÁN-VALENZUELA, Carolina et al.** «Learning analytics in higher education: a preponderance of analytics but very little learning?». *International Journal of Educational Technology in Higher Education* [en línea], 2021, 18(1), p.23. [Consulta: 13 mayo 2023]. ISSN 2365-9440. DOI 10.1186/s41239-021-00258-x. Disponible en: <https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-021-00258-x>

37. **HA, Jiawei et al.** *Data Mining: Concepts and Techniques* [en línea]. 3ª ed. Elsevier, 2011. ISBN 9780123814791. [Consulta: 7 noviembre 2023]. Disponible en: <http://www.sciencedirect.com:5070/book/9780123814791/data-mining-concepts-and-techniques>
38. **HAIR, Joe F. et al.** «PLS-SEM: Indeed a Silver Bullet». *Journal of Marketing Theory and Practice* [en línea], 2011, 19(2), pp.139-152. ISSN 1069-6679. DOI 10.2753/MTP1069-6679190202. Disponible en: <https://doi.org/10.2753/MTP1069-6679190202>
39. **HAIR, Joseph F. et al.** *Partial Least Squares Structural Equation Modeling (PLS-SEM) Using R* [en línea]. Classroom Companion: Business. Cham: Springer International Publishing, 2021. ISBN 978-3-030-80518-0. [Consulta: 31 julio 2023]. Disponible en: <https://link.springer.com/10.1007/978-3-030-80519-7>
40. **IIVARI, Juhani.** «An Empirical Test of the DeLone-McLean Model of Information System Success». *SIGMIS Database* [en línea], 2005, 36(2), pp.8-27. [Consulta: 4 junio 2023]. ISSN 0095-0033. DOI 10.1145/1066149.1066152. Disponible en: <https://dl.acm.org/doi/abs/10.1145/1066149.1066152>
41. **ISO/IEC 25010.** *Systems and software engineering — Systems and software Quality Requirements and Evaluation (SQuaRE) — System and software quality models.* 2011.
42. **JUHAŇÁK, Libor et al.** «Using process mining to analyze students' quiz-taking behavior patterns in a learning management system». *Computers in Human Behavior* [en línea], 2019, 92, pp.496-506. ISSN 0747-5632. DOI 10.1016/j.chb.2017.12.015. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0747563217306957>
43. **KELLY, Sloan.** «What Is Python?» *Python, PyGame and Raspberry Pi Game Development* [en línea]. Berkeley: Apress, 2016. ISBN 978-1-4842-2517-2, pp.3-5. [Consulta: 13 mayo 2023]. Disponible en: https://link.springer.com/chapter/10.1007/978-1-4842-2517-2_2
44. **KHALIL, Mohammad; & EBNER, Martin.** «Learning Analytics: Principles and Constraints». *World Conference on Educational Multimedia, Hypermedia and Telecommunications 2015* [en línea], 2015, (Chesapeake, Estados Unidos), pp.1326-1336. [Consulta: 11 mayo 2023]. DOI 10.13140/RG.2.1.1733.2083. Disponible en: <http://dx.doi.org/10.13140/RG.2.1.1733.2083>
45. **KIM, Jin Ju et al.** «A systematic literature review on humanitarian logistics using network analysis and topic modeling». *The Asian Journal of Shipping and Logistics* [en línea], 2022, 38(4), pp.263-278. ISSN 2092-5212. DOI <https://doi.org/10.1016/j.ajsl.2022.10.003>. Disponible en: <https://www.sciencedirect.com/science/article/pii/S2092521222000323>
46. **KIM, Junghee et al.** «A Study on User Satisfaction regarding the Clinical Decision Support System (CDSS) for Medication». *Healthcare Informatics Research* [en línea], 2012, 18(1), pp.35-43. [Consulta: 13 mayo 2023]. ISSN 2093-3681. DOI 10.4258/hir.2012.18.1.35. Disponible en: <http://e-hir.org/journal/view.php?number=668>
47. **LEE, Yang W. et al.** «AIMQ: a methodology for information quality assessment». *Information & Management* [en línea], 2002, 40(2), pp.133-146. ISSN 0378-7206. DOI 10.1016/S0378-7206(02)00043-5. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0378720602000435>
48. **LEITNER, Philipp et al.** «Learning Analytics in Higher Education-A Literature Review». *Learning Analytics: Fundamentals, Applications, and Trends: A View of the Current State of the Art to Enhance e-Learning* [en línea]. Cham: Springer International Publishing, 2017. ISBN 978-3-319-52977-6, pp.1-23. [Consulta: 13 mayo 2023]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-52977-6_1

49. **LEPENIOTI, Katerina et al.** «Prescriptive analytics: Literature review and research challenges». *International Journal of Information Management* [en línea], 2020, 50, pp.57-70. ISSN 0268-4012. DOI 10.1016/j.ijinfomgt.2019.04.003. Disponible en: <https://www.sciencedirect.com/science/article/pii/S0268401218309873>
50. **LUTZ, M.** *Programming Python* [en línea]. Nutshell handbook. O'Reilly Media, Incorporated, 2011. ISBN 9780596158101. Disponible en: <https://books.google.com.ec/books?id=etibAgAAQBAJ>
51. **MCDONALD, Andy.** *Getting Started With Streamlit Web Based Applications* [blog]. 2022. [Consulta: 3 marzo 2023]. Disponible en: <https://towardsdatascience.com/getting-started-with-streamlit-web-based-applications-626095135cb8>.
52. **MENDOZA, María; & PÁRRAGA, María.** Sistema de información web sobre las características socioeconómicas del alumnado de la ESPAM MFL como apoyo a la toma de decisiones (Trabajo de titulación) (Pregrado) [en línea]. Escuela Superior Politécnica Agropecuaria de Manabí Manuel Félix López. Calceta, Ecuador. 2017. [Consulta: 10 diciembre 2022]. Disponible en: <http://repositorio.espam.edu.ec/handle/42000/476>.
53. **MIMNO, David et al.** «Optimizing Semantic Coherence in Topic Models». *Proceedings of the 2011 Conference on Empirical Methods in Natural Language Processing* [en línea], 2011, (Edinburgh, Scotland, UK.), pp.262-272. Disponible en: <https://aclanthology.org/D11-1024>
54. **MITTAL, Mamta et al.** «Clustering approaches for high-dimensional databases: A review». *WIREs Data Mining and Knowledge Discovery* [en línea], 2019, 9(3), pp.1-14. DOI 10.1002/widm.1300. Disponible en: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1300>
55. **MOSCOSO-ZEA, Oswaldo et al.** «Evaluation of algorithms to predict graduation rate in higher education institutions by applying educational data mining». *Australasian Journal of Engineering Education* [en línea], 2019, 24(1), pp.4-13. [Consulta: 19 noviembre 2022]. ISSN 13254340. DOI 10.1080/22054952.2019.1601063. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/22054952.2019.1601063>
56. **MUNIER, Nolberto et al.** *Strategic Approach in Multi-Criteria Decision Making* [en línea]. International Series in Operations Research & Management Science. 1ª ed. Cham: Springer International Publishing, 2019. ISBN 978-3-030-02725-4. [Consulta: 15 abril 2023]. Disponible en: <https://link.springer.com/book/10.1007/978-3-030-02726-1>
57. **MURRAY, John.** «Critical issues facing school leaders concerning data-informed decision-making». *Professional Educator* [en línea], 2014, 38(1), [Consulta: 29 mayo 2023]. ISSN 0196-786X. Disponible en: <https://eric.ed.gov/?id=EJ1038162>
58. **NGUYEN, Thanh D. et al.** «Information Systems Success: A Literature Review». *Future Data and Security Engineering* [en línea], 2015, pp.242-256. [Consulta: 16 mayo 2023]. DOI 10.1007/978-3-319-26135-5_18. Disponible en: https://doi.org/10.1007/978-3-319-26135-5_18
59. **NIETO, Yuri et al.** «Supporting academic decision making at higher educational institutions using machine learning-based algorithms». *Soft Computing* [en línea], 2019, 23(12), pp.4145-4153. ISSN 1433-7479. DOI 10.1007/s00500-018-3064-6. Disponible en: <https://doi.org/10.1007/s00500-018-3064-6>
60. **PAZ, Fábio; & CAZELLA, Silvio.** «Academic analytics: a systematic review of literature». *International Journal of Development Research* [en línea], 2019, 9(11), pp.31710-31716. [Consulta: 29 mayo 2023]. ISSN 2230-9926. Disponible en: <http://www.journalijdr.com/academic-analytics-systematic-review-literature>

61. **PECK, Roxy et al.** *Introduction to Statistics and Data Analysis*. 6ª ed. Boston: Cengage Learning, 2019. ISBN 9781337793612.
62. **PETTER, Stacie et al.** «Measuring information systems success: models, dimensions, measures, and interrelationships». *European Journal of Information Systems* [en línea], 2008, 17(3), pp.236-263. ISSN 1476-9344. DOI 10.1057/ejis.2008.15. Disponible en: <https://doi.org/10.1057/ejis.2008.15>
63. **PIATETSKY-SHAPIRO, Gregory.** «Knowledge discovery in databases: 10 years after». *ACM SIGKDD Explorations Newsletter* [en línea], 2000, 1(2), pp.59-61. [Consulta: 21 febrero 2023]. ISSN 1931-0145. DOI 10.1145/846183.846197. Disponible en: <https://dl.acm.org/doi/10.1145/846183.846197>
64. **PUTLER, Daniel S.; & KRIDER, Robert E.** «A Process Model for Data Mining—CRISP-DM». *Customer and Business Analytics*. CRC Press, 2012. ISBN 978-1-4665-0398-4, pp.17-29.
65. **RAMALHO, Lucca de Farias; & SEGUNDO, Washington R. de Carvalho.** «R-Shiny as an Interface for Data Visualization and Data Analysis on the Brazilian Digital Library of Theses and Dissertations (BDTD)». *Publications* [en línea], 2020, 8(2), pp.1-8. ISSN 2304-6775. DOI 10.3390/publications8020024. Disponible en: <https://www.mdpi.com/2304-6775/8/2/24>
66. **RASCHKA, Sebastian et al.** *Machine Learning with PyTorch and Scikit-Learn: Develop machine learning and deep learning models with Python* [en línea]. Birmingham: Packt Publishing, 2022. ISBN 9781801816380. Disponible en: <https://books.google.com.ec/books?id=SVxaEAAAQBAJ>
67. **RASHIDI, Maria et al.** «Decision Support Systems». *Management of Information Systems* [en línea]. Rijeka: IntechOpen, 2018. Disponible en: <https://doi.org/10.5772/intechopen.79390>
68. **ŘEHUŘEK, Radim; & SOJKA, Petr.** «Software Framework for Topic Modelling with Large Corpora». *Proceedings of the LREC 2010 Workshop on New Challenges for NLP Frameworks* [en línea], 2010, (Valletta, Malta), pp.45-50. [Consulta: 8 agosto 2023]. Disponible en: <https://repozitar.cz/publication/15725/?lang=cs;kod=S530>
69. **RINGLE, Christian M. et al.** *SmartPLS 4* [en línea]. Oststeinbek: SmartPLS, 2022. Disponible en: <https://www.smartpls.com/>
70. **RISING, L.; & JANOFF, N. S.** «The Scrum software development process for small teams». *IEEE Software* [en línea], 2000, 17(4), pp.26-32. DOI 10.1109/52.854065. Disponible en: <https://ieeexplore.ieee.org/document/854065>
71. **RÖDER, Michael et al.** «Exploring the Space of Topic Coherence Measures». *Proceedings of the Eighth ACM International Conference on Web Search and Data Mining* [en línea], 2015, (New York, Estados Unidos), pp.399-408. DOI 10.1145/2684822.2685324. Disponible en: <https://doi.org/10.1145/2684822.2685324>
72. **RODRÍGUEZ, Alberto; & BERNAL, Elizabeth.** *Gestión de la información cuantitativa en las universidades. Pistas para su abordaje en la era de la sobreinformación* [en línea]. 1ª ed. Bogotá: Universidad Nacional de Colombia, 2019. ISBN 978-958-783-805-3. [Consulta: 21 febrero 2023]. Disponible en: https://estadisticaun.github.io/L_Conceptual/
73. **ROMERO, Cristobal; & VENTURA, Sebastian.** «Data mining in education». *WIREs Data Mining and Knowledge Discovery* [en línea], 2013, 3(1), pp.12-27. DOI <https://doi.org/10.1002/widm.1075>. Disponible en: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1075>

74. **ROMERO, Cristobal; & VENTURA, Sebastian.** «Educational data mining and learning analytics: An updated survey». *WIREs Data Mining and Knowledge Discovery* [en línea], 2020, 10(3), DOI <https://doi.org/10.1002/widm.1355>. Disponible en: <https://wires.onlinelibrary.wiley.com/doi/abs/10.1002/widm.1355>
75. **RUBIN, Kenneth S.** *Essential Scrum: A Practical Guide to the Most Popular Agile Process*. 1stª ed. Addison-Wesley Professional, 2012. ISBN 0137043295.
76. **RUNKLER, Thomas A.** *Data Analytics* [en línea]. 3ª ed. Wiesbaden: Springer Fachmedien Wiesbaden, 2020. ISBN 978-3-658-29778-7. [Consulta: 7 diciembre 2022]. Disponible en: <http://link.springer.com/10.1007/978-3-658-29779-4>
77. **SALIHOUN, Mohammed.** «State of Art of Data Mining and Learning Analytics Tools in Higher Education». *International Journal of Emerging Technologies in Learning (iJET)* [en línea], 2020, 15(21), pp.58–76. DOI 10.3991/ijet.v15i21.16435. Disponible en: <https://online-journals.org/index.php/i-jet/article/view/16435>
78. **SALVATORE, Dominick; & REAGLE, Derrick.** *Schaum's Outline of Statistics and Econometrics*. 2ª ed. McGraw-Hill, 2011. ISBN 9780071755474.
79. **SANTOS, Alejandra et al.** «Identification of Characteristics and Functionalities for the Design of an Academic Analytics Model for Higher Education». *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality* [en línea], 2021, (New York, Estados Unidos), pp.997-1003. DOI 10.1145/3434780.3436578. Disponible en: <https://doi.org/10.1145/3434780.3436578>
80. **SCHILDKAMP, Kim.** «Data-based decision-making for school improvement: Research insights and gaps». *Educational Research* [en línea], 2019, 61(3), pp.257-273. [Consulta: 7 diciembre 2022]. ISSN 14695847. DOI 10.1080/00131881.2019.1625716. Disponible en: <https://www.tandfonline.com/doi/abs/10.1080/00131881.2019.1625716>
81. **SCHRÖER, Christoph et al.** «A Systematic Literature Review on Applying CRISP-DM Process Model». *Procedia Computer Science* [en línea], 2021, 181, pp.526-534. ISSN 1877-0509. DOI <https://doi.org/10.1016/j.procs.2021.01.199>. Disponible en: <https://www.sciencedirect.com/science/article/pii/S1877050921002416>
82. **SCHWABER, Ken; & SUTHERLAND, Jeff.** *The Scrum Guide The Definitive Guide to Scrum: The Rules of the Game*. s.n., 2020.
83. **SHANNON, C. E.; & WEAVER, W.** *The Mathematical Theory of Communication* [en línea]. Illinois: University of Illinois Press, 1949. ISBN 9780252725487. [Consulta: 21 mayo 2023]. Disponible en: https://books.google.com.ec/books/about/The_Mathematical_Theory_of_Communication.html?id=k0n_eGcqsUC&redir_esc=y
84. **SIEMENS, George; & LONG, Phil.** «Penetrating the fog: Analytics in learning and education». *EDUCAUSE review* [en línea], 2011, 46(5), pp.31-40. Disponible en: <https://er.educause.edu/articles/2011/9/penetrating-the-fog-analytics-in-learning-and-education>
85. **ŠIKŠNYŠ, Laurynas; & PEDERSEN, Torben Bach.** «Prescriptive Analytics». *Encyclopedia of Database Systems* [en línea]. New York: Springer New York, 2018. ISBN 978-1-4614-8265-9, pp.2792-2793. Disponible en: https://doi.org/10.1007/978-1-4614-8265-9_80624
86. **SIMON, H. A.** *Administrative Behavior: A Study of Decision-Making Processes in Administrative Organization* [en línea]. 4ª ed. Free Press, 1997. ISBN 9780684835822. Disponible en: <https://books.google.com.ec/books?id=jmzWLn8pBKUC>

87. **SOLTANPOOR, Reza; & SELLIS, Timos.** «Prescriptive Analytics for Big Data». *Databases Theory and Applications. ADC 2016* [en línea]. Cham: Springer International Publishing, 2016. ISBN 978-3-319-46922-5, pp.245-256. [Consulta: 11 mayo 2023]. Disponible en: https://link.springer.com/chapter/10.1007/978-3-319-46922-5_19
88. **SOMMERVILLE, Ian.** *Ingeniería de software* [en línea]. 9ª ed. México D.F.: Pearson Educacion, 2011. ISBN 9786073206044. Disponible en: <https://elibro.net/es/lc/epoch/titulos/37857>
89. **SUTHERLAND, Jeff; & COPLIEN, James O.** *A scrum book: the spirit of the game.* Pragmatic Bookshelf, 2019. ISBN 9781680506716.
90. **TAULÉ, Mariona et al.** «AnCora: Multilevel Annotated Corpora for Catalan and Spanish». *Proceedings of the Sixth International Conference on Language Resources and Evaluation (LREC'08)* [en línea], 2008, (Marrakech, Morocco), Disponible en: http://www.lrec-conf.org/proceedings/lrec2008/pdf/35_paper.pdf
91. **URBACH, Nilsand; & MÜLLER, Benjamin.** «The Updated DeLone and McLean Model of Information Systems Success». *Information Systems Theory. Integrated Series in Information Systems, vol 28* [en línea]. New York: Springer, 2012. ISBN 978-1-4419-6108-2, pp.1-18. Disponible en: https://doi.org/10.1007/978-1-4419-6108-2_1
92. **VAN ROSSUM, Guido et al.** «PEP 8 - Style Guide for Python Code». Python Enhancement Proposals, 2001. [Consulta: 10 junio 2023]. Disponible en: <https://peps.python.org/pep-0008/>
93. **VÁZQUEZ-INGELMO, Andrea et al.** «C4 model in a Software Engineering subject to ease the comprehension of UML and the software». *2020 IEEE Global Engineering Education Conference (EDUCON)* [en línea], 2020, pp.919-924. [Consulta: 10 junio 2023]. DOI 10.1109/EDUCON45650.2020.9125335. Disponible en: <https://ieeexplore.ieee.org/abstract/document/9125335>
94. **VON KROGH, Georg; & SPAETH, Sebastian.** «The open source software phenomenon: Characteristics that promote research». *The Journal of Strategic Information Systems* [en línea], 2007, 16(3), pp.236-253. [Consulta: 7 noviembre 2023]. ISSN 0963-8687. DOI 10.1016/J.JSIS.2007.06.001. Disponible en: <https://www.sciencedirect.com/science/article/abs/pii/S096386870700025X>
95. **WANG, Hongzhi et al.** «Progress in Outlier Detection Techniques: A Survey». *IEEE Access* [en línea], 2019, 7, pp.107964-108000. ISSN 21693536. DOI 10.1109/ACCESS.2019.2932769. Disponible en: <https://ieeexplore.ieee.org/document/8786096>
96. **WEBBER, Karen L.; & ZHENG, Henry Y.** *Data analytics and the imperatives for data-informed decision-making in higher education* [en línea]. Johns Hopkins University Press, 2020. ISBN 9781421439044. [Consulta: 15 febrero 2023]. Disponible en: <https://muse.jhu.edu/book/78960>
97. **WISE, Alyssa; & JUNG, Yeonji.** «Teaching with analytics: Towards a situated model of instructional decision-making». *Journal of Learning Analytics* [en línea], 2019, 6(2), pp.53-69. Disponible en: <https://doi.org/10.18608/jla.2019.62.4>
98. **ZABEO, Alex et al.** «A Decision Support System for preclinical assessment of nanomaterials in medical products: the REFINE DSS». *Drug Delivery and Translational Research* [en línea], 2022, 12(9), pp.2101-2113. ISSN 2190-3948. DOI 10.1007/s13346-022-01145-2. Disponible en: <https://doi.org/10.1007/s13346-022-01145-2>
99. **ZOYA et al.** «Analyzing LDA and NMF Topic Models for Urdu Tweets via Automatic Labeling». *IEEE Access* [en línea], 2021, 9, pp.127531-127547. [Consulta: 8 agosto 2023]. ISSN 21693536. DOI

10.1109/ACCESS.2021.3112620.

Disponible

en:

<https://ieeexplore.ieee.org/abstract/document/9536731>

100. **ZVIRAN, M.; & ERLICH, Z.** «Measuring IS User Satisfaction: Review and Implications». *Communications of the Association for Information Systems* [en línea], 2003, 12(1), p.5. [Consulta: 21 mayo 2023]. ISSN 1529-3181. DOI 10.17705/1CAIS.01205. Disponible en: <https://aisel.aisnet.org/cais/vol12/iss1/5>
101. **ZVORNICANIN, Enes.** *When Coherence Score Is Good or Bad in Topic Modeling? - Baeldung on Computer Science* [blog]. 31 de mayo de 2023. Disponible en: <https://www.baeldung.com/cs/topic-modeling-coherence-score>.

ANEXOS

ANEXO A: CUESTIONARIO PARA EVALUACIÓN DE SATISFACCIÓN DE USUARIO

CUESTIONARIO PARA LA EVALUACIÓN DE LA SATISFACCIÓN DE USUARIO

Descripción general

El presente cuestionario pretende valorar la satisfacción del usuario respecto al uso, calidad del sistema y calidad de la información de una aplicación web para el apoyo a la toma de decisiones.

Instrucciones

Evalúe en una escala de 1 al 5, de acuerdo con su experiencia utilizando la aplicación, su satisfacción respecto a cada parámetro descrito, siendo:

1: “Completamente en desacuerdo”; 2: “En desacuerdo”; 3: “Indiferente”; 4: “De acuerdo”; 5: “Completamente de acuerdo”

			1	2	3	4	5
Calidad del sistema	SQ1	La aplicación web es flexible al cambio en respuesta a nuevas demandas.					
	SQ2	El tiempo de respuesta de la aplicación web es rápido.					
Calidad de la información	IQ1	La información que provee la aplicación web es completa.					
	IQ2	La información que provee la aplicación es lo suficientemente actual.					
	IQ3	La información que provee la aplicación web es fácilmente interpretable.					
	IQ4	Las unidades de medida de la información que provee la aplicación web son claras.					
	IQ5	La información que provee la aplicación web es útil.					
	IQ6	La información que provee la aplicación web es relevante.					
Uso	USO1	Utiliza esta aplicación para tomar decisiones					
	USO2	Utiliza esta aplicación para ver la información de una manera innovadora, por ejemplo, para ver nuevas opciones y alternativas.					
	USO3	Utiliza esta aplicación para identificar problemas en el proceso de titulación.					
	USO4	Utiliza esta aplicación para apoyar la toma de sus decisiones					
	USO5	Utiliza esta aplicación para tomar decisiones más rápidamente					

	USO6	Utiliza esta aplicación para mejorar la efectividad y eficiencia del proceso de toma de decisiones.					
Satisfacción de usuario	STF1	En términos generales, ¿su experiencia con el sistema es satisfactoria?					
	STF2	En términos generales, ¿su experiencia con la información que provee el sistema es satisfactoria?					

ANEXO B: MANUAL TÉCNICO - SCRUM

MANUAL TÉCNICO

Scrum

IMAK - DSS

Junio, 2023

Introducción

La presente guía tiene como objetivo documentar la gestión del proyecto con el marco de trabajo Scrum. Dicha metodología permitió gestionar el desarrollo de la aplicación web para el apoyo a la toma de decisiones “IMAK”, así como de los modelos de analítica de datos implementados que se construyeron de forma paralela utilizando CRISP-DM (Para más información revise el **MANUAL TÉCNICO – CRISP-DM – IKMA DSS**).

1. Análisis preliminar

A través de entrevistas a los miembros de la Comisión de Titulación y el análisis de documentos, se pudo establecer los equipos y roles necesarios, así como los requerimientos iniciales del proyecto.

1.1 Módulos de la aplicación

La aplicación web para el apoyo a la toma de decisiones se ha dividido en cuatro módulos, tal y como se muestra en la *Ilustración 2*:

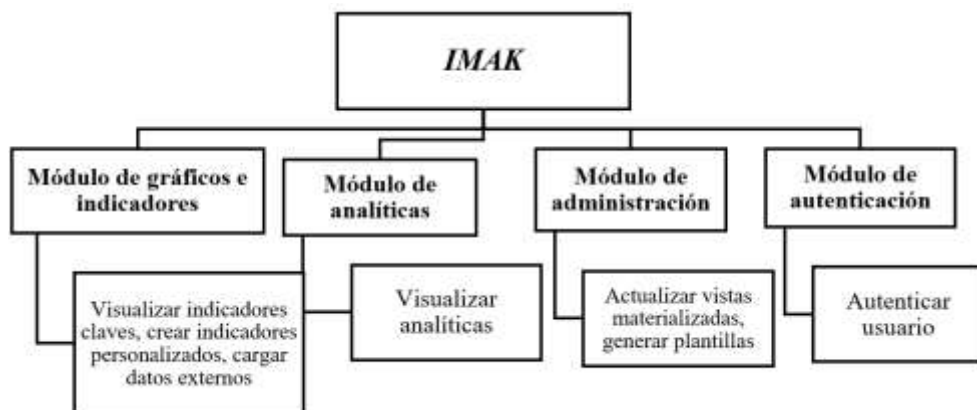


Ilustración 2: Módulos de la aplicación web "IKMA"

Módulo de gráficos e indicadores

El módulo de gráficos e indicadores agrupa las funcionalidades de la aplicación relacionadas con la visualización de indicadores clave en forma de gráficos y tablas, así como la construcción de indicadores personalizados.

Módulo de analíticas

El módulo de analíticas incluye los modelos desarrollados con CRISP-DM y las estadísticas descriptivas que se han propuesto de acuerdo con los requerimientos de Comisión y los datos que esta dispone.

Módulo de administración

El módulo de administración permite actualizar las vistas materializadas que almacenan los datos utilizados en la aplicación. También, incluye la opción de generar plantillas en formato .xlsx para cargar datos externos.

Módulo de autenticación

El módulo de autenticación permite a los usuarios iniciar sesión con las credenciales registradas en el sistema de gestión de trabajos de titulación de la carrera.

1.2 Diagramas de caso de uso

Los miembros de la Comisión de Titulación utilizarán la aplicación web en un contexto determinado, tal y como se muestra en los diagramas presentados a continuación:

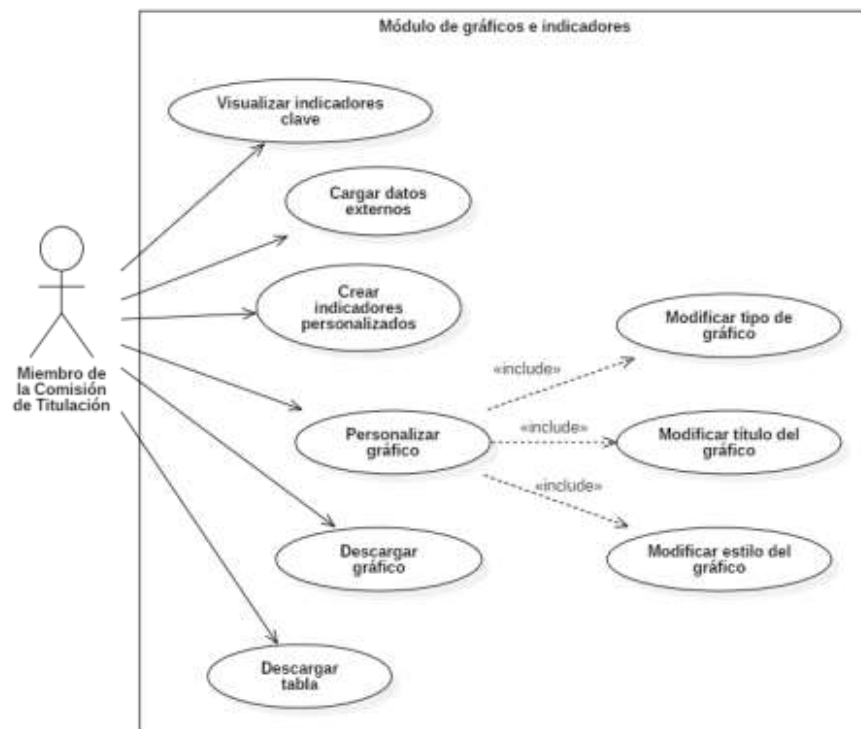


Ilustración 3: Diagrama de caso de uso del módulo de gráficos e indicadores

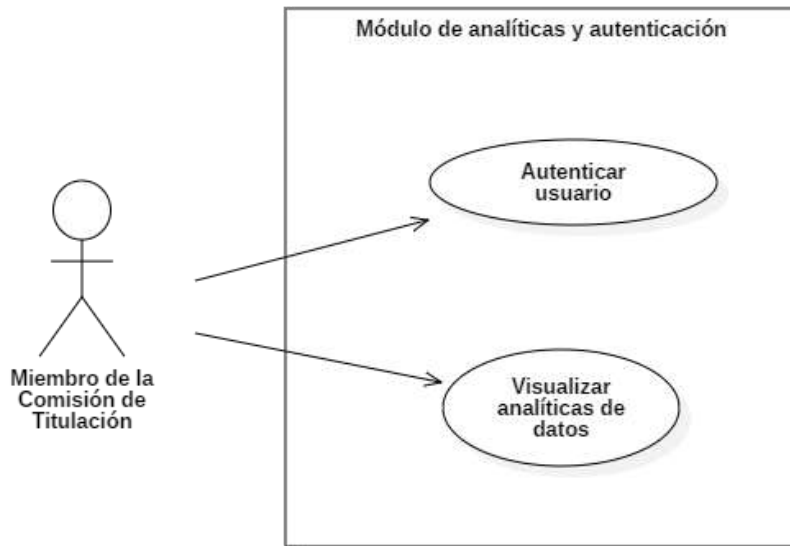


Ilustración 4: Diagrama de caso de uso de los módulos de analíticas y autenticación

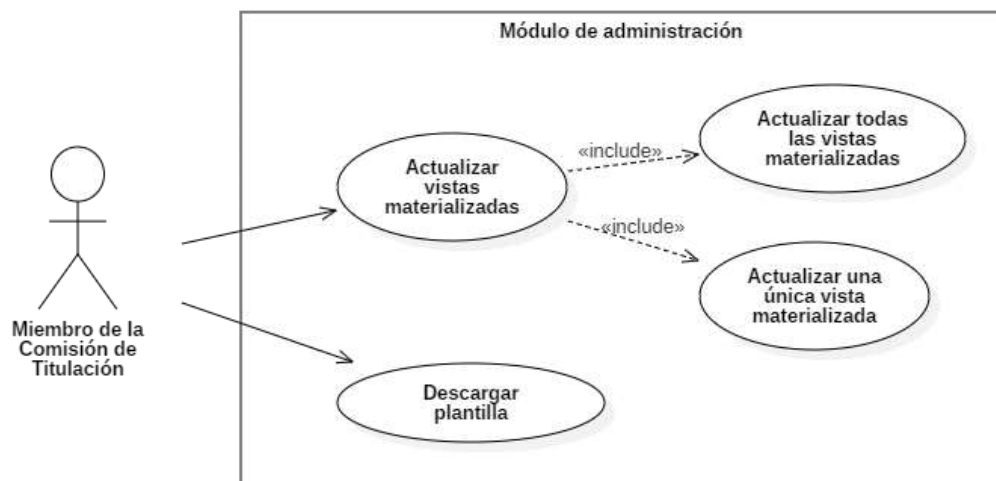


Ilustración 5: Diagrama de caso de uso del módulo de administración

1.3 Equipo y roles

En la construcción de la aplicación web intervino un grupo de actores conocidos como Scrum team, conformados por un Product owner, un Scrum master y el equipo de desarrollo (Tabla 8). Dicho equipo estuvo conformado por las siguientes personas y roles:

Tabla 8: Equipo Scrum para el desarrollo del proyecto

Persona	Contacto	Rol	Descripción
---------	----------	-----	-------------

Ing. Ivonne Rodríguez	ivonne.rodriguez@epoch.edu.ec	Scrum Master	Encargada de liderar el equipo y promover los valores y principios que guían el proceso de Scrum entre el resto de los integrantes.
Homero Barragán	homero.barragan@epoch.edu.ec	Product Owner	Su papel es servir de punto de conexión entre los intereses de los miembros de la Comisión de Titulación y el Scrum team, de tal forma que pueda canalizar las prioridades y necesidades de ambas partes.
Homero Barragán	homero.barragan@epoch.edu.ec	Equipo de desarrollo	Cumple la tarea de desarrollar la aplicación web de apoyo a la toma de decisiones correctamente.

2 Planificación

En esta sección se describen las actividades ejecutadas respecto a la planificación del proyecto, incluyendo el estudio de factibilidad, análisis de riesgos, estimaciones y organización del desarrollo de las funcionalidades de la solución.

2.1 Estudio de factibilidad técnica

El estudio de factibilidad técnica permite estimar la posibilidad de ejecución del proyecto. Se realiza un análisis respecto al *software* y *hardware* necesarios, comparándolos con los equipos y aplicaciones disponibles.

2.1.1 *Software*

Las herramientas de software que se disponen para el proyecto se presentan en la *Tabla 9* a continuación:

Tabla 9: Software disponible para el proyecto

Lenguaje de programación	Python
Librerías y frameworks	Streamlit
	Pandas
	Numpy
IDE	Visual Studio Code
Servidor y base de datos	Streamlit Cloud
	Postgresql
Sistema operativo	Windows 11
	macOS Ventura 13
Otros	Microsoft Office 365
	Github
	StarUML
	PgAdmin4
	Knime
	Navegador web
	Aplicación de notas de voz

El software requerido para el desarrollo y ejecución del proyecto se detalla en la *Tabla 10*:

Tabla 10: Software requerido para el proyecto

Lenguaje de programación	Python
Librerías y frameworks	Streamlit
	Pandas
	Scikit-learn
	Numpy
IDE	Visual Studio Code
Servidor y base de datos	Streamlit Cloud
	Postgresql
	Azure Data Services
Sistema operativo	Windows 11
	macOS Ventura 13
Otros	Microsoft Office 365
	Github

	StarUML
	PgAdmin4
	Knime
	Navegador web
	Aplicación de notas de voz

Para completar los requerimientos técnicos respecto al *software*, fue necesario descargar la librería Scikit-learn y contratar el servicio de Azure Data Services para el alojamiento de la base de datos.

2.1.2 Hardware

Para la ejecución del proyecto, se dispone de dos computadores y un dispositivo móvil. Sus características se detallan en la *Tabla 11*:

Tabla 11: Hardware disponible para el proyecto

Descripción	Características	Cantidad
MacBook Air M1 2020	Chip M1 de 8 núcleos	1
	GPU de 7 núcleos	
	8 GB RAM	
	512 GB SSD	
PC de escritorio	AMD Ryzen 7 5700G	1
	Radeon Graphics	
	16 GB RAM	
	512 GB HDD; 1 TB SSD	
iPhone 12 Pro Max	A14 Bionic	1
	6 GB RAM	
	128 GB de almacenamiento	

El *hardware* requerido y sus características mínimas para ejecutar el proyecto se presentan la siguiente tabla:

Tabla 12: Hardware requerido para el proyecto

Descripción	Características mínimas	Cantidad
Computadora	Intel Core i3 7th Gen	2

	Gráficos integrados	
	8 GB RAM	
	128 GB de almacenamiento	
Teléfono celular	Procesador de 1 Ghz	1
	500 MB RAM	
	32 GB de almacenamiento	

Una vez detallados los recursos con los que se cuenta y los requerimientos mínimos para el desarrollo del proyecto, se puede concluir que el proyecto es técnicamente factible.

2.2 Análisis de riesgos

Utilizando la estructura de descomposición de riesgos (EDR) que se muestra en la *Ilustración 6*, se han identificado las situaciones fortuitas que representan un mayor peligro para el desarrollo del proyecto:

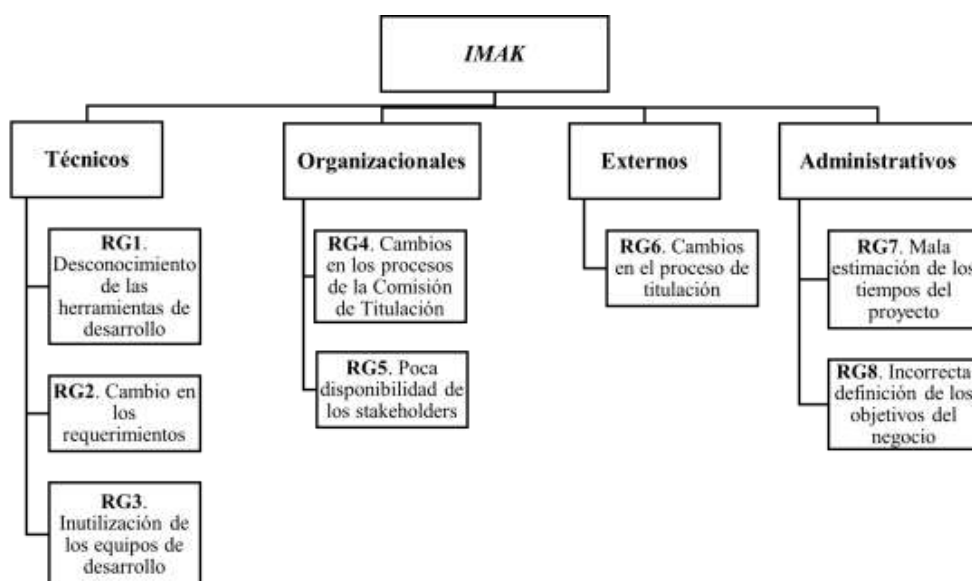


Ilustración 6: Estructura de descomposición de riesgos

Para estimar la probabilidad de ocurrencia de los riesgos identificados, se construyó la *Tabla 13* que reúne los valores numéricos asignados a cada rango de probabilidad:

Tabla 13: Valores asignados a los rangos de probabilidad

Rango de probabilidades	Nivel	Valor
1% al 20%	Muy baja	1
21% al 40%	Baja	2

41% al 60%	Media	3
61% al 80%	Alta	4
81% al 99%	Muy alta	5

De forma similar, se ha establecido los valores para el impacto que representa cada riesgos en el desarrollo del proyecto en la *Tabla 14*:

Tabla 14: Valores asignados a los niveles de impacto

Impacto	Impacto técnico	Valor
Muy bajo	No repercute en el desarrollo del sistema web	1
Bajo	Tiene un impacto ligero en el desarrollo del sistema	2
Moderado	Genera un impacto considerable en el proyecto	3
Alto	Tiene un efecto serio en el desarrollo del sistema web	4
Crítico	No permite el desarrollo del sistema web	5

Se construyó la *Tabla 15* con los niveles de gravedad estimados para cada riesgos:

Tabla 15: Equivalencias en colores de los niveles de gravedad

Nivel de gravedad	Color
Bajo	Verde
Moderado	Amarillo
Alto	Rojo

La matriz de gravedad del riesgos (*Ilustración 7*) muestra la prioridad de los riesgos respecto a su probabilidad de ocurrencia y nivel de impacto. La franja de color rojo, que indica un nivel de gravedad Alto, se extiende hacia la esquina inferior derecha ya que se ha priorizado el impacto sobre la probabilidad de ocurrencia.

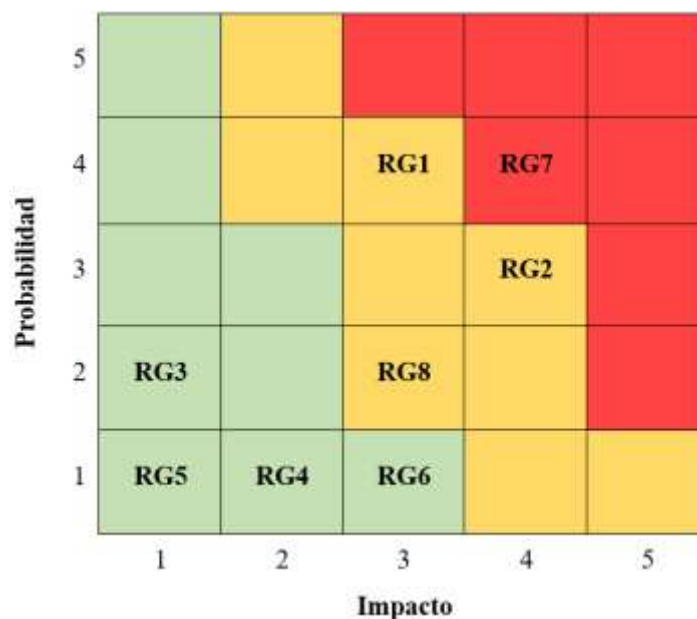


Ilustración 7: Matriz de gravedad de riesgos

Posteriormente, en la *Tabla 16*, se representaron los riesgos ordenados por prioridad, empezando con aquellos cuyo nivel de gravedad es Alto.

Tabla 16: Priorización de los riesgos identificados

Id	Descripción	Probabilidad			Impacto		Gravedad
		%	Nivel	Valor	Impacto	Valor	
RG7	Mala estimación de los tiempos del proyecto	80%	Alta	4	Alto	4	Alto
RG2	Cambio en los requerimientos	50%	Media	3	Alto	4	Moderado
RG1	Desconocimiento de las herramientas de desarrollo	75%	Alta	4	Moderado	3	Moderado
RG8	Incorrecta definición de los objetivos del negocio	25%	Baja	2	Moderado	3	Moderado
RG6	Cambios en el proceso de titulación	10%	Muy baja	1	Alto	3	Bajo
RG4	Cambios en los procesos de la Comisión de Titulación	10%	Muy baja	1	Bajo	2	Bajo

Id	Descripción	Probabilidad			Impacto		Gravedad
		%	Nivel	Valor	Impacto	Valor	
RG3	Inutilización de los equipos de desarrollo	25%	Baja	2	Muy bajo	1	Bajo
RG5	Poca disponibilidad de los <i>stakeholders</i>	10%	Muy baja	1	Muy bajo	1	Bajo

Finalmente, se diseñaron fichas con información y las acciones para prevenir y/o mitigar, supervisar y gestionar cada riesgo.

Tabla 17: Hoja de gestión del riesgo RG1

Ficha de gestión de riesgos		
ID: RG1		Fecha: 07/05/2023
Descripción: Desconocimiento de las herramientas de desarrollo		
Probabilidad de ocurrencia: Alta	Nivel de impacto: Moderado	Nivel de gravedad: Moderado
Refinamiento Causas: <ul style="list-style-type: none"> Poca familiarización con el <i>framework</i>, librerías y lenguaje de programación. Falta de práctica de las herramientas de desarrollo. Consecuencias: <ul style="list-style-type: none"> Incremento en los tiempos del proyecto. Mayor posibilidad de encontrar errores en el código. 		
Acciones de prevención y/o mitigación: <ul style="list-style-type: none"> Capacitación autónoma previo al inicio del proyecto y durante los fines de semana. Recurrir a foros y ayuda de expertos para solventar dudas respecto a las herramientas. Búsqueda y selección de herramientas alternas. 		
Supervisión: <ul style="list-style-type: none"> Evaluación continua de las habilidades respecto a las herramientas necesarias para el desarrollo del proyecto. 		
Acciones de gestión del riesgo: <ul style="list-style-type: none"> Incluir en las estimaciones los tiempos empleados en capacitación y corrección de errores. 		

Estado actual	Fase de reducción iniciada	
	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables: Homero Barragán		

Tabla 18: Hoja de gestión del riesgo RG2

Ficha de gestión de riesgos		
ID: RG2		Fecha: 07/05/2023
Descripción: Cambio en los requerimientos		
Probabilidad de ocurrencia: Media	Nivel de impacto: Alto	Nivel de gravedad: Moderado
Refinamiento		
Causas: <ul style="list-style-type: none"> • Necesidades cambiantes de los <i>stakeholders</i>. • Mala elicitación de requisitos. • Interpretación errónea de los requerimientos. • Falta de comunicación con los <i>stakeholders</i>. 		
Consecuencias: <ul style="list-style-type: none"> • Desarrollo de las funcionalidades incorrectas. • Incrementos en los tiempos del proyecto para corregir errores. • Retrasos en la entrega de incrementos. • Documentación errónea. 		
Acciones de prevención y/o mitigación: <ul style="list-style-type: none"> • Reuniones periódicas involucrando a los <i>stakeholders</i>. • Uso de más de una técnica para elicitación de requisitos. • Revisión constante del documento de requerimientos. • Revisión de los incrementos con los <i>stakeholders</i>. • Negociación de requisitos. 		
Supervisión: <ul style="list-style-type: none"> • Validación y verificación de requerimientos. 		
Acciones de gestión del riesgo:		

<ul style="list-style-type: none"> • Reuniones con los <i>stakeholders</i> para ajustar los requerimientos. • Refinamiento del documento de requisitos. 		
Estado actual	Fase de reducción iniciada	
	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables: Homero Barragán		

Tabla 19: Hoja de gestión del riesgo RG3

Ficha de gestión de riesgos		
ID: RG3		Fecha: 07/05/2023
Descripción: Inutilización de los equipos de desarrollo		
Probabilidad de ocurrencia: Baja	Nivel de impacto: Muy bajo	Nivel de gravedad: Bajo
Refinamiento Causas: <ul style="list-style-type: none"> • Posibles defectos físicos en los equipos de desarrollo. • Accidentes que involucren daños. • Hurto de los equipos de desarrollo. • Errores en el software. Consecuencias: <ul style="list-style-type: none"> • Potencial pérdida de la información del proyecto. • Retrasos en los tiempos del proyecto. • Incremento en los costos. 		
Acciones de prevención y/o mitigación: <ul style="list-style-type: none"> • Creación de copias de seguridad locales y en la nube de forma constante. • Utilización de un segundo computador para el desarrollo. 		
Supervisión: <ul style="list-style-type: none"> • Constatación regular del estado de las copias de seguridad y equipos alternos. 		
Acciones de gestión del riesgo: <ul style="list-style-type: none"> • Modificación del estudio de recursos para incluir un equipo alternativo para el desarrollo. 		
Estado actual	Fase de reducción iniciada	

	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables: Homero Barragán		

Tabla 20: Hoja de gestión del riesgo RG4

Ficha de gestión de riesgos		
ID: RG4		Fecha: 07/05/2023
Descripción: Cambios en los procesos de la Comisión de Titulación		
Probabilidad de ocurrencia: Muy baja	Nivel de impacto: Bajo	Nivel de gravedad: Bajo
Refinamiento		
Causas: <ul style="list-style-type: none"> • Reformas en el Reglamento del Régimen Académico. • Cambio de roles o personas de la Comisión. • Cambios en las tareas de la Comisión. 		
Consecuencias: <ul style="list-style-type: none"> • Necesidad de ajuste en los objetivos del negocio • Posibles cambios en los requerimientos 		
Acciones de prevención y/o mitigación: <ul style="list-style-type: none"> • Ajuste inmediato del proyecto ante posibles cambios. 		
Supervisión: <ul style="list-style-type: none"> • Revisión regular del Reglamento del Régimen Académico en busca de cambios. 		
Acciones de gestión del riesgo: <ul style="list-style-type: none"> • Reuniones con los <i>stakeholders</i> para identificar soluciones potenciales. 		
Estado actual	Fase de reducción iniciada	
	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables: Homero Barragán		

Tabla 21: Hoja de gestión del riesgo RG5

Ficha de gestión de riesgos	
ID: RG5	Fecha: 07/05/2023

Descripción: Poca disponibilidad de los <i>stakeholders</i>		
Probabilidad de ocurrencia: Muy baja	Nivel de impacto: Muy bajo	Nivel de gravedad: Bajo
Refinamiento Causas: <ul style="list-style-type: none"> • Los <i>stakeholders</i> no disponen del tiempo suficiente debido a sus ocupaciones. • Los <i>stakeholders</i> gozan de sus días de vacaciones. • Razones de causa mayor. Consecuencias: <ul style="list-style-type: none"> • Posibles retrasos en los tiempos del proyecto. • Potencial insatisfacción de los <i>stakeholders</i> respecto a ciertas funcionalidades. 		
Acciones de prevención y/o mitigación: <ul style="list-style-type: none"> • Reuniones con los demás <i>stakeholders</i> para negociar los requerimientos. • Conversar con los <i>stakeholders</i> para conocer sus horarios y disponibilidad de tiempo. 		
Supervisión: <ul style="list-style-type: none"> • Registro de las reuniones y asistencia de los <i>stakeholders</i>. 		
Acciones de gestión del riesgo: <ul style="list-style-type: none"> • Reuniones de priorización de requerimientos con los <i>stakeholders</i>. 		
Estado actual	Fase de reducción iniciada	
	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables: Homero Barragán		

Tabla 22: Hoja de gestión del riesgo RG6

Ficha de gestión de riesgos		
ID: RG6		Fecha: 07/05/2023
Descripción: Cambios en el proceso de titulación		
Probabilidad de ocurrencia: Muy baja	Nivel de impacto: Alto	Nivel de gravedad: Bajo

Refinamiento		
Causas:		
<ul style="list-style-type: none"> • Modificaciones en el proceso de titulación a nivel institucional. • Nuevos requerimientos para titularse a nivel de facultad. 		
Consecuencias:		
<ul style="list-style-type: none"> • Posibles cambios en el proyecto. • Retrasos en los tiempos del proyecto. • Aumento o reducción en el número de entregables. 		
Acciones de prevención y/o mitigación:		
<ul style="list-style-type: none"> • Incluir en las estimaciones un tiempo de holgura en el caso de presentarse inconvenientes. 		
Supervisión:		
<ul style="list-style-type: none"> • Constatación regular de los reglamentos institucionales e internos respecto a los procesos de titulación. 		
Acciones de gestión del riesgo:		
<ul style="list-style-type: none"> • Modificación de las estimaciones de tiempo para incluir el periodo de holgura. 		
Estado actual	Fase de reducción iniciada	
	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables:		
Homero Barragán		

Tabla 23: Hoja de gestión del riesgo RG7

Ficha de gestión de riesgos		
ID: RG7		Fecha: 07/05/2023
Descripción:		
Mala estimación de los tiempos del proyecto		
Probabilidad de ocurrencia:	Nivel de impacto:	Nivel de gravedad:
Alta	Alto	Alto

Refinamiento		
Causas:		
<ul style="list-style-type: none"> • Incorrecto análisis de los requerimientos. • Falta de capacitación respecto al uso de las tecnologías. • Situaciones adversas. 		
Consecuencias:		
<ul style="list-style-type: none"> • Extensión del tiempo requerido para culminar el proyecto. • Necesidad de reajustes en los entregables. • <i>Stakeholders</i> insatisfechos. 		
Acciones de prevención y/o mitigación:		
<ul style="list-style-type: none"> • Evaluación constante de la planificación del proyecto. • Incluir en las estimaciones un tiempo de holgura. 		
Supervisión:		
<ul style="list-style-type: none"> • Evaluación del documento de planificación y reajustes. 		
Acciones de gestión del riesgo:		
<ul style="list-style-type: none"> • Cambios en la planificación. • Modificación de las estimaciones de tiempo para incluir el periodo de holgura. 		
Estado actual	Fase de reducción iniciada	
	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables:		
Homero Barragán		

Tabla 24: Hoja de gestión del riesgo RG8

Ficha de gestión de riesgos		
ID: RG8		Fecha: 07/05/2023
Descripción:		
Incorrecta definición de los objetivos del negocio		
Probabilidad de ocurrencia:	Nivel de impacto:	Nivel de gravedad:
Baja	Moderado	Moderado

Refinamiento		
Causas:		
<ul style="list-style-type: none"> • Mala elicitación de requerimientos. • Cambios en los objetivos de los <i>stakeholders</i>. 		
Consecuencias:		
<ul style="list-style-type: none"> • Posibles cambios en el proyecto. • Retrasos en los tiempos del proyecto. 		
Acciones de prevención y/o mitigación:		
<ul style="list-style-type: none"> • Revisión y modificación de los documentos del proyecto. • Reuniones con los <i>stakeholders</i> para redefinir los objetivos del proyecto. 		
Supervisión:		
<ul style="list-style-type: none"> • Revisión de la planificación. 		
Acciones de gestión del riesgo:		
<ul style="list-style-type: none"> • Reunión con los <i>stakeholders</i> con el propósito de llegar a un acuerdo mutuo sobre los objetivos a cumplir. 		
Estado actual	Fase de reducción iniciada	
	Fase de supervisión iniciada	X
	Gestionando el riesgo	
Responsables:		
Homero Barragán		

2.3 Historias de usuario

Los requerimientos se recopilar utilizando historias de usuario, las cuales se presentan a continuación:

Tabla 25: Historia de usuario HU-01

ID	HU-01	Usuario	Miembro de la Comisión de Titulación
Nombre	Desarrollar analíticas de datos		
Descripción	Analizar la información recolectada para generar analíticas de datos para la Comisión		
Validación Los miembros de la Comisión pueden visualizar analíticas de datos basadas en la información que disponen.	Prioridad	Alta	
	Puntos	20	
	Estimación (Horas)	100	

Tabla 26: Historia de usuario HU-02

ID	HU-02	Usuario	Miembro de la Comisión de Titulación
Nombre	Autenticar usuario		
Descripción	Iniciar sesión en la aplicación web		
Validación Los miembros de la Comisión pueden iniciar sesión en la aplicación con las credenciales que se encuentran guardadas en la base de datos del sistema de gestión de trabajos de titulación.	Prioridad	Baja	
	Puntos	2	
	Estimación (Horas)	10	

Tabla 27: Historia de usuario HU-03

ID	HU-03	Usuario	Miembro de la Comisión de Titulación
Nombre	Visualizar analíticas de datos		
Descripción	Visualizar los gráficos de las analíticas de datos desarrolladas		
Validación Los miembros de la Comisión pueden visualizar los gráficos de las analíticas de datos en la aplicación web.	Prioridad	Media	
	Puntos	5	
	Estimación (Horas)	25	

Tabla 28: Historia de usuario HU-04

ID	HU-04	Usuario	Miembro de la Comisión de Titulación
Nombre	Visualizar indicadores clave		
Descripción	Visualizar los indicadores clave (vistas materializadas) en forma de gráficos y tablas.		
Validación Los miembros de la Comisión pueden visualizar los indicadores claves almacenados en vistas materializadas.	Prioridad	Alta	
	Puntos	5	
	Estimación (Horas)	25	

Tabla 29: Historia de usuario HU-05

ID	HU-05	Usuario	Miembro de la Comisión de Titulación
Nombre	Cargar datos externos		
Descripción	Seleccionar un archivo en formato .xlsx y cargar sus datos en la aplicación web.		
Validación Los miembros de la Comisión podrán seleccionar un archivo .xlsx que se unirá (JOIN) con una de las columnas del indicador escogido.	Prioridad	Media	
	Puntos	10	
	Estimación (Horas)	50	

Tabla 30: Historia de usuario HU-06

ID	HU-06	Usuario	Miembro de la Comisión de Titulación
Nombre	Crear indicadores personalizados		
Descripción	Permite crear un gráfico personalizado con los datos almacenados por la información.		
Validación Los miembros de la Comisión pueden crear su propio gráfico personalizando su contenido.	Prioridad	Media	
	Puntos	5	
	Estimación (Horas)	25	

Tabla 31: Historia de usuario HU-07

ID	HU-07	Usuario	Miembro de la Comisión de Titulación
Nombre	Modificar tipo de gráfico		
Descripción	Permite elegir el tipo de gráfico que desea crear.		
Validación Los usuarios pueden elegir el tipo de gráfico que desean crear.	Prioridad	Baja	
	Puntos	1	
	Estimación (Horas)	5	

Tabla 32: Historia de usuario HU-08

ID	HU-08	Usuario	Miembro de la Comisión de Titulación
Nombre	Modificar título del gráfico		

Descripción	Personalizar el título del gráfico creado por el usuario.		
Validación Los miembros de la Comisión pueden personalizar el título del gráfico.	Prioridad	Baja	
	Puntos	1	
	Estimación (Horas)	5	

Tabla 33: Historia de usuario HU-09

ID	HU-09	Usuario	Miembro de la Comisión de Titulación
Nombre	Modificar estilo del gráfico		
Descripción	El usuario puede cambiar el estilo del gráfico seleccionado.		
Validación Los miembros de la Comisión pueden cambiar el estilo de los gráficos en la aplicación.	Prioridad	Baja	
	Puntos	2	
	Estimación (Horas)	10	

Tabla 34: Historia de usuario HU-10

ID	HU-10	Usuario	Miembro de la Comisión de Titulación
Nombre	Descargar gráficos en formato .png		
Descripción	Permite descargar el gráfico en formato .png		
Validación El usuario puede descargar el gráfico en formato .png.	Prioridad	Media	
	Puntos	2	
	Estimación (Horas)	10	

Tabla 35: Historia de usuario HU-11

ID	HU-11	Usuario	Miembro de la Comisión de Titulación
Nombre	Descargar tablas en formato .xlsx		
Descripción	Permite descargar las tablas de los indicadores en formato .xlsx		
Validación Los miembros de la Comisión pueden descargar las tablas de los indicadores en formato .xlsx.	Prioridad	Media	
	Puntos	2	
	Estimación (Horas)	10	

Tabla 36: Historia de usuario HU-12

ID	HU-12	Usuario	Miembro de la Comisión de Titulación
Nombre	Actualizar vistas materializadas		
Descripción	Actualiza los registros de las vistas materializadas en la base de datos.		
Validación Los miembros de la Comisión pueden actualizar las vistas materializadas cuando encuentren conveniente.	Prioridad	Baja	
	Puntos	2	
	Estimación (Horas)	10	

Tabla 37: Historia de usuario HU-13

ID	HU-13	Usuario	Miembro de la Comisión de Titulación
Nombre	Generar plantillas en formato .xlsx		
Descripción	Permite descargar una plantilla en formato .xlsx con todos los valores de una columna en la base de datos.		
Validación El usuario puede descargar una plantilla de cualquiera de las columnas disponibles en la aplicación web.	Prioridad	Baja	
	Puntos	2	
	Estimación (Horas)	10	

2.4 Product backlog

Para estimar las actividades a realizar en cada uno de los ítems del product backlog se utilizó la técnica de *T-Shirt Sizing*, cuyas equivalencias con los puntos estimados y las horas de trabajo se han establecido acorde a lo presentado en la *Tabla 38*:

Tabla 38: Equivalencias para T-Shirt Sizing

Talla	Iteración	Puntos estimados	Horas de trabajo
XS	1/10 iteración	1	5
S	1/5 iteración	2	10
M	1/2 iteración	5	25
L	1 iteración	10	50
XL	2 iteraciones	20	100

Las estimaciones se realizaron para una sola persona, tomando en cuenta que una jornada laboral representa 5 horas de trabajo y una iteración equivale a dos semanas o 10 días laborales. Con estas consideraciones, se efectuaron los cálculos correspondientes para cada talla.

Posteriormente, se construyó el *product backlog* (Tabla 39), recopilando las estimaciones y priorizando cada historia técnica y de usuario.

Tabla 39: Product backlog

Código	Descripción	Talla	Puntos	Horas	Prioridad
HT-01	Diseñar la arquitectura de la aplicación	S	2	10	Alta
HT-02	Diseñar la base de datos	XS	1	5	Alta
HT-03	Diseñar las vistas materializadas	S	2	10	Alta
HT-04	Cargar los datos	M	5	25	Alta
HT-05	Establecer el estándar de codificación	XS	1	5	Baja
HT-06	Desplegar la aplicación web.	M	5	25	Media
HU-01	Desarrollar analíticas de datos	XL	20	100	Alta
HU-02	Autenticar usuario	S	2	10	Baja
HU-03	Visualizar analíticas de datos	M	5	25	Media
HU-04	Visualizar indicadores clave	M	5	25	Alta
HU-05	Cargar datos externos	L	10	50	Media
HU-06	Crear indicadores personalizados	M	5	25	Media
HU-07	Modificar tipo de gráfico	XS	1	5	Baja
HU-08	Modificar título del gráfico	XS	1	5	Baja
HU-09	Modificar estilo del gráfico	S	2	10	Baja
HU-10	Descargar gráficos en formato .png	S	2	10	Media
HU-11	Descargar tablas en formato .xlsx	S	2	10	Media
HU-12	Actualizar vistas materializadas	S	2	10	Baja
HU-13	Generar plantillas en formato .xlsx	S	2	10	Baja

2.5 Sprint backlog

Las actividades planificadas se ejecutaron 8 *sprints* de dos semanas, a excepción del último, cuyos puntos estimados totales fueron 5, correspondientes a una semana (media iteración). La Tabla 40 muestra el *sprint backlog* para el desarrollo de las actividades descritas en las historias técnicas y de usuario.

Tabla 40: Sprint backlog

N°	Código	Descripción	Fecha Ini.	Fecha Fin	Pts. estim.	Pts. totales
1	HT-01	Diseñar la arquitectura de la aplicación	24/04/23	25/04/23	2	10
	HT-02	Diseñar la base de datos	26/04/23	26/04/23	1	
	HT-03	Diseñar las vistas materializadas	27/04/23	28/04/23	2	
	HT-04	Cargar los datos	01/05/23	05/05/23	5	
2	HU-05	Cargar datos externos	08/05/23	19/05/23	10	10
3	HU-01	Desarrollar analíticas de datos	22/05/23	02/06/23	10	10
4	HU-01	Desarrollar analíticas de datos	05/06/23	16/06/23	10	10
5	HT-05	Establecer el estándar de codificación	19/06/23	19/06/23	1	10
	HU-02	Autenticar usuario	20/06/23	21/06/23	2	
	HU-04	Visualizar indicadores clave	22/06/23	28/06/23	5	
	HU-10	Descargar gráficos en formato .png	29/06/23	30/06/23	2	
6	HU-06	Crear indicadores personalizados	03/07/23	07/07/23	5	10
	HU-08	Modificar título del gráfico	10/07/23	10/07/23	1	
	HU-09	Modificar estilo del gráfico	11/07/23	12/07/23	2	
	HU-11	Descargar tablas en formato .xlsx	13/07/23	14/07/23	2	
7	HU-12	Actualizar vistas materializadas	17/07/23	18/07/23	2	10
	HU-13	Generar plantillas en formato .xlsx	19/07/23	20/07/23	2	

	HU-07	Modificar tipo de gráfico	21/07/23	21/07/23	1	
	HU-03	Visualizar analíticas de datos	24/07/23	28/07/23	5	
8	HT-06	Desplegar la aplicación web.	31/07/23	04/08/23	5	5

3 Desarrollo

La aplicación web para el apoyo a la toma de decisiones “IMAK” se desarrolló mediante la ejecución de tareas técnicas que incluyen la definición de la arquitectura, el establecimiento de estilos de codificación, el diseño de la base de datos y de las interfaces de usuario.

3.1 Arquitectura de la aplicación

Para documentar la arquitectura de la aplicación, se ha optado por emplear el modelo C4, que define cuatro vistas o niveles, cada una centrado en una perspectiva diferente del sistema.

El primer diagrama presentado en la *Ilustración 8*, y que corresponde al contexto de la aplicación, evidencia la interacción de “IMAK” con el sistema de gestión de trabajos de titulación de la carrera.

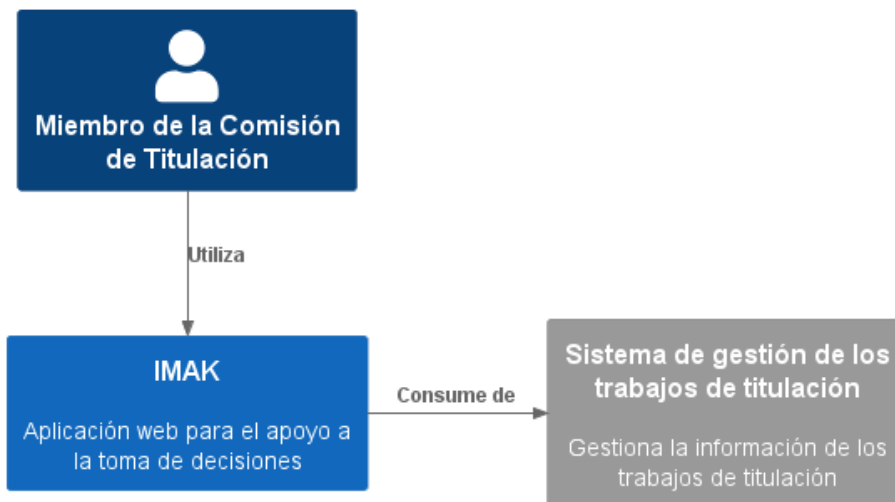


Ilustración 8: Diagrama de contexto de la aplicación

En la segunda vista, representada a través de un diagrama de contenedor (*Ilustración 9*), se observa que la aplicación web “IMAK” obtiene los datos de las vistas almacenadas en el esquema “kpi” de la base de datos del sistema de gestión de trabajos de titulación de la carrera a través de una arquitectura cliente-servidor.

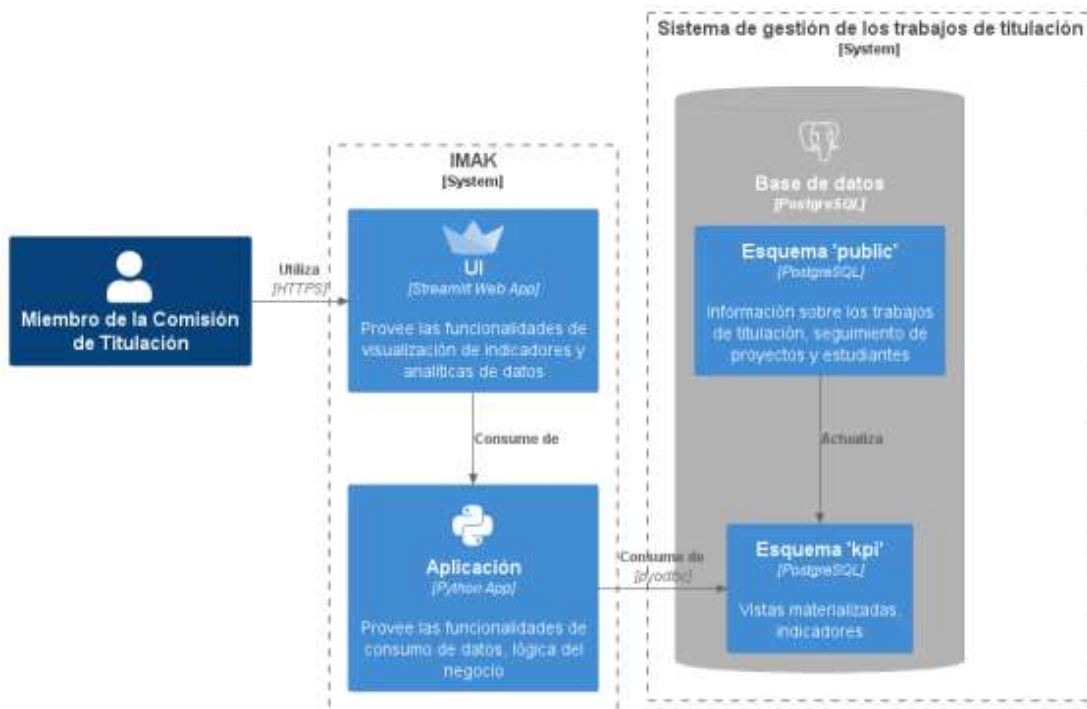


Ilustración 9: Diagrama de contenedor de la aplicación

3.2 Estilo de codificación

Se utilizó la guía de estilo PEP 8, que provee la orientación necesaria para mejorar la lectura del código en Python y su consistencia. Las convenciones completas pueden encontrarse en la guía oficial disponible en el sitio web de PEPS (<https://peps.python.org/pep-0008/>)

3.3 Diseño de la base de datos

La aplicación web “IMAK” utiliza una base de datos propia en el esquema “kpi” para obtener la información referente a los indicadores clave requeridos por la Comisión de Titulación, que son almacenados en vistas materializadas y posteriormente consultadas para presentarlos en forma de gráficos. La *Ilustración 10* corresponde al modelo relacional de la base de datos.

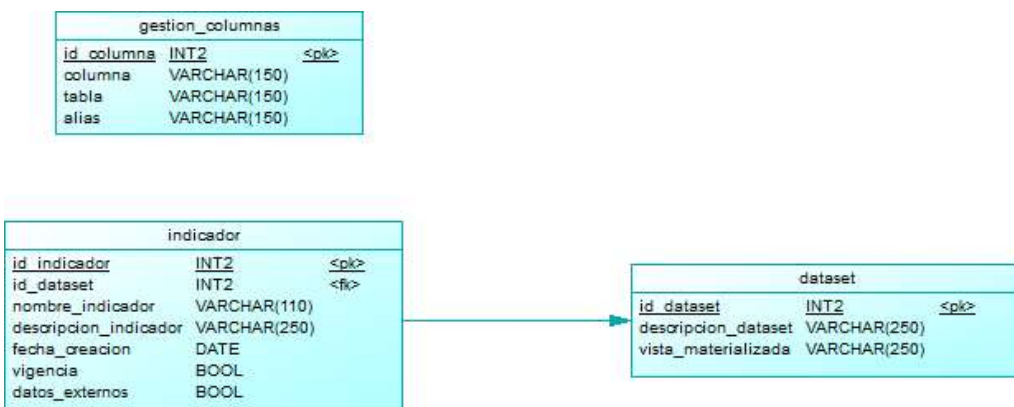


Ilustración 10: Diseño de la base de datos

El diccionario de datos, que documenta y define la información principal de los campos de cada tabla, incluyendo el nombre de la columna, restricciones, tipo, si permite valores nulos, valor predeterminado y descripción. A continuación, se presenta el diccionario de datos para las tablas presentadas en el modelo anterior.

Tabla 41: Diccionario de datos para la tabla "gestion_columnas"

Nombre de la tabla: gestion_columnas				
Descripción: Alias asignados a las columnas de una tabla				
Columna	Descripción	Tipo	Nulos	Valor permitido
id_columna (PK)	Identificador único de la columna	smallint	No	*0 a 32767*
columna	Nombre de la columna	varchar(150)	No	{ [A-Z a-z] }
tabla	Nombre de la tabla	varchar(150)	No	{ [A-Z a-z] }
alias	Alias asignado a la columna para su manejo en la aplicación web	varchar(150)	No	{ [A-Z a-z] }

Tabla 42: Diccionario de datos para la tabla "indicador"

Nombre de la tabla: indicador				
Descripción: Indicador manejado por la Comisión de Titulación				
Columna	Descripción	Tipo	Nulos	Valor permitido
id_indicador (PK)	Identificador único del indicador	smallint	No	*0 a 32767*
id_dataset (FK)	Identificador del <i>dataset</i> que almacena los datos del indicador	smallint	No	*0 a 32767*
nombre_indicador	Nombre corto asignado al indicador para identificarlo	varchar(110)	No	{ [A-Z a-z] }

descripcion_indicador	Nombre largo (descripción) asignado al indicador	varchar(250)	Si	{ [A-Z a-z] }
fecha_creacion	Fecha de creación del indicador en la BD de la CT	date	No	* formato: aaaa – mm – dd *
vigencia	Vigencia del indicador en la CT	boolean	No	[true false]
datos_externos	Indica si el indicador requiere la carga de datos externos	boolean	No	[true false]

Tabla 43: Diccionario de datos para la tabla “dataset”

Nombre de la tabla: dataset				
Descripción: Información de la vista materializada (<i>dataset</i>) almacenada en el esquema “kpi”				
Columna	Descripción	Tipo	Nulos	Valor permitido
id_dataset (PK)	Identificador único del <i>dataset</i>	smallint	No	*0 a 32767*
descripcion_dataset	Descripción del <i>dataset</i>	varchar(250)	No	{ [A-Z a-z] }
vista_materializada	Nombre de la vista materializada, tal y como aparece en el esquema “kpi”	varchar(250)	No	{ [a-z 0-9] }

3.4 Diseño de las interfaces

Se diseñaron *wireframes* de las pantallas a incluir en la aplicación web para visualizar su estructura y los elementos que la constituyen. Los diagramas, presentado a continuación, fueron desarrollados utilizando el software Balsamiq Wireframes.

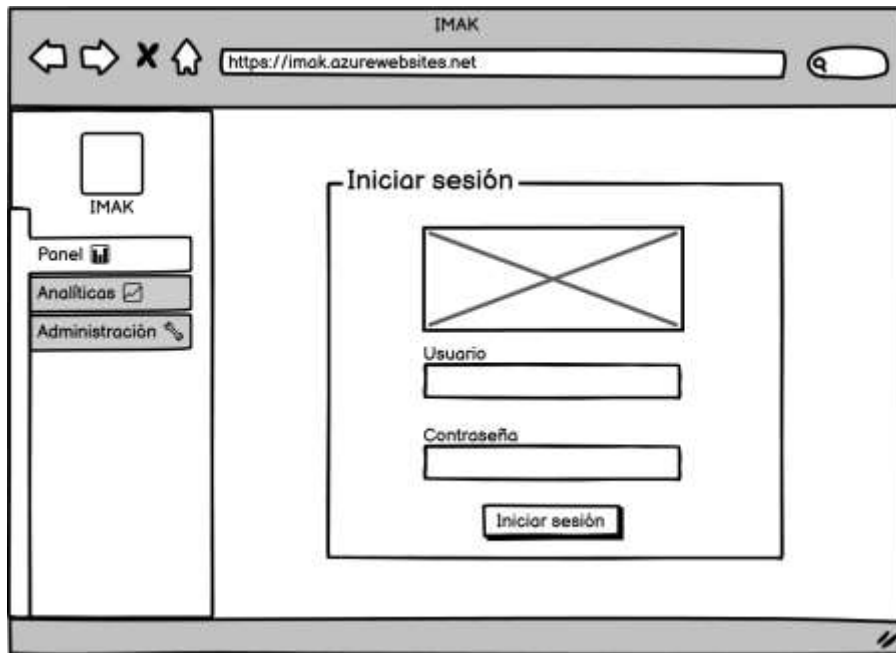


Ilustración 11: Wireframe de la pantalla “Iniciar sesión”

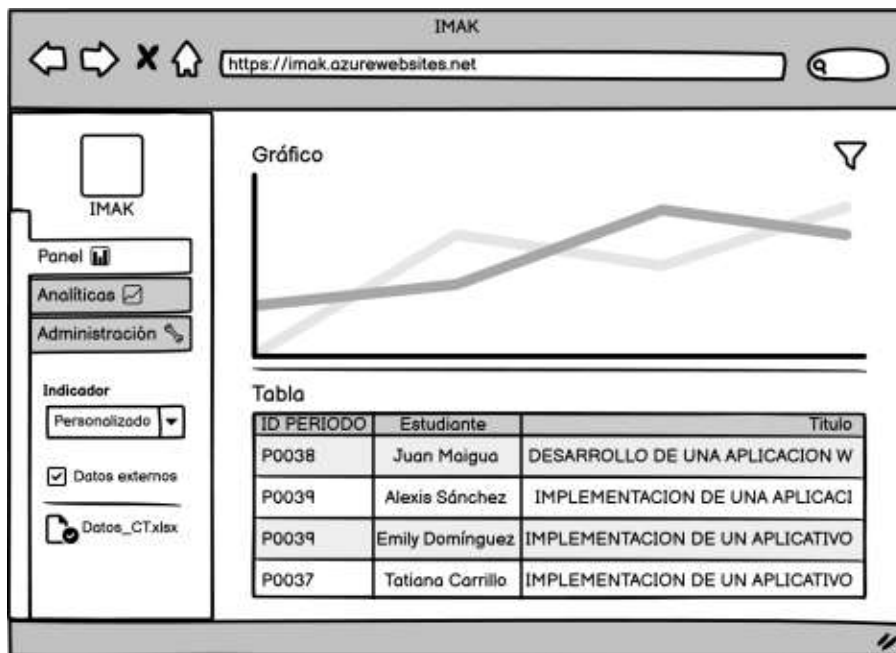


Ilustración 12: Wireframe de la pantalla “Panel”

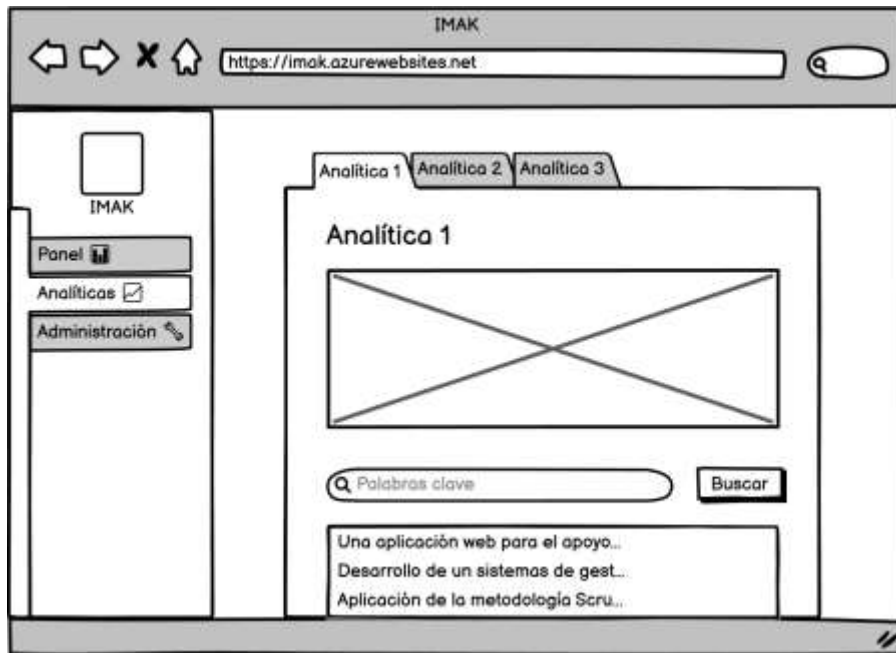


Ilustración 13: Wireframe de la pantalla “Analíticas”

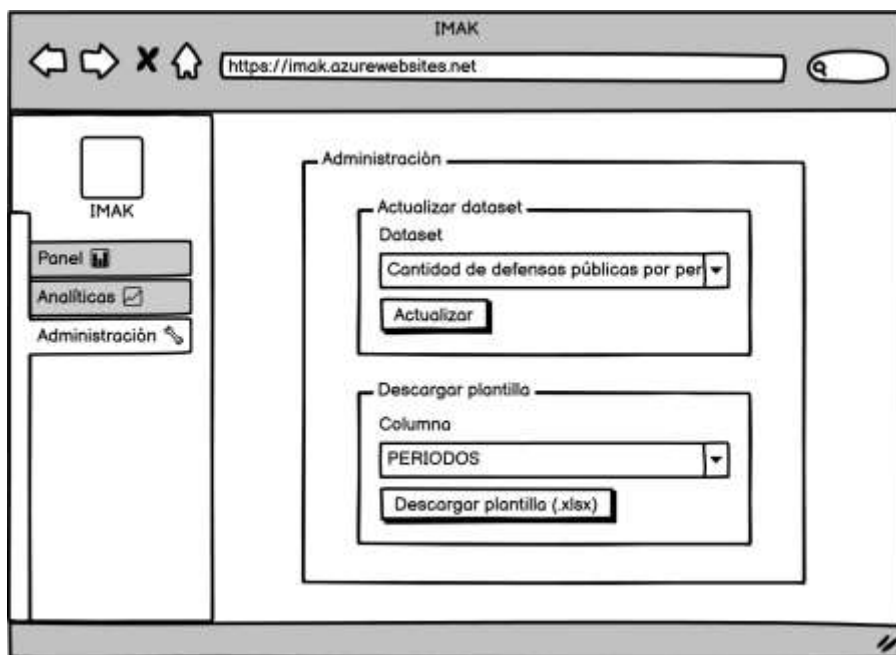


Ilustración 14: Wireframe de la pantalla “Administración”

4 Cierre

Se describen las actividades realizadas que marcaron la finalización de la gestión y el desarrollo del proyecto.

4.1 Burndown chart

Para visualizar la diferencia entre los puntos de historia estimados y los puntos reales completados en cada *sprint*, se empleó el *burndown chart*, presentado en la *Ilustración 15*:

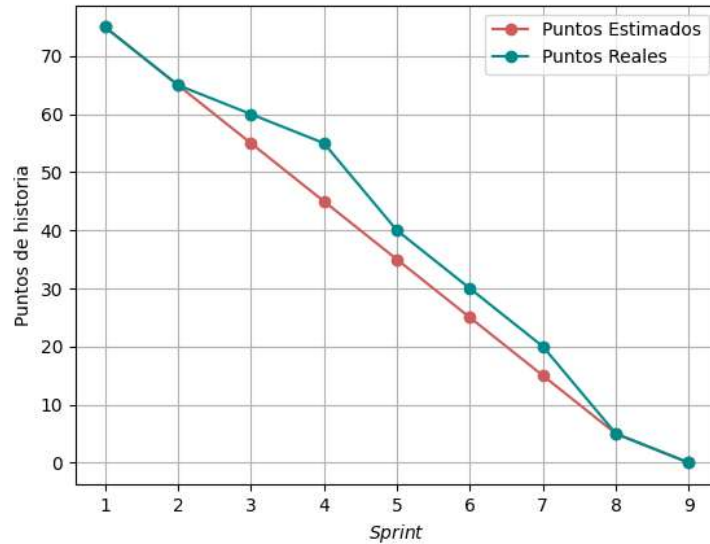


Ilustración 15: Burndown chart

Luego de finalizar el desarrollo, se desplegó la aplicación en un servidor de la carrera de Software. La *Ilustración 16* muestra el diagrama de despliegue de la aplicación.

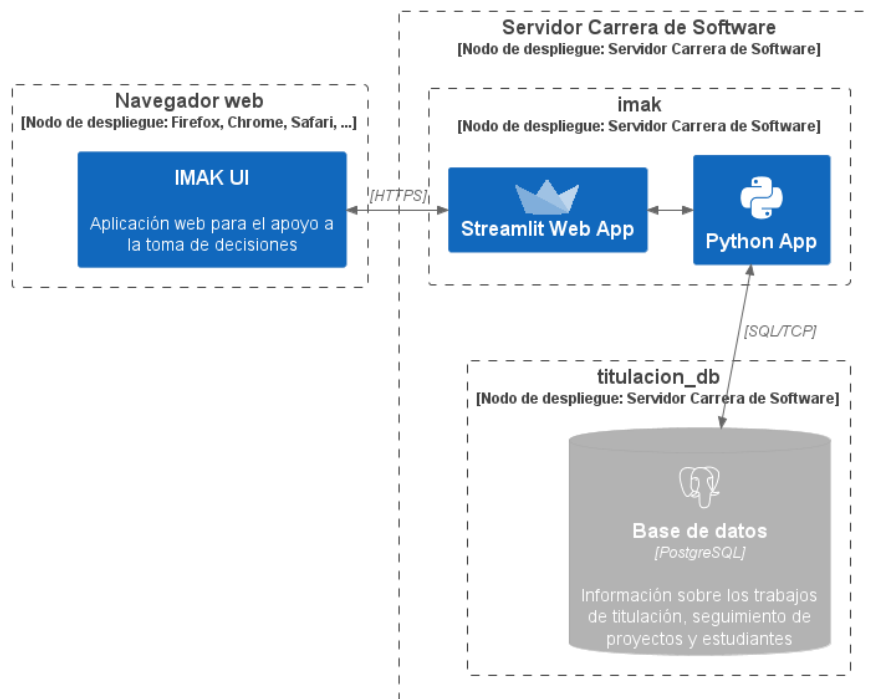


Ilustración 16: Diagrama de despliegue

ANEXO C: MANUAL TÉCNICO – CRISP-DM

MANUAL TÉCNICO

CRISP-DM

IMAK - DSS

Junio, 2023

Introducción

El modelo referencial CRISP-DM se divide en seis fases: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue. Se trata de un proceso flexible, en el que se puede mover entre cada fase dependiendo de los resultados obtenidos al finalizar cualquiera de las etapas. La presente guía tiene como objetivo documentar las fases del proceso del modelo estándar CRISP-DM empleado en la generación de analíticas de datos para la aplicación web de apoyo a la toma de decisiones “IMAK”.

1. Comprensión del negocio

En esta primera etapa, las tareas realizadas incluyeron la determinación de los objetivos del negocio, evaluación de la situación, determinación de los objetivos de minería y la elaboración de un plan del proyecto.

1.1 Determinación de los objetivos del negocio

Antecedentes

La Comisión de Unidad de Integración Curricular (Comisión de Titulación) es un comité conformado por el Coordinador de la Carrera, un profesor de la Unidad Profesional y un profesor de la Unidad de Integración Curricular, quienes ejercen sus funciones por un mínimo de dos periodos académicos ordinarios. Sus atribuciones incluyen la planificación y estructuración de actividades, además de la evaluación de las solicitudes de los estudiantes para acceder a las diferentes modalidades de titulación. La Tabla 44 muestra los miembros actuales y su función dentro de la Comisión de Titulación.

Tabla 44: Estructura de la Comisión de Titulación

Miembro	Función
Patricio Moreno	Presidente CT
Danilo Pástor	Miembro CT
Ivonne Rodríguez	Miembro CT

La Comisión de Titulación de la carrera de Software ha almacenado datos referentes al proceso de titulación de los estudiantes. Sin embargo, al no contar con la información suficiente para facilitar la toma de decisiones que conduzcan a solucionar problemas dentro de este contexto, buscan extraer valor de los datos recolectados de manera automatizada. La solución temporal a este problema ha sido generar gráficos de los indicadores principales del proceso de titulación

utilizando Power BI. Esta información le ha permitido a la Comisión construir los informes necesarios, aprovechando las ventajas de dicha herramienta de reportería. Aun así, las limitaciones del programa no permitirían satisfacer el requerimiento de disponer de analíticas de datos más sofisticadas que generen valor para la Comisión. También, se ha propuesto el desarrollo de un sistema que almacene de forma organizada la información referente al seguimiento de los estudiantes y sus trabajos. El principal beneficio de la implementación de esta solución es que permitirá contar con una estructura fija de los datos, lo que facilitará su posterior análisis. Pero el contexto de este sistema se limita a la manipulación de los, mas no a la extracción de valor de la información.

Objetivos del negocio

La Comisión de Titulación, en base a los datos que almacenan, buscan:

- Disponer de información útil que les permita tomar mejores decisiones respecto a los trabajos de titulación.
- Darles un valor agregado a los indicadores claves que manejan, al incluir un análisis más sofisticado de los datos.
- Comunicar de manera oportuna y precisa la información que se modele, de forma tal que individuos externos a la Comisión puedan comprender su valor.
- Migrar los datos almacenados de forma provisional en SQL Server a la base de datos definitiva en PostgreSQL del sistema de gestión de los trabajos de titulación de la carrera.

Criterios de éxito del negocio

Los criterios de éxito para el presente consisten en que la Comisión de Titulación considere:

- Que la información generada (modelos de analítica) les permitirá tomar mejores decisiones en las labores que ejecutan.
- Que la información generada aporta a los informes que realizan, siendo un valor añadido a los indicadores claves con los que trabajan.
- Que la forma de presentar la información sea clara y comprensible para cualquier individuo en el contexto de su trabajo.

1.2 Evaluación de la situación

Los entregables para este apartado han sido documentados en la gestión del proyecto con Scrum. Para más información, consulte el **MANUAL TÉCNICO – SCRUM – IMAK DSS**.

1.3 Determinación de los objetivos de la minería de datos

Objetivos de la minería de datos

En base a los requerimientos del negocio, se plantean los siguientes objetivos de minería:

- **DM1:** Utilizar los títulos de los trabajos de titulación para extraer las palabras clave en un rango de periodos académicos (procesamiento del lenguaje natural para etiquetado gramatical).
- **DM2:** Identificar los temas y términos recurrentes de los trabajos de titulación (procesamiento del lenguaje natural para modelado de temas).
- **DM3:** Visualizar los proyectos no finalizados y el tiempo en meses desde su aprobación (estadística descriptiva).

1.4 Plan del proyecto

De acuerdo con la planificación descrita en el **MANUAL TÉCNICO – SCRUM – IMAK DSS**, los modelos de analítica se desarrollarán en un lapso de media iteración, es decir, una semana, en conformidad con la *Tabla 45* mostrada a continuación.

Tabla 45: Vista general del plan del proyecto

Fases CRISP-DM	Tiempo	Recursos	Riesgos
Comprensión del negocio	2 días	Analista, Miembros de la Comisión	Cambio en los requerimientos.
Comprensión de los datos	3 días	Analista	Cambio en los requerimientos, desconocimiento de las herramientas.
Preparación de los datos	5 días	Analista	Desconocimiento de las herramientas.
Modelado	5 días	Analista	Desconocimiento de las herramientas.

Fases CRISP-DM	Tiempo	Recursos	Riesgos
Evaluación	2 día	Analista, Miembros de la Comisión	Desconocimiento de las herramientas.
Despliegue	3 días	Analista/Desarrollador	Desconocimiento de las herramientas.

2. Comprensión de los datos

La fase de comprensión de los datos agrupó las tareas de recolección, descripción, exploración y verificación de la calidad de los datos iniciales.

2.1 Recolección de los datos iniciales

La fuente para el proyecto corresponde a una base de datos provisional en SQL Server, “DB_SeguimientoUIC”, que ha sido construida durante varios periodos al recopilar información referente al proceso de titulación, específicamente el seguimiento histórico de los proyectos y estudiantes. Dicha base de datos cuenta con 12 tablas, de las cuales 5 son de interés para el análisis:

- **Tabla de proyectos aprobados (dsa.proy_aprobado):** almacena los datos de los proyectos aprobados por la Comisión de Titulación.
- **Tabla de docentes por proyecto (dsa.proy_docente):** Almacena la información del docente por proyecto según su función.
- **Tabla de estudiantes por proyecto (dsa.proy_estud):** Almacena la información del estudiante en relación con su proyecto.
- **Tabla de catálogo de procesos (dsa.catProcesos):** en esta tabla se mantiene un catálogo de los procesos por los que atraviesa un proyecto durante su ejecución.
- **Tabla de seguimiento de proyectos (dsa.segProy):** registra el seguimiento de los proyectos a través de los procesos establecidos por la Comisión.

2.2 Descripción, exploración y verificación de la calidad de los datos

A continuación, se presenta el perfilado de datos de las columnas de interés en cada una de las tablas señaladas. La información se obtuvo utilizando el módulo de Python “*pandas_profiling*”, que brinda un reporte completo de las características principales de los datos.

dsa.proy_aprobado

Como se observa en la *Ilustración 17*, la tabla cuenta con un total de 7 columnas (variables), distribuidas en 4 columnas numéricas, 2 categóricas y 1 de tipo *unsupported*, lo que implica que todos sus valores son nulos. Además, cuenta con 210 observaciones (número de filas).

Dataset statistics		Variable types	
Number of variables	7	Numeric	4
Number of observations	210	Categorical	2
Missing cells	212	Unsupported	1
Missing cells (%)	14.4%		
Total size in memory	11.6 KiB		
Average record size in memory	56.6 B		

Ilustración 17: Descripción general de la tabla dsa.proy_aprobado

La *Ilustración 18* confirma la presencia de valores nulos en la variable MODALIDAD.

Alerts		
PROYECTO	has a high cardinality: 210 distinct values	High cardinality
MODALIDAD	has 210 (100.0%) missing values	Missing
COD_PROYECTO_UIC	has unique values	Unique
PROYECTO	has unique values	Unique
MODALIDAD	is an unsupported type, check if it needs cleaning or further analysis	Unsupported
ID_PROCESO	has 8 (3.8%) zeros	Zeros

Ilustración 18: Características generales de las columnas de la tabla dsa.proy_aprobado

Sin embargo, son de interés únicamente las columnas PROYECTO, COD_PERIODO y ID_PROCESO.

La columna **PROYECTO** representa el nombre del trabajo de titulación. El 100% de los valores son distintos, lo que implica que no hay valores repetidos. Tampoco existen registros faltantes. La *Ilustración 19* muestra más características de esta columna.

PROYECTO	
Categorical	
HIGH CARDINALITY UNIQUE	
Distinct	210
Distinct (%)	100.0%
Missing	0
Missing (%)	0.0%
Memory size	1.8 KiB

Ilustración 19: Características de la columna PROYECTO de la tabla dsa.proy_aprobado

La columna **COD_PERIODO** almacena el código del periodo de aprobación del proyecto. Es de tipo categórica y cuenta con únicamente 8 valores distintos, que representarían un total de 8 periodos en los que se han registrado proyectos aprobados. La *Ilustración 20* muestra las características de la columna.



COD_PERIODO	
Categorical	
Distinct	8
Distinct (%)	3.8%
Missing	0
Missing (%)	0.0%
Memory size	1.8 KIB

Ilustración 20: Características de la columna COD_PERIODO en dsa.proy_aprobado

Como se presenta en la *Ilustración 21*, los proyectos aprobados se han registrado desde el periodo P0031 hasta el P0038, siendo el periodo P0037 el que mayor número de proyectos aprobados ha tenido.



Value	Count	Frequency (%)
P0034	37	17.6%
P0038	30	14.3%
P0033	29	13.8%
P0035	29	13.8%
P0032	27	12.9%
P0037	24	11.4%
P0036	18	8.6%
P0031	16	7.6%

Ilustración 21: Frecuencia de registro de proyectos aprobados por periodo académico

Finalmente, la columna **ID_PROCESO** representa el proceso registrado en la tabla. Si observamos la *Ilustración 22* notaremos que existen 3 valores distintos

ID_PROCESO

Real number (R)

Distinct	3	Minimum	0
Distinct (%)	1.4%	Maximum	13
Missing	0	Zeros	8
Missing (%)	0.0%	Zeros (%)	3.8%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	4.885714286	Memory size	1.8 KiB

Ilustración 22: Características de la columna ID_PROCESO en dsa.proy_aprobado

La Ilustración 23 indica que estos valores son 5, 13 y 0. Los dos primeros corresponden a los procesos “Aprobación del proyecto” y “Cambio en el tribunal”, respectivamente, mientras que al valor 0 no se le relaciona con ningún proceso. Se deben considerar los registros cuyo ID_PROCESO sea distinto a 5, ya que no corresponden a procesos válidos dentro de la tabla.

Value	Count	Frequency (%)
5	200	95.2%
0	8	3.8%
13	2	1.0%

Ilustración 23: Frecuencia de los valores distintos para ID_PROCESO en la tabla dsa.proy_aprobado

dsa.proy_docente

Esta tabla relaciona la información de los docentes con los proyectos asignados según su función. Como se muestra en la Ilustración 24, cuenta con 8 columnas y un total de 562 observaciones.

Dataset statistics		Variable types	
Number of variables	8	Numeric	3
Number of observations	562	Categorical	4
Missing cells	672	Unsupported	1
Missing cells (%)	14.9%		
Total size in memory	35.2 KiB		
Average record size in memory	64.2 B		

Ilustración 24: Descripción general de la tabla dsa.proy_docente

El principal inconveniente detectado es la falta de datos en la columna CEDULA, ya que ningún docente tiene registrado dicho valor, tal y como se presenta en la *Ilustración 25*.

Alert	Action
OBSERVACION is highly imbalanced (88.9%)	Imbalance
CEDULA has 562 (100.0%) missing values	Missing
OBSERVACION has 110 (19.8%) missing values	Missing
CEDULA is an unsupported type, check if it needs cleaning or further analysis	Unsupported
VIGENCIA has 18 (3.2%) zeros	Zeros

Ilustración 25: Características generales de las columnas de la tabla dsa.proy_docente

DOCENTE y FUNCION son las columnas de interés dentro de esta tabla y que se analizarán detalladamente.

En la columna **DOCENTE** se presenta el nombre de profesor asignado al proyecto. Se han registrado un total de 49 docentes y no existen valores faltante, según lo presentado en la *Ilustración 26*.

DOCENTE	
Categorical	
Distinct	49
Distinct (%)	8.7%
Missing	0
Missing (%)	0.0%
Memory size	4.5 KiB

Ilustración 26: Características de la columna DOCENTE de la tabla dsa.proy_docente

A cada docente se le asigna un rol dentro del proyecto, que es almacenada en la columna **FUNCION**. Según la *Ilustración 27*, existen 4 valores distintos y ningún registro presenta datos faltantes en este campo.

FUNCION

Categorical

Distinct	4
Distinct (%)	0.7%
Missing	0
Missing (%)	0.0%
Memory size	4.5 KiB

Ilustración 27: Características de la columna FUNCION de la tabla dsa.proy_docente

Estos 4 valores corresponden a las funciones que puede desempeñar el docente dentro de un proyecto. La *Ilustración 28* muestra la frecuencia de dichos datos en la tabla.

Value	Count	Frequency (%)
DIR	229	40.7%
MIE	162	28.8%
DEL	104	18.5%
ASE	67	11.9%

Ilustración 28: Frecuencia de las funciones dentro de la tabla dsa.proy_docente

dsa.proy_estud

La tabla dsa.proy_estud almacena la información de los estudiantes y sus proyectos. Cuenta con un total de 12 columnas y 285 observaciones, como se presenta en la *Ilustración 29*.

Dataset statistics		Variable types	
Number of variables	12	Numeric	7
Number of observations	285	Categorical	5
Missing cells	854		
Missing cells (%)	25.0%		
Total size in memory	26.8 KiB		
Average record size in memory	96.4 B		

Ilustración 29: Descripción general de la tabla dsa.proy_estud

Existe una gran cantidad de valores faltantes en varias de las columnas de la tabla, tal y como se muestra en la *Ilustración 30*. Sin embargo, solamente son de interés las columnas NOMBRE y APELLIDOS.

Alerts

APELLIDOS has a high cardinality: 272 distinct values	High cardinality
NOMBRE has a high cardinality: 267 distinct values	High cardinality
SEDE is highly imbalanced (72.2%)	Imbalance
OBSERVACION has 274 (96.1%) missing values	Missing
ESTADO has 10 (3.5%) missing values	Missing
COD_TESIS_OASIS has 261 (91.6%) missing values	Missing
SEBE has 77 (27.0%) missing values	Missing
COD_EST_ANT has 232 (81.4%) missing values	Missing
ID_PROCESO has 9 (3.2%) zeros	Zeros
VIGENCIA has 6 (2.1%) zeros	Zeros

Ilustración 30: Características generales de las columnas de la tabla dsa.proy_estud

Las columnas **NOMBRE** y **APELLIDOS** guardan los nombres completos de los estudiantes. En ambos casos, no existen valores faltantes (*Ilustración 31* e *Ilustración 32*), sin embargo, existen nombres registrados en letras minúsculas y mayúsculas, por lo que es necesario normalizar dichos datos.

NOMBRE
Categorical

Distinct	267
Distinct (%)	93.7%
Missing	0
Missing (%)	0.0%
Memory size	2.4 KiB

Ilustración 31: Características de la columna NOMBRE de la tabla dsa.proy_estud

APELLIDOS
Categorical

Distinct	272
Distinct (%)	95.4%
Missing	0
Missing (%)	0.0%
Memory size	2.4 KiB

Ilustración 32: Características de la columna APELLIDOS de la tabla dsa.proy_estud

dsa.catProcesos

Al ser una tabla con pocos registros, se realizó el análisis ejecutando una consulta SQL. Observamos en la *Ilustración 33* que la tabla cuenta con 5 columnas. Son de interés únicamente ID_PROCESO, ACTIVIDAD y DESCRIPCIÓN.

ID_PROCESO	ACTIVIDAD	DESCRIPCIÓN	id_padre	orden
1	PROCESO TITULACION	NUL	0	0
2	INGRESO A LA UIC	INGRESO DE ESTUDIANTES A LA UIC - VIA SEMINARIO / B...	5	1
3	APROBACION DE OPCION TITULACION	APROBACION DE OPCION DE TITULACION - TIC / EXAMEN CB...	5	2
4	DEFENSA ANTEPROYECTO	PARA ANTEPROYECTO	3	3
5	APROBACION DEL PROYECTO	PROYECTOS APROBADOS POR COMISION UIC	8	4
6	DESIGNACION DE DOCENTES PARA TRIBUNAL DE GRADO	COMISION UIC SUGIERE DESIGNACION DE DOCENTES COMO TR...	5	5
7	EXAMEN COMPLETIVO	PARA EXAMEN COMPLETIVO	3	00
8	GRADO	SUBPROCESO DE GRADO - DEFENSA	5	8
9	PRODEFENSA	PRODEFENSA DE PROYECTO	8	7
10	DEFENSA PUBLICA	DEFENSA ORAL PUBLICA DEL PROYECTO	8	8
11	MATRICULA	MATRICULA DEL PROYECTO	5	8
12	AUTORIZACION DE MATRICULA	UIC AUTORIZA MATRICULA DEL PROYECTO IC	11	6
13	CAMBIO EN EL TRIBUNAL	CAMBIO DE DIRECTOR, MIEMBRO O DELEGADO DE PROYECTO	6	8
14	CAMBIO EN PROYECTO	CAMBIO EN EL PROYECTO POR CAUSAS DE FUERZA MAYOR, SD...	5	8
15	MATRICULA SISTEMA DE TITULACION - ORDINARIO	MATRICULA EN LA MATERIA DE SISTEMAS DE TITULACION DA...	2	8
16	CAMBIO POR SALIDA DE ESTUDIANTE INTEGRANTE	CAMBIO DE ESTUDIANTES EN EL PROYECTO PORI ABANDONA P...	5	0
17	CAMBIO POR INGRESO DE NUEVO ESTUDIANTE	CAMBIO DE ESTUDIANTES EN EL PROYECTO POR: INGRESO DE...	5	0
18	MATRICULA SISTEMA DE TITULACION -REINGRESO A ...	REINGRESO A LA CARRERA - UIC POR NO TERMINAR TRABAJO...	2	8
19	REINGRESO MODULO GRADUACION	REINGRESO AL MODULO DE GRADUACION - CONTINUAR COM TR...	8	0

Ilustración 33: Datos de la tabla dsa.catProcesos

Los procesos de interés para el análisis son: APROBACIÓN DEL PROYECTO, que inicia la ejecución de proyecto; y DEFENSA PÚBLICA, que marca el fin de un trabajo de titulación y puede considerarse como terminado. ID_PROCESO = 13 representa cambios en el tribunal, lo cual no tiene sentido en la tabla dsa.proy_aprobado, tal como se indicó con anterioridad. Tampoco existe un proceso cuyo ID sea 0, por lo que dichos datos deberán ser tratados previo al análisis.

dsa.segProy

Esta tabla cuenta con 7 columnas o variables, de las cuales 2 son numéricas, 3 categóricas y 2 de tipo fecha. El número de observaciones o filas es 962. En la *Ilustración 34* podemos ver las principales características descritas.

Dataset statistics		Variable types	
Number of variables	7	Numeric	2
Number of observations	962	Categorical	3
Missing cells	1401	DateTime	2
Missing cells (%)	20.8%		
Total size in memory	52.7 KIB		
Average record size in memory	56.1 B		

Ilustración 34: Descripción general de la tabla dsa.segProy

Al observar las alertas (*Ilustración 35*) notamos que tres columnas tienen valores nulos, especialmente FECHA_ACTIVIDAD y OBSERVACIONES.

Alerts	
SEDE is highly imbalanced (75.2%)	Imbalance
FECHA_ACTIVIDAD has 479 (49.8%) missing values	Missing
SEDE has 18 (1.9%) missing values	Missing
OBSERVACIONES has 904 (94.0%) missing values	Missing

Ilustración 35: Características generales de las columnas de la tabla dsa.segProy

Las columnas de interés en esta tabla son ID_PROCESO, COD_PROYECTO_UIC y FECHA_UIC, que se analizarán a continuación.

ID_PROCESO identifica el proceso al cual hace referencia el registro en la tabla. Como muestra la *Ilustración 36*, existen 10 valores distintos, lo que indica que no existe seguimiento de todos los procesos establecidos en el catálogo. Tampoco existen valores faltantes.

ID_PROCESO	
Real number (R)	
Distinct	10
Distinct (%)	1.0%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Mean	5.937629938
Minimum	3
Maximum	17
Zeros	0
Zeros (%)	0.0%
Negative	0
Negative (%)	0.0%
Memory size	7.6 KiB

Ilustración 36: Características de la columna ID_PROCESO de la tabla dsa.segProy

Observando la *Ilustración 37* podemos concluir que la mayoría de los registros en la tabla corresponden a los procesos 3, 5 y 6, que representan la aprobación de la opción de titulación, la aprobación del proyecto y la designación de docentes para el tribunal de grado, respectivamente.

Value	Count	Frequency (%)
3	203	21.1%
5	199	20.7%
6	199	20.7%
4	131	13.6%
10	94	9.8%
9	84	8.7%
13	43	4.5%
14	5	0.5%
16	3	0.3%
17	1	0.1%

Ilustración 37: Frecuencia de los procesos en la tabla dsa.segProy

La columna **COD_PROYECTO_UIC** identifica de manera única a cada proyecto. Al analizar sus principales características, presentadas en la *Ilustración 38*, observamos que no existen registros faltantes y que hay 212 valores distintos, es decir, se ha realizado el seguimiento a 212 proyectos.

COD_PROYECTO_UIC

Real number (R)

Distinct	212	Minimum	1
Distinct (%)	22.0%	Maximum	9012
Missing	0	Zeros	0
Missing (%)	0.0%	Zeros (%)	0.0%
Infinite	0	Negative	0
Infinite (%)	0.0%	Negative (%)	0.0%
Mean	218.1777547	Memory size	7.6 KIB

Ilustración 38: Características de la columna COD_PROYECTO_UIC de la tabla dsa.segProy

Finalmente, la columna **FECHA_UIC** corresponde a la fecha en la que la Comisión ejecutó el trámite correspondiente al proceso registrado. En la *Ilustración 39* observamos que hay 0 valores faltantes y 89 registros distintos. También podemos observar que la fecha mínima es el 2019-02-

17 y la máxima el 2022-08-1, lo que indica que se han registrado seguimientos desde inicios del 2019 hasta finales del 2022.

FECHA_UIC		Date	
Distinct	89	Minimum	2019-02-17 00:00:00
Distinct (%)	9.3%	Maximum	2022-08-01 00:00:00
Missing	0		
Missing (%)	0.0%		
Memory size	7.6 KiB		

Ilustración 39: Características de la columna FECHA_UIC de la tabla dsa.segProy

Tras el análisis de los resultados obtenidos en los reportes, se estableció las características de las columnas de interés para el desarrollo de las analíticas, que se describen en la *Tabla 46*.

Tabla 46: Descripción de las columnas de interés para el análisis

Tabla	Columna	Descripción	Inconvenientes
dsa.proy_aprobado	COD_PERIODO	Código del periodo académico	-
	PROYECTO	Título del proyecto	Caracteres especiales (~, ^) no deseados, no hay normalización entre mayúsculas y minúsculas
	COD_PROYECTO_UIC	Código asignado al proyecto en la base de datos de la CT	-
	VIGENCIA	Vigencia del proyecto	-
dsa.proy_docente	COD_PROYECTO_UIC	Código asignado al proyecto en la base de datos de la CT	-
	DOCENTE	Nombre del docente	Nombres mal escritos y repetidos
	FUNCION	Función que desempeña el docente en el proyecto	-
	VIGENCIA	Vigencia del docente en el proyecto	-
dsa.proy_estud	COD_PROYECTO_UIC	Código asignado al proyecto en la base de datos de la CT	-

Tabla	Columna	Descripción	Inconvenientes
	APELLIDOS	Apellidos del estudiante	Caracteres especiales (~, ^) no deseados, no hay normalización entre mayúsculas y minúsculas
	NOMBRE	Nombre del estudiante	Caracteres especiales (~, ^) no deseados, no hay normalización entre mayúsculas y minúsculas
dsa.catProcesos	ID_PROCESO	ID del proceso	-
	ACTIVIDAD	Nombre del proceso/actividad	-
dsa.segProy	ID_PROCESO	ID del proceso	-
	COD_PROYECTO_UIC	Código asignado al proyecto en la base de datos de la CT	-
	FECHA_UIC	Fecha de registro del seguimiento en la base de datos de la CT	-

Para el desarrollo de la analítica DM2, se utilizaron 13 trabajos de titulación descargados del repositorio DSPACE ESPOCH, que corresponden a proyectos elaborados entre los años 2019 y 2022. Además, se adicionaron 10 proyectos brindados por la CT y que fueron aprobados en el periodo septiembre 2022 - marzo 2023. De esta manera, se contó con un total de 23 proyectos de titulación para el análisis.

3. Preparación de los datos

Ya que los requerimientos del proyecto requerían migrar los datos de una base de datos SQL Server a la base de datos PostgreSQL del sistema de gestión de trabajos de titulación, se realizaron procesos ETL para ejecutar las actividades de esta etapa (selección, limpieza, construcción, integración y formateo de datos). Este tipo de procesos empiezan con la extracción de los datos desde una fuente u origen. Posteriormente, son transformados para adaptarlos al destino; esta transformación incluye la selección, limpieza y formateo de los datos, para finalmente ser cargados. A continuación, se presentan los procesos ETL diseñados para cada tabla de la base de datos en PostgreSQL, utilizando la herramienta Knime Analytics Platform en su versión 4.7.3. El diseño de los procesos ETL se realizó considerando los inconvenientes detectados en la etapa anterior.

Tabla estudiante_carrera

Nodos empleados:

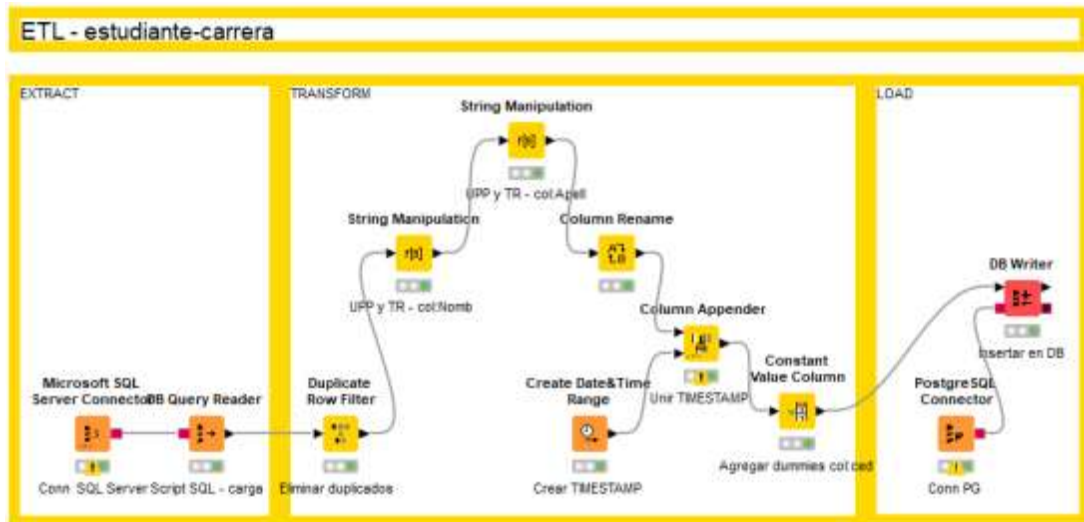


Ilustración 40: Proceso ETL para estudiante_carrera

Consulta SQL:

```
SELECT COD_EST, NOMBRE, APELLIDOS  
FROM dsa.antProy_estud
```

UNION

```
SELECT COD_EST, NOMBRE, APELLIDO  
FROM dsa.ingreso_ic
```

UNION

```
SELECT COD_EST, NOMBRE, APELLIDOS  
FROM dsa.proy_estud
```

Tabla catalogo_proceso

Nodos empleados:

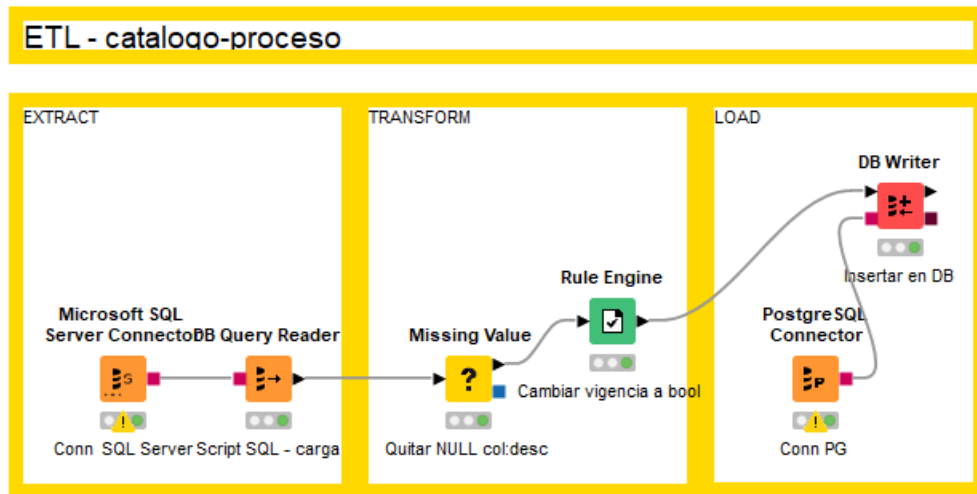


Ilustración 41: Proceso ETL para catalogo_proceso

Consulta SQL:

SELECT

```
ID_PROCESO AS id_proceso,  
ACTIVIDAD AS actividad,  
DESCRIPCION AS descripcion_proceso,  
orden,  
(SELECT 1) AS vigencia,  
id_padre
```

FROM dsa.catProcesos

Tabla periodo_academico

Nodos empleados:

ETL - periodo-academico

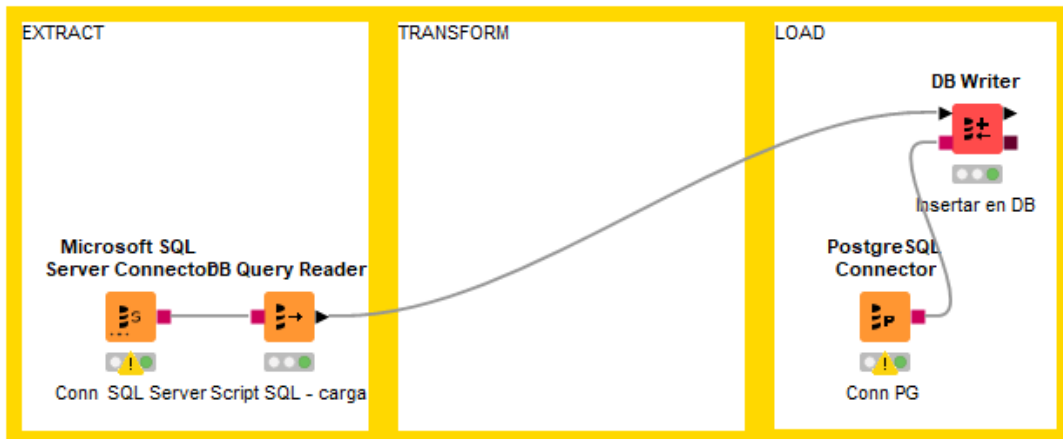


Ilustración 42: Proceso ETL para periodo_academico

Consulta SQL:

SELECT

```
COD_PERIODO AS id_periodo,  
ANIO AS anio,  
DESCRIPCION AS descripcion_periodo,  
FECHA_FIN AS fecha_fin,  
FECHA_INICIO AS fecha_inicio
```

FROM tit.dim_periodo

Tabla anteproyecto

Nodos empleados:

ETL - anteproyecto

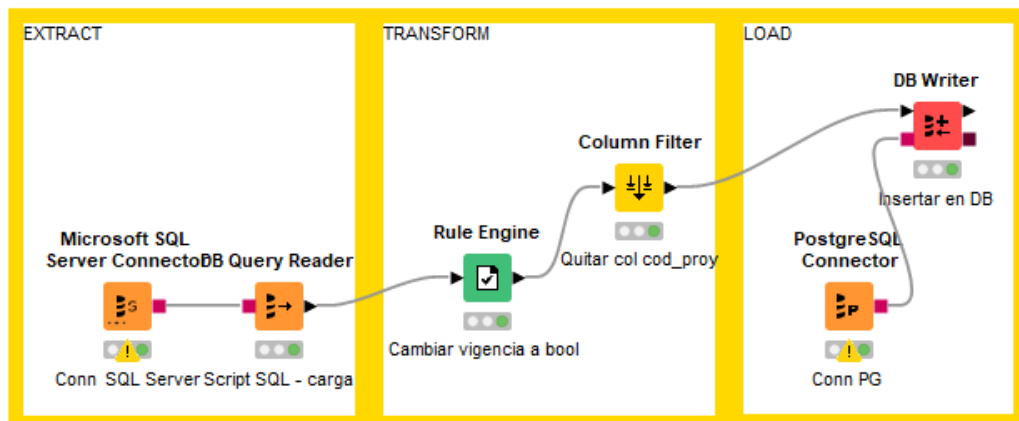


Ilustración 43: Proceso ETL para anteproyecto

Consulta SQL:

SELECT

```
DISTINCT UPPER(ANTEPROYECTO) AS titulo_anteproyecto,  
ESTADO AS id_estado,  
1 AS id_modalidad,  
0 AS id_solicitud,  
COD_PROYECTO_UIC
```

FROM dsa.antProy_estud

ORDER BY COD_PROYECTO_UIC

Tabla proyecto

Nodos empleados:

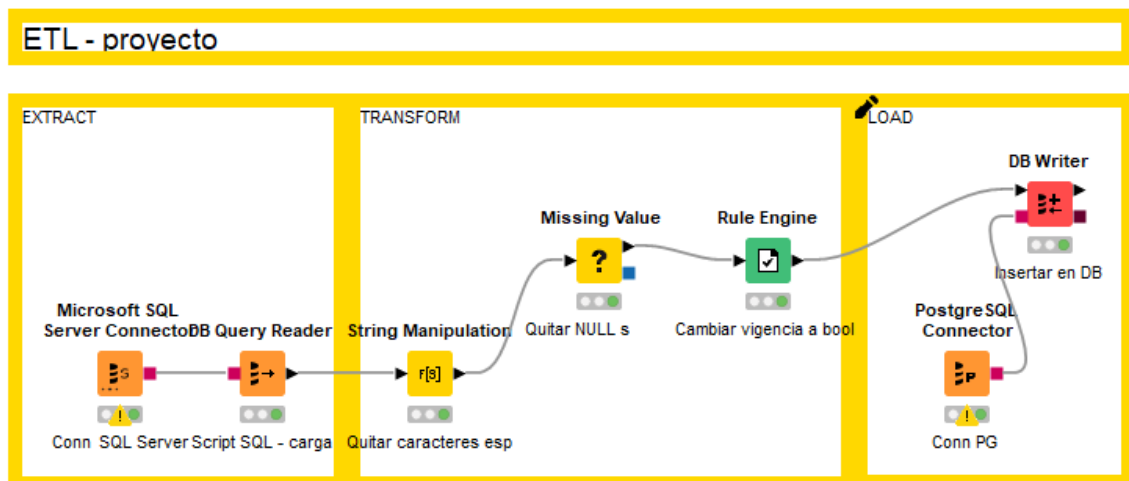


Ilustración 44: Proceso ETL para proyecto

Consulta SQL:

SELECT

```
COD_PROYECTO_UIC AS id_proyecto,  
CASE WHEN COD_PROYECTO_UIC BETWEEN 1 AND 131 THEN  
COD_PROYECTO_UIC ELSE 0 END AS id_anteproyecto,  
COD_PERIODO AS id_periodo,  
PROYECTO AS titulo_proyecto,  
(SELECT GETDATE()) AS date_created,  
(SELECT GETDATE()) AS date_updated,  
VIGENCIA AS proyecto_vigencia,  
ESTADO AS id_estado
```

FROM dsa.proy_aprobado

Tabla estado_proyecto

Nodos empleados:

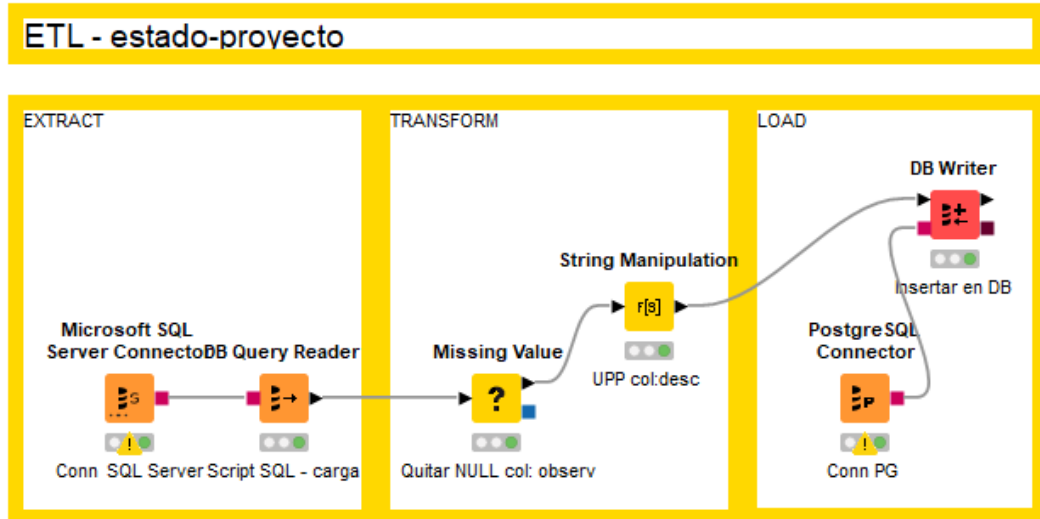


Ilustración 45: Proceso ETL para estado_proyecto

Consulta SQL:

SELECT

ESTADO AS id,
DESCRIPCION AS descripcion,
OBSERVACIONES AS observaciones

FROM dsa.estado

WHERE TABLA LIKE 'proy_aprobado'

Tabla estudiante_proyeto

Nodos empleados:

ETL - estudiante-proyecto

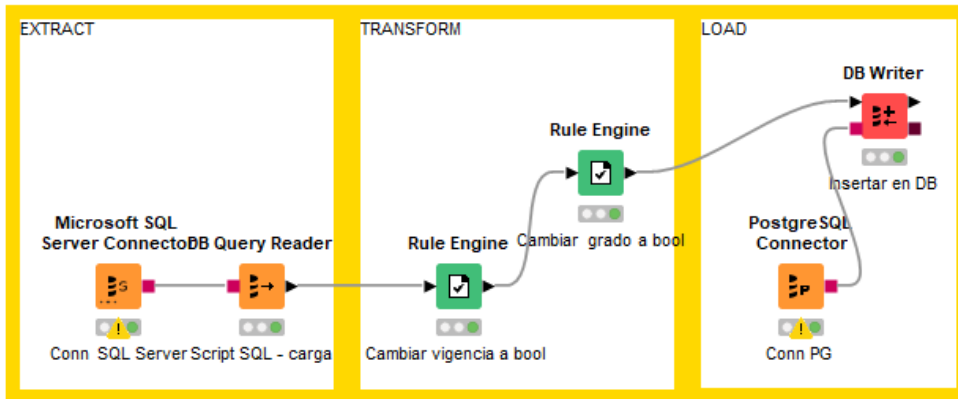


Ilustración 46: Proceso ETL para estudiante_proyecto

Consulta SQL:

SELECT

```
COD_EST AS codigo_estudiante,  
COD_PROYECTO_UIC AS id_proyecto,  
VIGENCIA AS estudiante_vigencia,  
0 AS grado
```

FROM dsa.proy_estud

Tabla seguimiento_proyecto

Nodos empleados:

ETL - seguimiento-proyecto

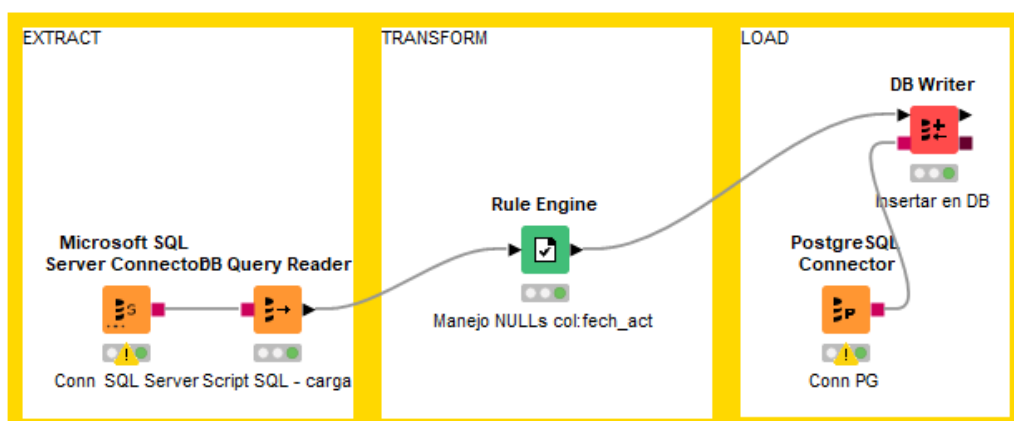


Ilustración 47: Proceso ETL para seguimiento_proyecto

Consulta SQL:

SELECT

```
" AS descripcion_seguimiento,  
FECHA_ACTIVIDAD AS fecha_actividad,  
FECHA_UIC AS fecha_uic,  
COD_PROYECTO_UIC AS id_proyecto,  
'proyecto' AS tipo,  
COD_PERIODO AS id_periodo,  
ID_PROCESO AS id_proceso
```

FROM dsa.segProy

Nodos empleados:

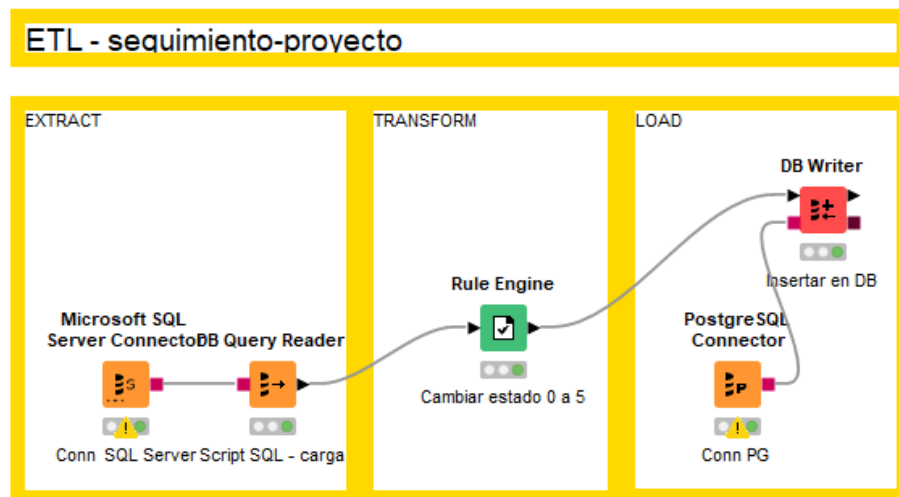


Ilustración 48: Proceso ETL para cargar datos faltantes en seguimiento_proyecto

Consulta SQL:

SELECT

```
COD_PROYECTO_UIC AS id_proyecto,  
ID_PROCESO AS id_proceso,  
COD_PERIODO AS id_periodo,  
" AS descripcion_seguimiento,  
CAST('1900-01-01' AS DATETIME) AS fecha_actividad,  
CAST('1900-01-01' AS DATETIME) AS fecha_uic,  
'proyecto' AS tipo
```

FROM dsa.proy_aprobado A

```
WHERE ID_PROCESO NOT IN (  
SELECT ID_PROCESO
```

FROM dsa.segProy B
WHERE A.COD_PROYECTO_UIC = B.COD_PROYECTO_UIC

)

Tabla docente

Nodos empleados:

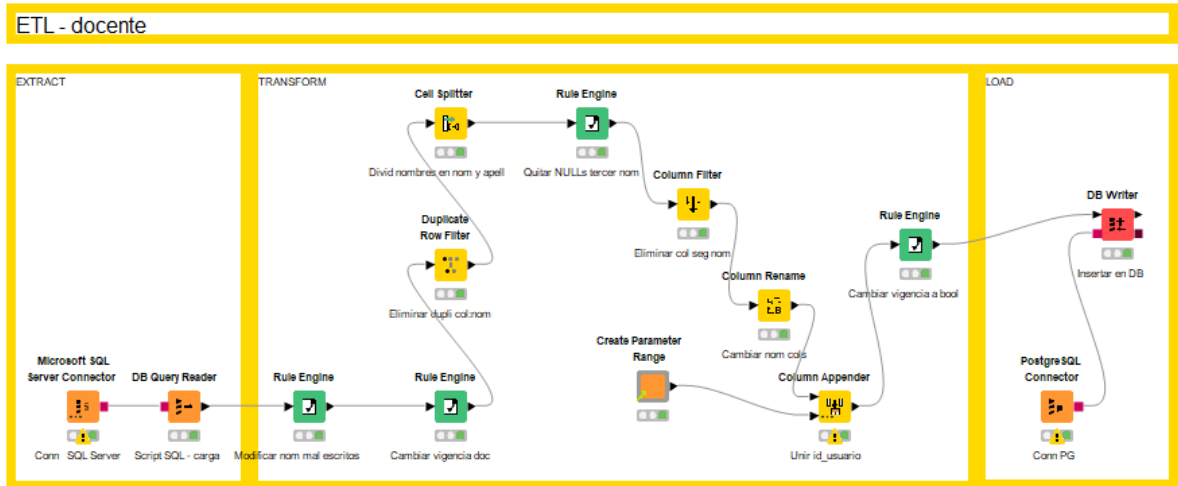


Ilustración 49: Proceso ETL para docente

Consulta SQL:

SELECT DISTINCT
 DOCENTE AS nombres,
 '0000000000' AS cedula_docente,
 1 AS vigencia
FROM dsa.proy_docente

Tabla docente_proyecto

Nodos empleados:

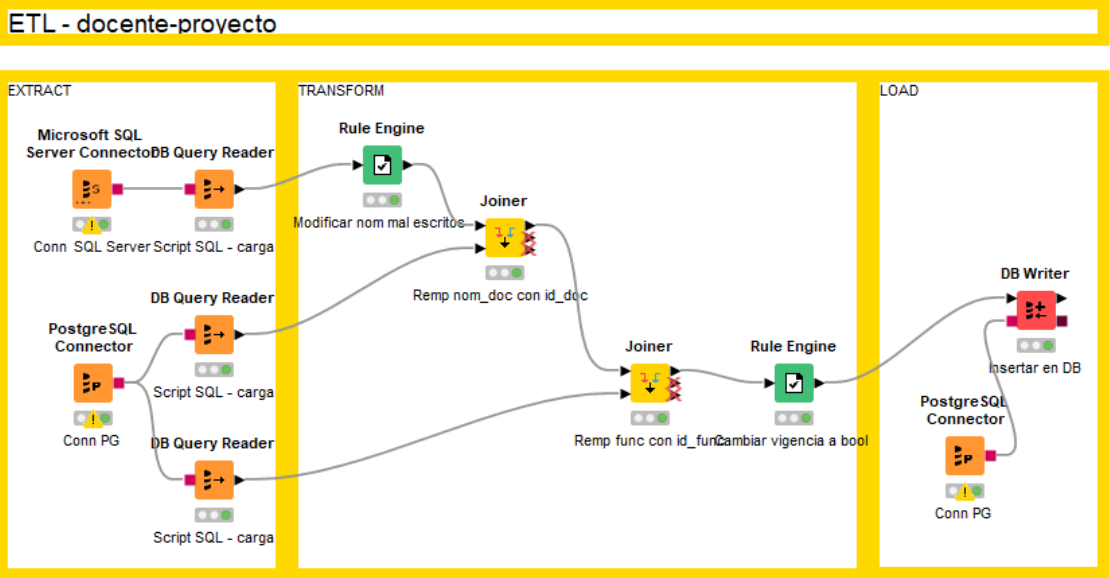


Ilustración 50: Proceso ETL para docente_proyecto

Consultas SQL:

```

SELECT
    DOCENTE AS id_docente,
    FUNCION AS id_funcion,
    COD_PROYECTO_UIC AS id_proyecto,
    VIGENCIA AS vigencia
FROM dsa.proy_docente
  
```

```

SELECT
    id_docente,
    nombre_docente || ' ' || apellido_docente AS nombres
FROM docente
  
```

```

SELECT
    id_funcion,
    funcion
FROM public.funcion
  
```

Tabla proyecto_linea

Nodos empleados:

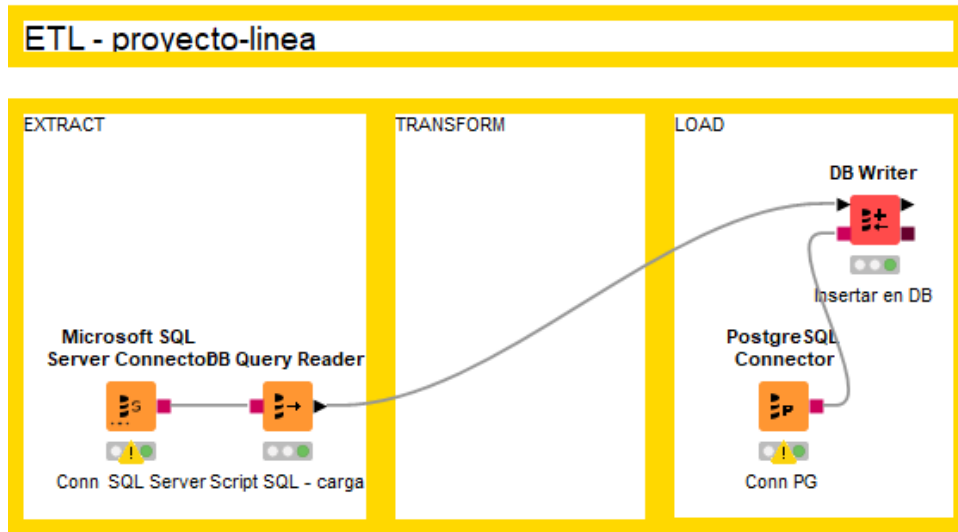


Ilustración 51: Proceso ETL para proyecto_linea

Consulta SQL:

SELECT

```
COD_PROG_INV AS codigo,  
COD_PROYECTO_UIC AS id_proyecto
```

FROM dsa.proy_lineaProgInv

Para la analítica DM2, se desarrollaron varias funciones en Python para la carga y limpieza de los trabajos de titulación. Se utilizó la librería Gensim, que proporciona las herramientas necesarias para el modelado de temas.

Como parte del proceso de limpieza, se cargaron los archivos y se removieron caracteres especiales; posteriormente, se extrajeron los lemas de las palabras; se removieron los nombres de personas para evitar que aparezcan como términos recurrentes; se dividieron las oraciones en palabras y se eliminaron las palabras vacías (artículos, preposiciones, conectores, etc.). La Ilustración 52 muestra el llamado a las funciones de limpieza.


```
#Clean text
lemmatized = lemmatization(pdfes)
no_names = remove_names(lemmatized)
words_sent = list(sent_to_words(no_names))
data_words = remove_stopwords(words_sent)

✓ 1m 29.8s
```

Ilustración 52: Proceso de limpieza de los trabajos de titulación

Con el módulo “corpora” de Gensim, se construyó un diccionario con las palabras obtenidas y un ID entero asignado a cada una. Finalmente, se transformó el diccionario a un formato de “bolsa de palabras” (*bag-of-words*), donde aparecen las distintas palabras y su frecuencia (*Ilustración 53*).

```
id2word = corpora.Dictionary(data_words)
texts = data_words
corpus = [id2word.doc2bow(text) for text in texts]

✓ 0.0s
```

Ilustración 53: Obtención del diccionario de palabras y corpus

4. Modelado

En esta etapa, se definió el diseño de evaluación y se construyeron los modelos de analítica de datos necesarios para cumplir con los objetivos de minería planteados en el proyecto. De las tres analíticas planteadas, solamente DM1 y DM2 requieren ser modeladas y evaluadas, ya que DM3 emplea únicamente estadística descriptiva.

4.1 Diseño de evaluación

En la evaluación del etiquetado gramatical (DM1) se calculó la exactitud (*Accuracy*). Se define como:

$$Exactitud = \frac{tp + tn}{tp + tn + fp + fn}$$

Donde:

tp: verdaderos positivos;

tn: verdaderos negativos;

fp: falsos positivos;

fn: falsos negativos.

La fórmula puede ser interpretada como:

$$\text{Exactitud} = \frac{\text{Número de palabras clasificadas correctamente}}{\text{Número total de palabras clasificadas}}$$

Ya que los títulos de los trabajos de titulación no se encuentran etiquetados, se evaluó el corpus AnCora, que consta de 500 000 palabras en español anotadas a diferentes niveles lingüísticos (<https://clic.ub.edu/corpus/es>).

Para la evaluación del modelado de temas, se empleó la medida de coherencia C_{UMass} , que cuantifica el grado de similitud semántica de las palabras más importantes en un tema o tópico. Este se calcula con la siguiente formula:

$$C_{UMass} = \sum_{m=2}^M \sum_{l=1}^{m-1} \log \frac{D(v_m^{(t)}, v_l^{(t)}) + 1}{D(v_l^{(t)})}$$

Sin embargo, Gensim cuenta con un módulo para determinar esta medida de forma automática.

4.2 Construcción de los modelos

En el etiquetado gramatical de los títulos de los trabajos de titulación, se empleó la librería spaCy, que cuenta con modelos pre entrenados en más de 60 idiomas, entre ellos español. A diferencia de otras librerías para procesamiento del lenguaje natural como NLTK, CoreNLP o ZPar, spaCy está enfocado en código de producción y puede llegar a ser más eficiente. En la *Ilustración 54* se muestra la importación del modelo en Python para su uso.

```
nlp = spacy.load("es_core_news_lg", exclude=['ner', 'parser'])  
d:\ESPOCH\IT\Analytics\venv\Lib\site-packages\tqdm\auto.py:21: TqdmWa  
from .autonotebook import tqdm as notebook_tqdm
```

Ilustración 54: Carga del modelo pre-entrenado para etiquetado gramatical

Se diseñó una función para extraer las palabras claves de cada título, filtrando únicamente los sustantivos, nombres propios y adjetivos. La función se presenta en la *Ilustración 55*.

```
def extract_keywords(title: str):
    doc = nlp(title)
    keywords = []
    for word in doc:
        if not word.is_stop and (word.pos_ in ["NOUN", "PROPN", "ADJ"]):
            keywords.append(word)
    return keywords
```

Ilustración 55: Función para extraer las palabras clave

El modelo *Latent Dirichlet Allocation* (LDA), disponible en la librería Gensim, fue utilizado en el desarrollo de la analítica DM2. Se seleccionó LDA debido a que es la técnica de modelado de temas más utilizada. La *Ilustración 56* muestra los parámetros escogidos para el entrenamiento del modelo.

```
lda_model = gensim.models.LdaModel(corpus=corpus,
    id2word=id2word,
    num_topics=10,
    random_state=100,
    chunksize=100,
    passes=10,
    per_word_topics=True)
```

Ilustración 56: Entrenamiento del modelo LDA

5. Evaluación

El resultado obtenido en el cálculo de la exactitud de la analítica DM1 se muestra en la *Ilustración 57*.

```
accuracy = results['pos_acc']
print(f'Exactitud del modelo "es_core_news_lg": {accuracy:.2f}')
✓ 0.0s
Exactitud del modelo "es_core_news_lg": 0.97
```

Ilustración 57: Precisión obtenida del modelo de etiquetado gramatical

La exactitud obtenida es de 0.97, lo que indica que 97% del corpus de prueba pudo ser etiquetado correctamente.

Al calcular la coherencia del modelo de la analítica DM2, obtenemos los resultados presentados en la *Ilustración 58*.

```
coherence_model_lda = CoherenceModel(model=lda_model,
                                     texts=data_words,
                                     dictionary=id2word,
                                     coherence='u_mass')
coherence_lda = coherence_model_lda.get_coherence()
print(f'Puntaje de coherencia: {coherence_lda:.2f}')
✓ 0.0s
Puntaje de coherencia: -4.21
```

Ilustración 58: Coherencia obtenida del modelo LDA

Se realizó un ajuste de los hiper parámetros α (densidad de tópicos en un documento), β (densidad de palabras en un tópico) y K (número de tópicos) del modelo para obtener distintos puntajes de coherencia y determinar la mejor configuración. Los valores de α y β se iteraron de 0.1 a 1 con un incremento de 0.3, mientras que K se incrementó de 2 a 30. Adicionalmente, se evaluaron los resultados para el 75% del corpus original, con el objetivo de determinar si existía una mejora en contraste al corpus completo. Los resultados del ajuste de hiper parámetros se muestra en la *Ilustración 59*, presentada a continuación.

```
results.loc[results['Coherencia'].idxmax()]
✓ 0.0s
Corpus          75% Corpus
Num_topicos      11
Alpha          0.9099999999999999
Beta            0.31
Coherencia      -1.602758
Name: 286, dtype: object
```

Ilustración 59: Resultados del ajuste de hiperparámetros

Se han seleccionado los hiper parámetros que acercan a 0 el valor de coherencia, pues esto indica un mayor grado de similitud semántica de las palabras en cada tópico.

6. Despliegue

Las analíticas desarrolladas se implementaron en la aplicación web “IMAK” durante el séptimo *sprint* (Para más información detallada sobre el proceso de despliegue, consulte el **MANUAL TÉCNICO – SCRUM – IMAK DSS**).

Para la primera analítica, DM1, se diseñó una nube de palabras que el usuario puede ajustar en un rango de periodos académicos, como se muestra en la *Ilustración 60*.



Ilustración 60: Nube de palabras con los términos recurrentes

La segunda analítica (DM2) se implementó utilizando la librería pyLDavis para Python, que permite visualizar los tópicos y términos más importantes generados por el modelo de forma interactiva. La *Ilustración 61* presenta el reporte obtenido.

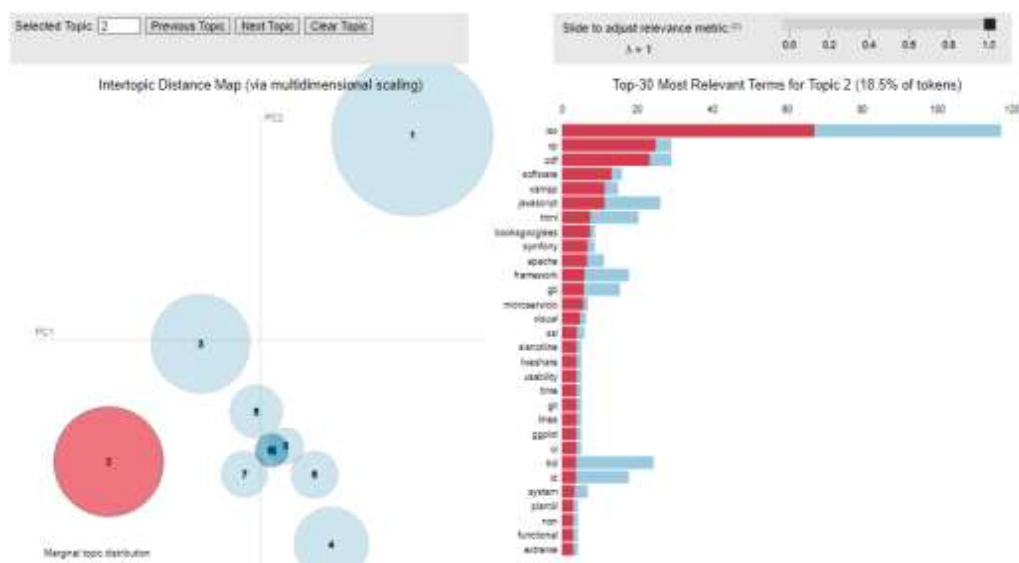


Ilustración 61: Visualización de los tópicos y términos más relevantes

La tercera analítica se representó utilizando un diagrama de violín (*Ilustración 62*), mostrando la distribución de la duración en meses de los trabajos de titulación. Además, se agregó una tabla con la información detallada de cada proyecto y la opción de filtrar las columnas según los requerimientos del usuario.

DISTRIBUCIÓN DE LA DURACIÓN EN MESES DE LOS PROYECTOS NO FINALIZADOS

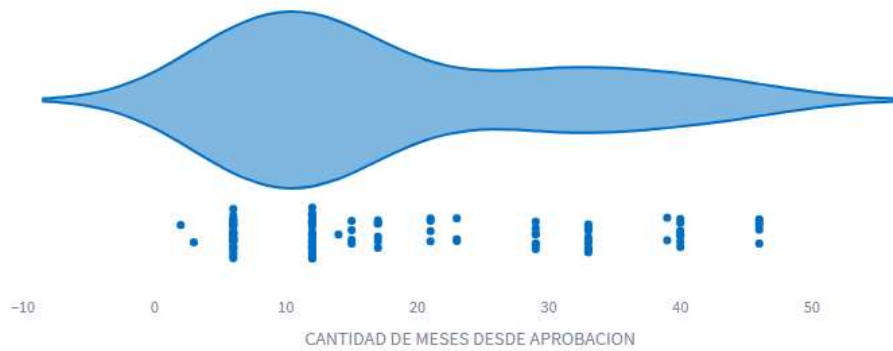


Ilustración 62: Distribución de la duración en meses de los proyectos no finalizados

ANEXO D: MANUAL DE USUARIO

MANUAL DE USUARIO

IMAK - DSS

Junio, 2023

Introducción

La presente guía tiene como objetivo describir las funcionalidades de la aplicación web para el apoyo a la toma de decisiones “IMAK”, brindando un instructivo para su uso. Para más información relativa a cuestiones técnicas de la aplicación y el contenido de la misma revise el **MANUAL TÉCNICO – CRISP-DM – IKMA DSS** y/o el **MANUAL TÉCNICO – SCRUM – IKMA DSS**).

1. Descripción de la aplicación

“IMAK”, acrónimo para *Integrated Management for Application and Knowledge*, es una aplicación web para el apoyo a la toma de decisiones de la Comisión de Titulación de la carrera de Software, ESPOCH. Cuenta con funcionalidades que le permiten visualizar información para la toma de decisiones a estudiantes, docentes y administrativos. “IMAK” utiliza los datos almacenados durante el proceso de titulación de los estudiantes para derivar información de valor que se presenta como gráficos personalizables y analíticas de datos que pueden ser explorados por los usuarios. Otras funcionalidades de la aplicación web son la capacidad de incluir datos externos en formato .xlsx, descargar los gráficos generados en formato .png y generar plantillas.

2. Estructura de la aplicación

“IMAK” cuenta con tres páginas, tal como se presentan a continuación:



- **Panel:** cuenta con la funcionalidad de visualización de gráficos personalizados e indicadores predeterminados, así como la carga de datos de una fuente externa. En caso de no estar autenticados, es también la página de inicio de sesión.

- **Analíticas:** muestra las analíticas de datos que incluyen las palabras recurrentes en los títulos de los trabajos de titulación, modelado de temas y la distribución de la duración en meses de los proyectos no finalizados.
- **Administración:** incluye funcionalidades para la actualización de los *datasets* (vistas materializadas) y la generación de plantillas en formato .xlsx para la carga de datos externos. Esta página es exclusiva para miembros de la Comisión de Titulación.

En los apartados siguientes se describirán a detalle cada una de las páginas mencionadas con las funcionalidades que incluyen.

3. “Panel”

La página “Panel” cuenta con la funcionalidad de visualización de gráficos personalizados e indicadores predeterminados, así como la carga de datos de una fuente externa. En caso de no estar autenticados, es también la página de inicio de sesión.

3.1 Inicio de sesión

Si es la primera vez que ingresa a la aplicación web o ha cerrado sesión, se le solicitará que ingrese un usuario y contraseña para poder utilizar las funcionalidades de “IMAK”.



The image shows a login form for the IMAK system. The form is contained within a light blue rounded rectangle. At the top left of the form, the text "IMAK - Inicio de sesión" is displayed in a bold, dark blue font. Below this, there are two input fields. The first is labeled "Usuario" in a small, light blue font, and the second is labeled "Contraseña" in the same font. The "Contraseña" field has a blue eye icon on its right side, indicating a toggle for password visibility. At the bottom left of the form, there is a button with the text "Iniciar sesión" in a dark blue font.

Ingrese el usuario (correo electrónico) y contraseña que utiliza para autenticarse en el sistema de gestión de los trabajos de titulación de la carrera.

¿Cómo obtengo un usuario y contraseña para ingresar a la aplicación?

Si usted es docente o administrador y no dispone de usuario y contraseña, solicítelos al administrador del sistema de gestión de los trabajos de titulación de la carrera.

Si usted es estudiante y no dispone de usuario y contraseña, puede registrarse en el sistema de gestión de los trabajos de titulación de la carrera y utilizar las mismas credenciales para autenticarse en “IMAK”.

3.2 Panel

Una vez autenticado en el sistema, usted verá una pantalla como la siguiente:



En la parte izquierda podrá encontrar el **menú desplegable**. En la parte superior de dicho menú, podrá cambiar de páginas mediante el **menú de navegación**; en la zona inferior se encuentra la **configuración del gráfico**. La mayor parte de la pantalla ocupan la **gráfica** y la **tabla de datos**, además de los **filtros**, ubicados a la derecha del menú desplegable.

Configuración del gráfico

Cada gráfico puede ser personalizado en varios atributos. La zona de configuración del gráfico puede ser utilizada para cambiar las propiedades de los indicadores seleccionados.

Selección de indicadores

La primera sección “**Indicadores**” permite escoger los indicadores deseados:



Indicadores

Indicador

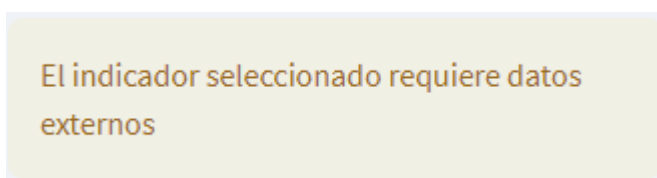
Personalizado

Cargar datos externos

Los indicadores predeterminados manejan campos y tipos de gráfico previamente configurados. Sin embargo, el indicador predeterminado **Personalizado** permite a cualquier usuario crear su propio gráfico al modificar los campos y características del mismo.

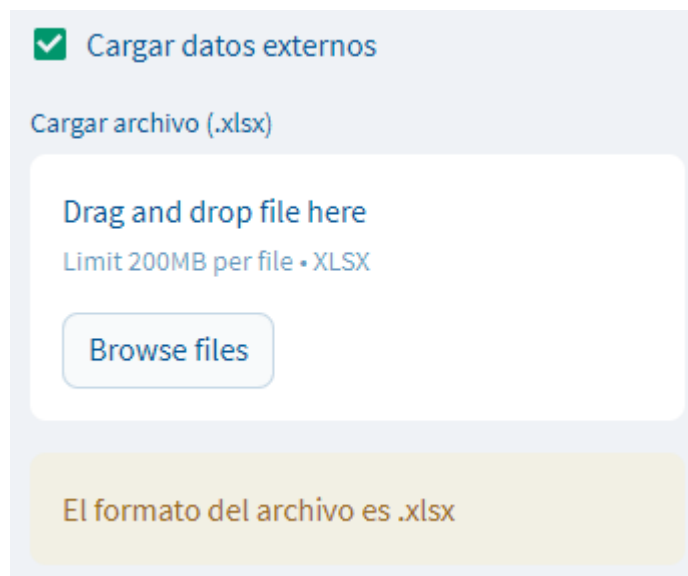
Carga de datos externos

Algunos indicadores podrían requerir la carga de datos externos. En estos casos, se mostrará el siguiente mensaje:



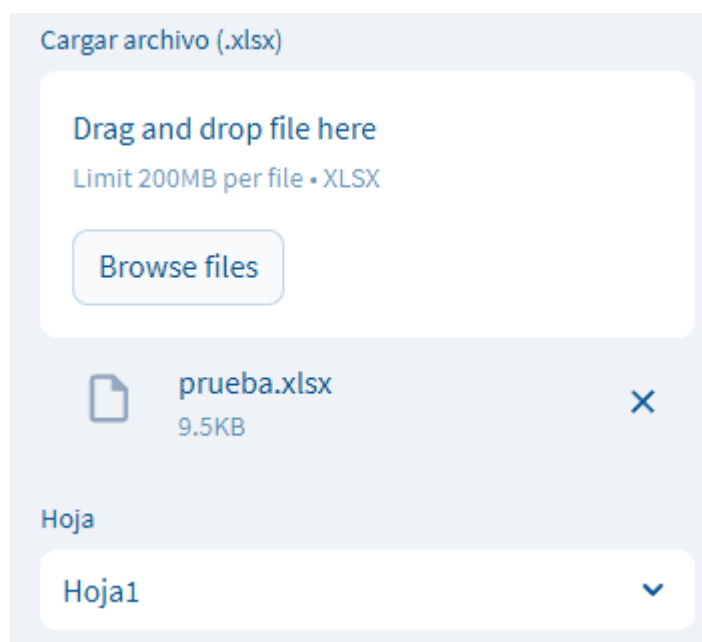
El indicador seleccionado requiere datos externos

El usuario podrá visualizar el gráfico, sin embargo, este no representará el indicador seleccionado en su total y deberá cargar los datos externos requeridos. Al marcar la casilla “**Cargar datos externos**”, el usuario podrá incluir un archivo cuyos registros se unirán con los datos del indicador seleccionado.



El archivo puede ser arrastrado directamente a dicha zona o cargado al presionar el botón “Browse file” que se muestra en la imagen anterior.

Una vez se haya cargado un archivo que cumpla con las condiciones (Ver siguiente apartado) el usuario podrá seleccionar la hoja de Excel que desea incorporar. Finalmente, los datos se habrán cargado y estarán disponibles para ser seleccionados en **Datos y visualización**.



Importante: Los datos externos NO se guardarán en la aplicación. Deberán ser cargados cada vez que se desee visualizar el indicador.

Condiciones para la carga de datos externos

- El formato del archivo debe ser .xlsx (Excel)
- El archivo puede contener varias hojas, pero en la aplicación podrá seleccionarse una a la vez.
- Debe existir una **única** columna cuyo nombre coincida con una de las columnas del indicador seleccionado. Así pues, si el indicador selecciona contiene una columna llamada “ID PERIODO”, el archivo de Excel deberá contener una columna con el mismo nombre.
- Las demás columnas del archivo no podrán coincidir en nombre con el resto de columnas del indicador.

Propiedades del gráfico

La siguiente sección corresponde a las propiedades del gráfico que se visualiza en la parte derecha de la pantalla:

Propiedades del gráfico

Tipo de gráfico

Barras

Título

Número de estudiantes con relación al número

Mostrar leyenda en X

Mostrar leyenda en Y

Mostrar leyenda (categorías)

Estilo

Por defecto

La configuración básica incluye:

- Cambiar el tipo de gráfico (barras, pastel, embudo, líneas o radar).
- El **título del gráfico**. En caso de haber seleccionado un indicador predeterminado, este ya tendrá un título por defecto, aunque sí puede ser modificado.
- **Mostrar leyenda en X**, que ocultará o mostrará el título del eje X en el gráfico (en algunos gráficos esta opción no tendrá efecto).
- **Mostrar leyenda en Y**, que ocultará o mostrará el título del eje Y en el gráfico (en algunos gráficos esta opción no tendrá efecto).
- **Mostrar leyendas**, que ocultará o mostrará un cuadro con las categorías del gráfico (dependerá del tipo de gráfico y si se ha seleccionado o no una columna como categoría).

- El **estilo** o color del gráfico (por defecto, harmony, monocromático o pastel).

Además, dependiendo del tipo de gráfico seleccionado, se mostrará una configuración específica que permite cambiar más aspectos estéticos de la figura.

Datos y visualización

En el último apartado del menú desplegable se pueden seleccionar los campos o columnas que se desean visualizar en el gráfico.

El menú desplegable 'Datos y visualización' contiene cuatro secciones de configuración:

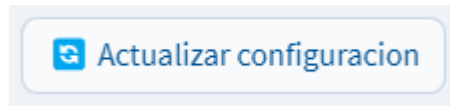
- Eje X:** Seleccionado 'ID PERIODO'.
- Categoría:** Seleccionado 'ID PERIODO'.
- Eje Y:** Seleccionado 'CANTIDAD DE PROYECTOS APROBADOS'.
- Función:** Seleccionado 'Ninguna'.

Las configuraciones en esta sección son:

- **Eje X:** campo o columna del indicador seleccionado que se ubicará en el eje X del gráfico. Usualmente está relacionado con variables categóricas, por lo que únicamente permitirá seleccionar columnas de este tipo.
- **Categoría:** campo o columna del indicador seleccionado que se desea emplear como las categorías del gráfico. Se trata de la segunda variable que puede ubicarse en el eje X. Si se coloca el mismo campo que el eje X entonces se omite la categoría.
- **Eje Y:** campo o columna del indicador seleccionado que se ubicará en el eje Y del gráfico. Permite tanto variables categóricas como numéricas, aunque se recomienda combinar estas últimas con una función.
- **Función:** función matemática que se aplica a la variable seleccionada en el eje Y. Se puede optar por el conteo (valor por defecto), suma, promedio, mínimo o máximo de valores numéricos en el eje Y. Para variables categóricas la única opción disponible es el conteo.

Actualizar configuración

Los miembros de la Comisión de Titulación y administradores pueden actualizar la configuración de los gráficos de los indicadores predeterminados.

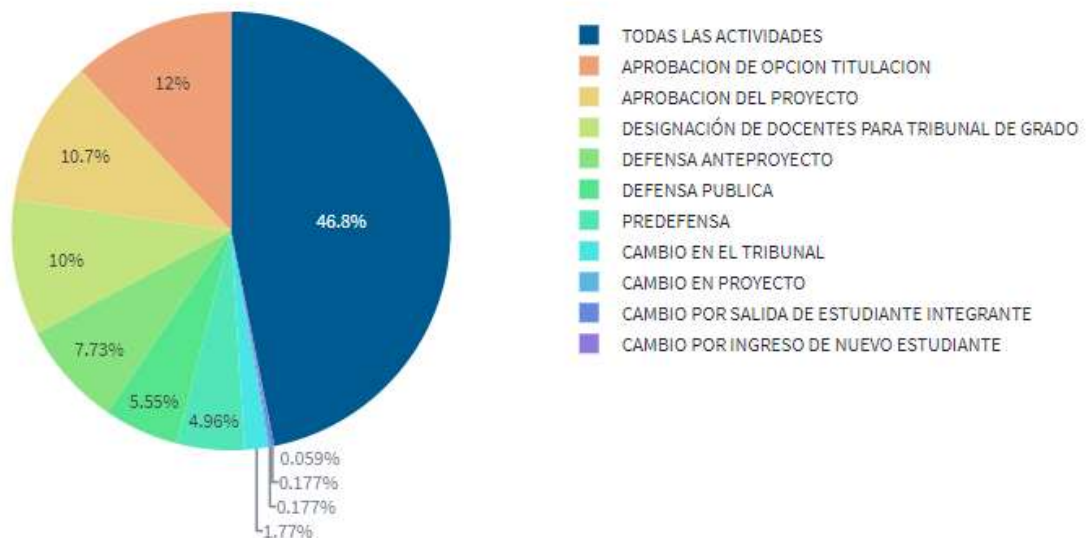


El botón “Actualizar configuración” aparecerá al final del menú desplegable de forma exclusiva para estos usuarios, quienes podrán modificar los valores y guardarlos para que otros usuarios puedan ver los gráficos con una configuración predeterminada.

Gráfico

El gráfico del indicador seleccionado aparecerá en la parte derecha de la pantalla con las configuraciones que hayamos realizado en el menú desplegable.

Número de trámites-actividades atendidas por la CT



La mayoría de gráficos cuentan con opciones adicionales para su manipulación. Dichas opciones están ubicadas en la parte superior derecha del gráfico y se encontrarán o no disponibles dependiendo de la configuración y los datos que hayamos seleccionado.



Estas opciones son, de izquierda a derecha:

- Descargar el gráfico como png.
- Desplazarse a través del gráfico (ideal cuando se amplía la imagen).
- Acercarse (zoom in).
- Alejarse (zoom out).
- Auto escalado del gráfico para que se ajuste a un tamaño adecuado.
- Expandir el gráfico para visualizarlo en pantalla completa.


Tabla

Cada gráfico viene acompañado de dos tablas que representan los datos visualizados. La **tabla completa** incluye todos los datos, tal y como se encuentran guardados en los *datasets* (vistas materializadas) de la base de datos. La **tabla filtrada** muestra únicamente los datos que han sido filtrados; estos son los que se visualizan en el gráfico.

Tabla

Completa Filtrada

	FECHA ACTIVIDAD SEGUIMIENTO PROYECTO	FECHA UIC SEGUIMIENTO PROYECTO	AÑO DE INICIO DEL PERIO
0	2020-03-09	2020-03-09	2019.0
1	2020-03-09	2020-03-09	2019.0
2	2020-03-09	2020-03-09	2019.0
3	2020-03-09	2020-03-09	2019.0
4	2020-03-09	2020-03-09	2019.0
5	2020-03-09	2020-03-09	2019.0
6	2020-03-09	2020-03-09	2019.0
7	2020-03-09	2020-03-09	2019.0
8	2020-03-09	2020-03-09	2019.0
9	2019-03-06	2019-03-06	2019.0

 Descargar tabla (.xlsx)

Al final de la página el usuario encontrará un botón “Descargar tabla (.xlsx)” que permite descargar en el computador la tabla seleccionada en un formato de hoja de Excel.

Carga de archivos

Para entrenar el modelo es necesario cargar los archivos de los trabajos de titulación que se deseen visualizar. El usuario puede arrastrar los documentos hasta la zona gris de la figura o dando clic en el botón “Browse files”:

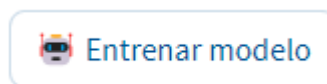
Cargar archivos (.pdf). Máximo 20 archivos



El máximo número de archivos aceptados es 20. En caso de cargar más de 20 archivos, el modelo se entrenará únicamente con los veinte primeros documentos seleccionados.

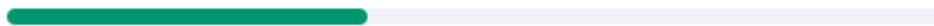
Entrenamiento del modelo

Una vez se hayan cargado los archivos aparecerá un botón para entrenar el modelo:



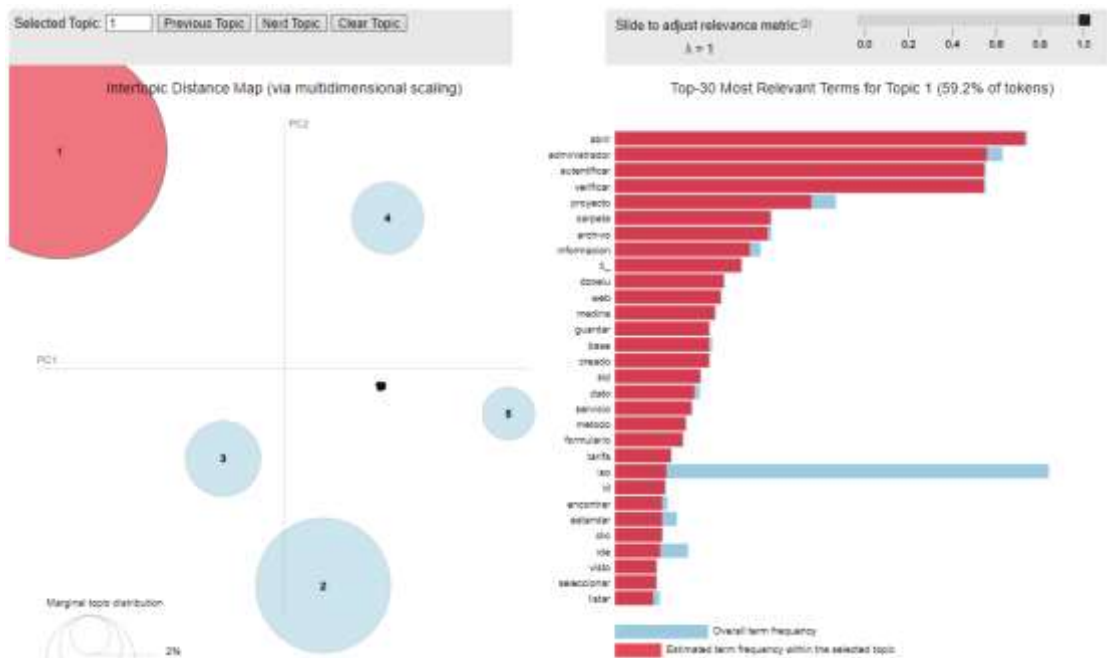
El proceso de entrenamiento del modelo puede tomar un tiempo considerable, especialmente si se trabaja con el número máximo de archivos, ya que primero se tendrá que transformar, limpiar y validar los archivos cargados.

Iniciando proceso de limpieza. Esta operación puede tardar varios minutos...



Reporte

Una vez culminado el proceso de entrenamiento del modelo, se mostrará un reporte que presenta de forma gráfica los tópicos en los que se han clasificado las palabras de los documentos cargados.



En la parte izquierda se visualizará los temas en los que se han clasificado los documentos. Cada círculo representa un tema y su tamaño está relacionado de forma proporcional a la importancia del tópico.

En la parte derecha se mostrarán los 30 términos más relevantes dentro de un tópico seleccionado. Se puede ajustar la métrica de relevancia para cambiar la disposición de los términos en la lista, deslizando la barra ubicada en la esquina superior derecha.

4.3 Proyectos no finalizados

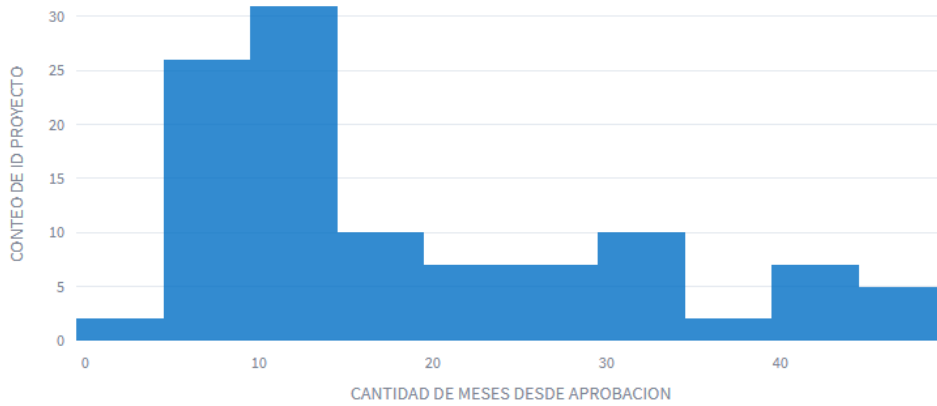
Esta analítica corresponde a información relacionada con los trabajos de titulación no finalizados, es decir, aquellos aprobados por la Comisión de Titulación pero que todavía no han presentado defensa pública.

Gráficos

La distribución de la duración en meses de los proyectos no finalizados se mostrará como dos gráficos: histograma y violín.

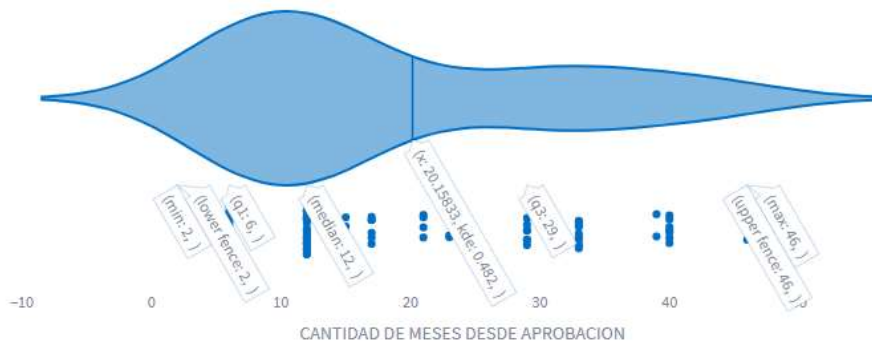
Histograma Violín

DISTRIBUCIÓN DE LA DURACIÓN EN MESES DE LOS PROYECTOS NO FINALIZADOS



Histograma Violín

DISTRIBUCIÓN DE LA DURACIÓN EN MESES DE LOS PROYECTOS NO FINALIZADOS



A diferencia del histograma, el diagrama de violín mostrará, además de la distribución de la duración en meses, los valores mínimo, máximo, mediana, rango intercuartílico, límites inferior y superior.

Tabla

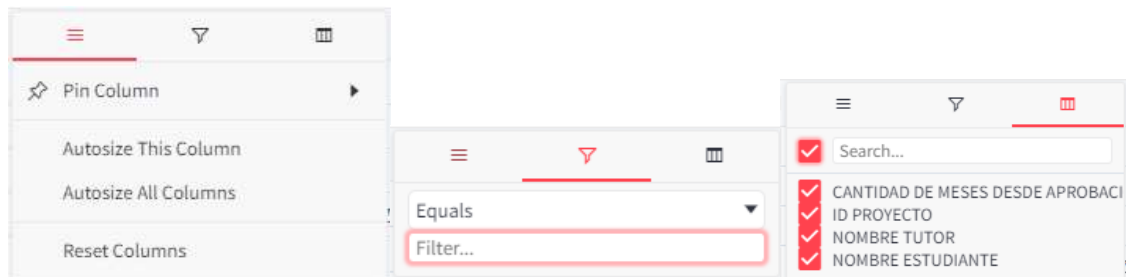
Debajo de los gráficos se encontrará la tabla con los datos principales de los proyectos no finalizados: cantidad de meses desde la aprobación del proyecto, ID, nombre del tutor y nombres de los estudiantes.

Tabla

CANTIDAD DE MESES DESDE APROBACION	ID PROYECTO	NOMBRE TUTOR	NOMBRE ESTUDIANTE	
<input type="checkbox"/>	2	232	OSCAR GAVILANEZ	GERMAN FRANCISCO RAMOS...
<input type="checkbox"/>	3	231	MIGUEL DUQUE	HILDA JHOANA CHICAIZA SU...
<input type="checkbox"/>	6	204	JULIO SANTILLAN	GERMAN FRANCISCO RAMOS...
<input type="checkbox"/>	6	205	IVONNE RODRIGUEZ	HOMERO VINICIO BARRAGA...
<input type="checkbox"/>	6	206	JORGE MENENDEZ	JEAN CARLOS CASTELO GUE...
<input type="checkbox"/>	6	207	GISEL BASTIDAS	CAROLINA BELEN CASTILLO ...
<input type="checkbox"/>	6	208	LORENA AGUIRRE	ANGELO ALE JANDRO CERNA...
<input type="checkbox"/>	6	209	IVONNE RODRIGUEZ	SEBASTIAN ANDRES COBOS ...
<input type="checkbox"/>	6	210	DIEGO AVILA	YANNIRA TATIANA CARRILLO ...
<input type="checkbox"/>	6	211	RAUL ROSERO	MARVIN RENATO ERAZO GIRON
<input type="checkbox"/>	6	212	ANGEL MENA	JOSE GREGORIO LOPEZ RIO...
<input type="checkbox"/>	6	214	MIGUEL DUQUE	WILSON ARMANDO LLUILEM...
<input type="checkbox"/>	6	215	GLORIA ARCOS	WASHINGTON DAVID LLUMIT...
<input type="checkbox"/>	6	216	VINICIO RAMOS	FRANCISCO XAVIER MESÍAS ...
<input type="checkbox"/>	6	217	GLORIA ARCOS	VICTOR MANUEL MONTALUI...

< 1 to 15 of 107 < < Page 1 of 8 > >

Al seleccionar los encabezados de cada columna, el usuario podrá acceder a una serie de configuraciones y filtros de la tabla.



Estos incluyen:

- Fijar columnas en la parte derecha o izquierda de la tabla.
- Ajustar automáticamente una columna en específico o todas las columnas.
- Filtrar datos. Si son datos numéricos entonces se deberá especificar el valor exacto que se desea filtrar; en el caso de los datos categóricos el usuario seleccionará o quitará aquellos valores que desee agregar o remover de la tabla respectivamente.
- Mostrar u ocultar columnas específicas.

Información

Al seleccionar cualquier fila de la tabla, se mostrará información más detallada en la parte inferior de la misma. Solo se podrá seleccionar un registro a la vez.

Información

	ID PROYECTO	ID PERIODO	NOMBRE TUTOR	NOMBRE ESTUDIANTE	TITULO PROYECTO
104	209	P0039	IVONNE RODRIGU	SEBASTIAN ANDRES COBOS MALDONA	DESARROLLO DE UNA /
105	209	P0039	IVONNE RODRIGU	JUAN CARLOS MAIGUA RIZO	DESARROLLO DE UNA /

La nueva tabla mostrará un registro por cada estudiante del proyecto. La información presentada incluye: ID del proyecto, ID del periodo académico en el que se aprobó el proyecto, nombre del tutor, nombre del estudiante, título del proyecto, fecha de aprobación y cantidad de meses desde la aprobación.

5 Administración

Esta página incluye funcionalidades para la actualización de los *datasets* (vistas materializadas) y la generación de plantillas en formato .xlsx para la carga de datos externos. Esta página es exclusiva para miembros de la Comisión de Titulación.

5.1 Actualizar *datasets*

Permite a los miembros de la Comisión de Titulación actualizar los datasets que almacenan los indicadores de la aplicación cada vez que se haya integrado nuevos datos a través del sistema de gestión de los trabajos de titulación de la carrera. Al trabajar con vistas materializadas, es necesario actualizarlas únicamente cuando haya cambios en las tablas principales de la base de datos para mantener así la eficiencia de las mismas.

Actualizar datasets ^

Actualizar todos los datasets

Datasets

Periodos, actividades y número de trámites atendidos por la CT ▼

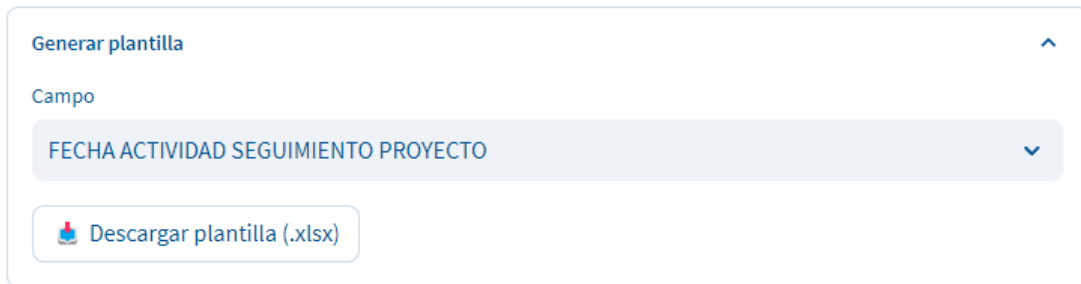
Vista materializada: ds_1

Actualizar

Al marcar la casilla “**Actualizar todos los datasets**”, se seleccionarán todos los datasets para ser actualizados. Si se mantiene esta casilla desmarcada y se escoge un dataset específico en la caja desplegable de selección, únicamente se actualizará la vista materializada señalada.

5.2 Generar plantilla

Los miembros de la Comisión de Titulación pueden generar plantillas en formato .xlsx con datos de las columnas específicas de su interés.



The image shows a user interface for generating a template. It features a title "Generar plantilla" with an upward arrow icon. Below the title is a label "Campo" followed by a dropdown menu. The dropdown menu is open, showing the selected option "FECHA ACTIVIDAD SEGUIMIENTO PROYECTO" with a downward arrow icon. Below the dropdown menu is a button labeled "Descargar plantilla (.xlsx)" with a download icon.

Al seleccionar un campo específico y dar clic sobre “**Descargar plantilla (.xlsx)**”, se guardará en el computador un archivo de Excel que contendrá una columna con el mismo nombre del campo seleccionado y todos los valores distintos que se almacenan en la base de datos para dicho campo.