EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA

TECNOLÓGICO NACIONAL DE MÉXICO

Tecnológico Nacional de México

Centro Nacional de Investigación y Desarrollo Tecnológico

Tesis de Maestría

Aplicando minería de datos a procesos de negocio para la mejora de indicadores de desempeño

presentada por

Ing. Emmanuel Rebolledo Torres

como requisito para la obtención del grado de

Maestro en Ciencias de la Ingeniería

Director de tesis

Dr. Hugo Estrada Esquivel

Codirector de tesis

Dra. Alicia Martínez Rebollar

Cuernavaca, Morelos, México. Febrero de 2025

Educación
Secretaría de Educación Pública

Centro Nacional de Investigación y Desarrollo tecnológico
Subdirección Académica

Cuernavaca Mor, 20/enero/2025
Oficio No. SAC/021/2025
Asunto: Autorización de impresión de tesis

EMMANUEL REBOLLEDO TORRES
CANDIDATO AL GRADO DE MAESTRO EN CIENCIAS DE LA INGENIERÍA
P R E S E N T E

Por este conducto, tengo el agrado de comunicarle que el Comité Tutorial asignado a su trabajo de tesis titulado **"Aplicando minería de datos a procesos de negocio para la mejora de indicadores de desempeño"**, ha informado a esta Subdirección Académica, que están de acuerdo con el trabajo presentado. Por lo anterior, se le autoriza a que proceda con la impresión definitiva de su trabajo de tesis.

Esperando que el logro del mismo sea acorde con sus aspiraciones profesionales, reciba un cordial saludo.

A T E N T A M E N T E
Excelencia en Educación Tecnológica®
"Conocimiento y Tecnología al Servicio de México"

EDUCACIÓN | TECNOLÓGICO NACIONAL DE MÉXICO
CENTRO NACIONAL DE INVESTIGACIÓN Y DESARROLLO TECNOLÓGICO
SUBDIRECCIÓN ACADÉMICA

CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO

c.c.p. Coordinación de Ingeniería
Departamento de Servicios Escolares

CMAZ/lmz

Educación
Secretaría de Educación Pública

Centro Nacional de Investigación y Desarrollo tecnológico
Coordinación de Ciencias de la Ingeniería

Cuernavaca, Mor., 20/enero/2025
OFICIO No. MCI/012/2025
Asunto: Aceptación de documento de tesis
CENIDET-AC-004-M14-OFICIO

DR. CARLOS MANUEL ASTORGA ZARAGOZA
SUBDIRECTOR ACADÉMICO
PRESENTE

Por este conducto, los integrantes de Comité Tutorial del C. Emmanuel Rebolledo Torres con número de control M23CE036, de la Maestría en Ciencias de la Ingeniería, le informamos que hemos revisado el trabajo de tesis de grado titulado **"Aplicando minería de datos a procesos de negocio para la mejora de indicadores de desempeño"** y hemos encontrado que se han atendido todas las observaciones que se le indicaron, por lo que hemos acordado aceptar el documento de tesis y le solicitamos la autorización de impresión definitiva.

Dr. Hugo Estrada Esquivel
Director de tesis

Dra. Alicia Martínez Rebollar
Codirector de tesis

Dr. Juan Reyes Reyes
Revisor 1

M.T.I. Rosa Olivia Maquinay Díaz
Revisor 2

C.c.p. Depto. Servicios Escolares
Expediente / Estudiante
Iniciales AMR/mrsr

Agradecimientos

Al Consejo Nacional de Ciencia y Tecnología (CONACYT) por el apoyo económico otorgado a través de la beca nacional No. 1267449, que hizo posible la realización de mis estudios de maestría.

Al Tecnológico Nacional de México (TecNM) y al Centro Nacional de Investigación y Desarrollo Tecnológico (CENIDET) por la oportunidad de realizar mis estudios de maestría y brindarme las herramientas para mi formación profesional.

Al Dr. Hugo Estrada Esquivel y la Dra. Alicia Martínez Rebollar, por su guía y apoyo constante durante el desarrollo de esta investigación. Les agradezco su confianza, sus enseñanzas y la paciencia que me demostraron mientras me guiaban en este camino de aprendizaje. Muchas gracias.

A la Mtra. Olivia Maquinay Diaz y al Dr. Juan Reyes Reyes, por su participación como miembros de mi comité revisor. Sus comentarios y sugerencias fueron fundamentales para mejorar esta investigación.

A mi madre, por ser la razón por la cual quiero superarme día con día, ser mi inspiración y el pilar de mi vida.

A mis amigos, por acompañarme en mi proceso de formación, y darme su apoyo para crecer personalmente.

A todos y cada uno de ustedes, mi más profundo agradecimiento por su contribución y apoyo.

Contenido

Capítulo 1. Introducción 1
1.1 Introducción/motivación 2
1.2 Planteamiento del problema. 3
1.3 Objetivos. 4
1.4 Propuesta de solución 4
1.5 Estructura de la tesis 5
Capítulo 2. Marco conceptual 6
2.1 Minería de datos 7
2.2 Procesos de negocio 9
2.3 Indicadores Clave de Desempeño 12
Capítulo 3. Estado del arte 13
3.1 Aplicación de minería de datos en procesos para la mejora de operaciones empresariales 15
3.2 Aplicación de minería de datos en procesos para la mejora de indicadores clave de desempeño 17
3.3 Aplicación de minería de datos en procesos a partir del procesamiento de registros de eventos 18
3.4 Conclusiones del estado del arte 22
Capítulo 4: Metodología para la mejora de indicadores clave de desempeño mediante minería de datos 23
4.1 Descripción de la metodología 24
4.2 Fase 1. Entendimiento del negocio 27
4.3 Fase 2. Entendimiento de los datos 29
4.4 Fase 3. Procesamiento de los datos. 32
4.5 Fase 4. Desarrollo del modelo 36
4.6 Fase 5. Interpretación y optimización 37
Capítulo 5. Plan de pruebas y resultados 38
5.1 Caso de estudio 39
5.2 Fase 1. Entendimiento del negocio 39
5.3 Fase 2. Entendimiento de los datos 42
5.4 Fase 3. Procesamiento de los datos. 47

5.5 Fase 4. Desarrollo del modelo 59
5.6 Fase 5. Interpretación y optimización 62
Capítulo 6. Conclusiones 69
6.1 Conclusiones generales 70
6.2 Trabajos futuros 72
6.3 Logros obtenidos 72
Referencias 73
Anexos 78
Anexo A. Diccionario de las variables de la base de datos del caso de estudio. 78
Anexo B. Código de la funcionalidad de eventos 83

Contenido Tablas

Tabla 1 Tabla de elementos de la notación BPMN 10
Tabla 2 Interpretación de la magnitud del coeficiente de Pearson (R)..... 34
Tabla 3 Interpretación del valor V de Cramer..... 35
Tabla 4 Atributos de la tabla “Bitacora_proyectos”..... 44
Tabla 5 Estructura MVC del algoritmo A 45
Tabla 6 Estructura MVC del algoritmo B 46
Tabla 7 Variables del primer conjunto de datos..... 47
Tabla 8 Rangos de vistas 48
Tabla 9 Rangos de la variable "Quincena de registro"..... 49
Tabla 10 Rangos de la variable "Mes de registro" 49
Tabla 11 Impactos encontrados en la base de datos 51
Tabla 12 Variables procesadas del conjunto de datos..... 53
Tabla 13 Asociaciones encontradas de la variable "Selección". 57
Tabla 14 Variables que están asociadas directamente con la selección de los proyectos 58
Tabla 15 Simbología del flujo de trabajo 1. 59
Tabla 16 Simbología del nodo Decision Tree Learner del flujo de trabajo 2 60
Tabla 17 Simbología de los nodos del flujo de trabajo 3 60
Tabla 18 Medidas estadísticas de las vistas de los proyectos seleccionados y no seleccionados 63
Tabla 19 Medidas estadísticas de las vistas de los proyectos seleccionados y no seleccionados 65
Tabla 20 Porcentajes de selección de proyectos por institución y modalidad..... 66

Contenido Figuras

Figura 1 Estructura de los árboles de decisión 8
Figura 2 Flujo de la metodología para la mejora de indicadores clave de desempeño mediante minería de datos 24
Figura 3 Metaproceso del caso de estudio 39
Figura 4 Proceso de negocio en notación BPMN: Selección de proyectos 40
Figura 5 Tasa de asignación de proyectos del año 2023 40
Figura 6 Tasa de asignación de proyectos del año 2024 41
Figura 7 Gráfico de barras de las preferencias de los proyectos por temática 43
Figura 8 Gráfico de barras de las preferencias de los proyectos por modalidad 43
Figura 9 Proceso de "Selección de proyectos" con la funcionalidad de eventos 46
Figura 10 Nube de palabras de requisitos y prerrequisitos frecuentes 50
Figura 11 Nube de palabras de las palabras clave de los proyectos 51
Figura 12 Flujo de trabajo 1 59
Figura 13 Flujo de trabajo 2 60
Figura 14 Flujo de trabajo 3 61
Figura 15 Métricas del modelo enfocado a las correlaciones de la variable: Selección 61
Figura 16 Número de vistas de proyectos seleccionados 62
Figura 17 Número de vistas de proyectos no seleccionados 62
Figura 18 Número total de vistas por temática. 63
Figura 19 Porcentaje de vistas realizadas por institución 64
Figura 20 Gráfica de barras de la selección de proyectos registrados en cada quincena desde que salió la convocatoria 64
Figura 21 Gráfica de dispersión: Días transcurridos desde que salió la convocatoria vs la cantidad de vistas 65
Figura 22 Gráfica de proyectos seleccionados por preferencia de institución 66
Figura 23 Preferencia de proyectos por modalidad 67

Capítulo 1. Introducción

Capítulo 1. Introducción

1.1 Introducción/motivación

En el entorno empresarial actual, las organizaciones enfrentan diversos factores externos, como la evolución tecnológica y la creciente competitividad del mercado. Sin embargo, también deben lidiar con factores internos, como la mala toma de decisiones, procesos ineficientes, resistencia al cambio y falta de comunicación, que pueden obstaculizar el logro de los objetivos establecidos. En este contexto, las empresas no solo deben centrarse en mejorar los productos o servicios que ofrecen, sino también en optimizar sus operaciones internas, que están estrechamente relacionadas con la gestión efectiva de sus procesos de negocio.

Los procesos de negocio consisten en una secuencia de actividades interrelacionadas que tienen como objetivo satisfacer las necesidades de todos los actores involucrados (clientes, empleados y gerentes) y establecer una ruta clara hacia el cumplimiento de los objetivos deseados (Ghasemi & Amyot, 2019; Van der Aalst, 2016). El grado de cumplimiento de estos objetivos se mide mediante indicadores clave de desempeño (KPI, Key Performance Indicators), que también reflejan la eficiencia de los procesos y las actividades que los componen. En la actualidad, muchas organizaciones utilizan plataformas y software para dar seguimiento a sus procesos de negocio con el objetivo de mejorar los tiempos de ejecución de cada actividad. Estas herramientas registran procesos, actividades, recursos y actores involucrados en bases de datos, las cuales son un recurso valioso para identificar patrones que revelen ineficiencias u oportunidades de mejora en los procesos.

Este trabajo de investigación propone un marco de trabajo que facilite el seguimiento y la extracción de conocimiento a partir de las bases de datos de los procesos de negocio, con el objetivo de que las organizaciones puedan desarrollar planes de mejora y optimizar sus indicadores clave de desempeño.

1.2 Planteamiento del problema.

Los procesos de negocio se conforman de una secuencia de actividades que le permiten a una empresa alcanzar sus objetivos estratégicos. El éxito de éstos puede ser medido mediante los indicadores clave de rendimiento (KPI, Key Performance Indicator), los cuales evalúan y cuantifican el desempeño de las actividades con respecto al cumplimiento de sus objetivos estratégicos, es decir, comparan los resultados de un proceso ya ejecutado con los objetivos deseados (Giraldo, Jiménez, & Montoya, 2015). Es fundamental que las empresas realicen un análisis de sus procesos de negocio para identificar problemas y monitorear el comportamiento de actividades realizadas y actividades en ejecución (zur Muehlen & Shapiro, 2015). En la actualidad existen metodologías que tienen la finalidad de mejorar el desempeño de los procesos con el fin de cumplir los objetivos deseados dentro de las organizaciones, sin embargo, implementarlas suele ser un desafío debido a la resistencia al cambio dentro de la organización, además de que solventan los problemas, pero no se enfocan en buscar la causa raíz (Becker, Pfeiffer, Räquers, Falk, & Czerwonka, 2014).

La falta de información para aquellos que gestionan la ejecución de los procesos representa un obstáculo para las organizaciones, debido a que estos no pueden administrar correctamente la priorización de las tareas que componen sus procesos y se les dificulta seleccionar sus indicadores de desempeño (Danilova, 2018).

Este trabajo de investigación propone una metodología que facilite la identificación de problemáticas y áreas de oportunidad en los procesos de negocio mediante el análisis de los datos de la ejecución de sus actividades.

1.3 Objetivos.

1.3.1 Objetivo General.

Desarrollar una metodología que permita el análisis de datos provenientes de la ejecución de procesos de negocio de una organización para obtener correlaciones o asociaciones entre los procesos, las personas que lo ejecutan, las áreas relacionadas y el desempeño de la ejecución de los procesos de negocio utilizando minería de datos.

1.3.2 Objetivos específicos.

- Crear una base de datos procesada que sirva de entrada para realizar el análisis de datos de la ejecución de procesos de negocio utilizando un caso de estudio específico.
- Desarrollar un modelo de análisis de datos que permita encontrar asociaciones entre las variables.
- Desarrollar un plan de pruebas.
- Documentar los pasos que componen la metodología para que pueda ser replicada

1.4 Propuesta de solución

Se propone realizar una metodología que proporcione información detallada del proceso de extracción de conocimiento de los datos generados de los procesos de negocio para mejorar los indicadores clave de desempeño. Para desarrollar la metodología se utiliza la siguiente metodología de solución.

Fase 1: Fundamentación conceptual

- Desarrollo del marco teórico
- Desarrollo del estado del arte

Fase 2: Selección y comprensión del caso de estudio

- Selección del caso de estudio
- Entendimiento del negocio del caso de estudio
- Entendimiento de los datos del caso de estudio
- Estudio de técnicas de minería de datos

Fase 3: Adquisición y preparación de los datos

- Creación de un conjunto de datos para el caso de estudio
- Desarrollo de un daemon que recopila información de actividad de los usuarios
- Selección de una técnica de minería de datos

Fase 4: Construcción del modelo

- Desarrollo de modelo de reconocimiento de patrones para el caso de estudio
- Desarrollo de un plan de pruebas para el caso de estudio

Fase 5: Interpretación y documentación

- Interpretación de resultados y generación de recomendaciones
- Documentar la metodología propuesta para detección de patrones en modelos de negocios

1.5 Estructura de la tesis

El presente documento está estructurado en seis capítulos: el capítulo 2 desarrolla el marco conceptual, proporcionando las bases teóricas relacionadas con el tema. El capítulo 3 aborda el estado del arte, destacando los trabajos más relevantes y relacionados con la temática en cuestión. En el capítulo 4 se presenta la metodología propuesta para la mejora de indicadores clave de desempeño mediante la aplicación de minería de datos. El capítulo 5 detalla un plan de pruebas donde se implementa la metodología en un caso de estudio específico. Por último, el capítulo 6 expone las conclusiones derivadas de la implementación y análisis de la metodología propuesta.

Capítulo 2. Marco conceptual

Capítulo 2. Marco conceptual

El Capítulo 2. Marco conceptual aborda los conceptos fundamentales necesarios para comprender este trabajo de investigación, incluyendo los métodos y herramientas relacionados con la minería de datos, los procesos de negocio y los indicadores clave de desempeño.

2.1 Minería de datos

La minería de datos es un proceso que implica la búsqueda de patrones en bases de datos para la extracción de conocimiento útil (Bose & Mahapatra, 2001). Se define como un proceso debido a que la implementación de técnicas y tareas para el procesamiento de datos se realiza de manera secuencial. Este enfoque ha dado lugar al desarrollo de metodologías específicas que estructuran y guían las etapas de la minería de datos (Moine, Haedo, & Gordillo, 2011).

2.1.1 Metodología CRISP-DM

Una de las metodologías más utilizadas en minería de datos es CRISP-DM (Cross-Industry Standard Process for Data Mining) que traslada problemas empresariales a tareas de minería de datos, guiando los proyectos desde su inicio hasta su conclusión (Huber, Wiemer H, Shneider, & Ihlenfeldt, 2019). Esta metodología se compone de las siguientes etapas:

1. Entendimiento del negocio: Esta etapa implica comprender los objetivos del proyecto, las necesidades desde una perspectiva comercial y transformar este conocimiento en una problemática que pueda resolverse con técnicas de minería de datos (Wegener & Rüping , 2010).

2. Entendimiento de los datos: Esta fase implica conocer los datos e identificar los problemas (Wegener & Rüping , 2010). También se formulan hipótesis para la información oculta con respecto a los del proyecto de minería de datos (Huber, Wiemer H, Shneider, & Ihlenfeldt, 2019).

3. Preparación de los datos: En esta etapa se filtran y reducen los datos, dejando solo los datos relevantes con respecto a los objetivos definidos (Huber, Wiemer H, Shneider, & Ihlenfeldt, 2019).

4. Modelado: En esta fase se define un flujo de trabajo de minería de datos para encontrar los parámetros deseados para los algoritmos seleccionados (Huber, Wiemer H, Shneider, & Ihlenfeldt, 2019).

5. Evaluación: Se evalúa que tan acertado es el modelo construido con respecto a los objetivos deseados (Huber, Wiemer H, Shneider, & Ihlenfeldt, 2019).

6. Despliegue: Finalmente, los resultados de la minería de datos y los conocimientos adquiridos se entregan al usuario, llegando desde generar un simple informe hasta una implementación compleja de un proceso repetitivo de minería de datos (Wegener & Rüping , 2010) .

2.1.2 Modelos de clasificación

Los modelos de clasificación permiten categorizar elementos en grupos definidos previamente, lo cual es útil en tareas como la segmentación de clientes o la clasificación de productos. Los modelos más comunes son los árboles de decisión, los cuales consisten en un conjunto de reglas de clasificación que permita determinar la clase (variable objetivo) a la que pertenecen una serie de objetos (instancias de la base de datos) dado los valores de sus variables (Coll, Rodríguez, Gutiérrez, Moreno, & Llanes-Santiago, 2022). La estructura de los árboles de decisión se muestra en la Figura 1.

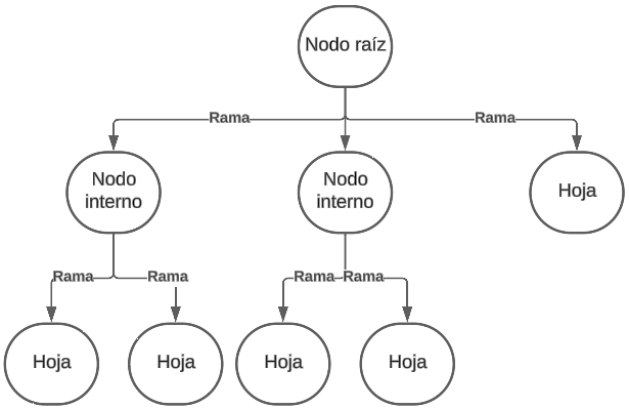


Figura 1 Estructura de los árboles de decisión

- Nodo raíz: El nodo raíz se caracteriza por ser el único nodo que no tiene una rama en la entrada, pero si varias ramas a la salida
- Nodo interno: Los nodos internos son aquellos que tienen una rama en la entrada y varias ramas a la salida.
- Hoja: Las hojas también conocidas como nodos hoja se caracterizan por tener una rama a la entrada y ninguna rama a la salida.
- Rama: Es la unión entre nodos.

Cada nodo de los árboles de decisión representa una variable de la base de datos. Además, dentro de este modelo existen diferentes algoritmos como el C4.5 e ID3. El método C4.5 emplea un enfoque recursivo para generar un árbol de decisión a partir de los datos. Utiliza

la estrategia de profundidad-primero para construir el árbol. Durante este proceso, el algoritmo considera todas las pruebas posibles para dividir el conjunto de datos y elige aquella variable que resulta con mayor ganancia de información. Para los atributos discretos, se evalúa una prueba con múltiples resultados, correspondientes al número de valores que el atributo puede tomar. En el caso de atributos continuos, se realiza una prueba binaria sobre cada valor del atributo en los datos. En cada nodo del árbol, se toma la decisión sobre qué prueba utilizar para dividir los datos (Ruggieri, S., 2002).

2.2 Procesos de negocio

Los procesos de negocio una serie de actividades que se llevan a cabo dentro de una organización con el objetivo de transformar insumos en productos o servicios con un valor añadido. Estas actividades están diseñadas para cumplir con los objetivos y metas establecidos por la organización.

2.2.1 Gestión de procesos de negocio

La gestión de procesos de negocio (Business Process Management, BPM) es un conjunto de prácticas de análisis y gestión enfocado a los procesos de negocio, ayuda a mejorar la eficiencia de los servicios que generan valor (Hitpass, 2017). La gestión de procesos de negocio administra cadenas enteras de eventos, actividades y decisiones que aportan valor a la empresa y a sus clientes (Doumas, La Rosa, Mendling, & Reijers, 2013). Este concepto surgió a partir de los conceptos de Reingeniería de Procesos de Negocio (BPR) y Total Quality Management (TQM) (Gras, 2019).

2.2.2 Notación de gestión de procesos de negocio

La notación de gestión de procesos de negocio (Business Process Modeling Notation, BPMN) es un medio de modelado utilizado por la gestión de procesos (BPM). Este esquematiza los procesos de negocio de una organización, de esta manera se facilita la comprensión y visualización del inicio, desarrollo y final de los procesos (SYDLE, 2022).

Se pueden distinguirlos tipos de modelado en esta notación cuando cumplen los siguientes criterios:

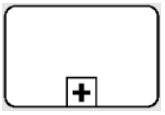
- Mapas de procesos: Diagramas simples del flujo de actividades un proceso, no tienen más detalles que los nombres de las actividades y condiciones de decisiones.

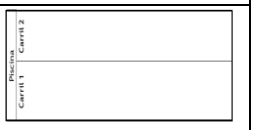
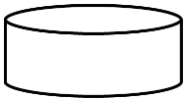
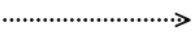
- Descripción de procesos: Proporcionan más información de los procesos como los roles involucrados, intercambio de datos e información, etc.

- Modelos de proceso: Proporcionan suficiente información como para analizar el proceso y simularlo. Esta clase de proceso permite ejecutar el modelo.

La notación BPMN tiene distintos elementos que permiten representar las actividades de los procesos (White, S. A., & Miers, D, 2009). La Tabla 1 muestra la simbología de los más recurrentes en la notación BPMN.

Tabla 1 Tabla de elementos de la notación BPMN

Elemento	Descripción	Imagen
Actividad o tarea	Las actividades son el elemento más básico en los modelos de procesos de negocio, por lo general requieren una entrada y produce una salida, dentro de ellos debe estar escrito al inicio con un verbo en infinitivo dependiendo del contexto del proceso.	
Subproceso	Un subproceso es un elemento que se utiliza para encapsular y organizar un conjunto de actividades dentro de un proceso más grande.	
Eventos		
Evento de inicio	Un evento de inicio muestra donde comienza un proceso	
Evento de inicio por mensaje	Este tipo de eventos indican que el proceso inicia al recibir un mensaje	
Evento de inicio por tiempo	El evento inicia por una fecha y hora especifico	
Eventos intermedios	Los eventos intermedios se colocan dentro del flujo del proceso para representar cosas que suceden durante las operaciones normales del proceso y generalmente ocurren entre las actividades	
Evento final	Un evento final indica cuando termino un proceso	
Evento final por mensaje	Este elemento indica que el proceso terminó con un mensaje	
Evento final por	El proceso termina en una fecha y hora especifica	
Compuertas		
Compuerta exclusiva	Representa una decisión en el flujo del proceso donde solo se puede elegir un camino. Se utiliza para modelar situaciones de "si-entonces", donde solo una de varias condiciones es verdadera.	

Compuerta paralela	Representa una convergencia o divergencia en el flujo del proceso donde múltiples caminos pueden ejecutarse en paralelo o converger. Se utiliza para modelar situaciones donde varias tareas pueden ocurrir simultáneamente.	
Compuerta inclusiva	Representa una decisión en el flujo del proceso donde múltiples caminos pueden ser elegidos pero un camino debe ser obligatorio. Se utiliza para modelar situaciones donde más de una condición puede ser verdadera.	
Compuerta compleja	Representa una decisión más compleja donde se pueden modelar condiciones más sofisticadas utilizando expresiones. Permite expresar lógica compleja para tomar decisiones.	
Piscinas		
Piscina	Una piscina es una forma de agrupar actividades y procesos que involucran a múltiples participantes o entidades. Esta piscina puede ser dividida en particiones a los cuales se les llama carriles.	
Objetos		
Objeto de datos	Se utiliza para representar información que es consumida o producida por una actividad en un proceso de negocio. Un objeto de datos en BPMN puede ser cualquier pieza de información relevante para el proceso y puede representar tanto datos estructurados como no estructurados.	
Almacén de datos	Se refiere a la representación de un lugar donde se almacenan o recuperan datos durante el transcurso de un proceso de negocio. La notación de almacenamiento de datos se utiliza para mostrar cómo los procesos interactúan con bases de datos, sistemas de archivos u otros repositorios de datos.	
Conectores		
Flujo de secuencia	Representa la secuencia lógica y direccional entre dos elementos en un proceso. Indica el orden en el que las actividades deben ejecutarse.	
Flujo de mensaje	Representa el intercambio de mensajes entre dos participantes o procesos. Indica la comunicación entre entidades o procesos en el contexto de colaboración.	
Asociación	Representa una relación semántica o de asociación entre elementos sin indicar flujo de control. Utilizado para mostrar conexiones significativas entre elementos, como información, documentos o comentarios.	

2.3 Indicadores Clave de Desempeño

Los indicadores clave de desempeño, también conocidos como KPI (Key Performance Indicator), son una herramienta que le permite a las organizaciones medir el rendimiento de los objetivos estratégicos y operativos a lo largo del tiempo. Se caracterizan por evaluar el desempeño de los procesos y/u operaciones de una organización para medir el logro de metas operativas y estratégicas (Giraldo, Jiménez, & Montoya, 2015).

Existe una gran variedad de KPI, ya que dependiendo del contexto de la organización y en qué sector realizan sus actividades se definen los indicadores pueden tener los procesos (Santander Universidades, 2022), algunos ejemplos son:

- KPI de producción: Volumen de producción, rotación del inventario, consecución de la producción.
- KPI de logística: rotación de inventario, stock de producto, margen entre coste de materia prima y total de ventas, errores de facturación.
- KPI de ventas: Posicionamiento en el mercado, el índice de fidelización, el volumen de compra por cliente.

Capítulo 3. Estado del arte

Capítulo 3. Estado del arte

En este capítulo se presentan diversos trabajos de investigación relacionados con el tema desarrollado. La búsqueda se centró en publicaciones en revistas y congresos científicos, con un enfoque en la aplicación de minería de datos en procesos de negocio con el objetivo de mejorar los indicadores clave de desempeño (KPI).

Para estructurar el análisis de estos trabajos en el estado del arte, se consideraron los siguientes aspectos descriptivos:

- Objetivo del trabajo: Describe la finalidad de cada investigación, ya sea enfocada en la mejora de los indicadores clave de desempeño o en la optimización de los procesos de negocio.

- Caso de estudio: Destaca la versatilidad de la aplicación de minería de datos en diferentes procesos de negocio, mostrando su adaptabilidad a diversos contextos organizacionales.

- Proceso de extracción de datos: Detalla los pasos seguidos para extraer conocimiento significativo de los casos de estudio analizados.

- Herramientas y técnicas utilizadas: Identifica los softwares y métodos empleados durante el proceso de extracción de datos, lo que proporciona un panorama de las soluciones tecnológicas aplicadas.

Los trabajos relacionados con esté proyecto de investigación han sido clasificados en tres categorías, de acuerdo al objetivo de esta tesis, las cuales son:

Aplicación de minería de datos en procesos para la mejora de operaciones empresariales

En esta categoría se analizan trabajos de investigación enfocados en la mejora de procesos de negocio mediante la aplicación de técnicas de minería de datos. Estos estudios buscan optimizar las operaciones empresariales, identificando patrones y áreas de oportunidad que permitan aumentar la eficiencia y eficacia de los procesos organizacionales.

Aplicación de minería de datos en procesos para la mejora de indicadores clave de desempeño

Esta categoría incluye investigaciones desarrolladas con el propósito de mejorar los indicadores clave de desempeño (KPI) a través de la minería de datos. Los trabajos analizan cómo la extracción de conocimiento y la identificación de relaciones significativas en los datos contribuyen a evaluar y optimizar el rendimiento de los procesos de negocio.

Aplicación de minería de datos en procesos a partir del procesamiento del registro de eventos

En esta categoría se consideran trabajos de investigación que aplican minería de datos a partir del procesamiento de registros de eventos. Estos estudios se enfocan en la utilización de diferentes formatos de registros para extraer información útil, modelar procesos de negocio y descubrir patrones que permitan mejorar la toma de decisiones y el rendimiento organizacional.

3.1 Aplicación de minería de datos en procesos para la mejora de operaciones empresariales

3.1.1 Una metodología para la extracción y análisis integrados de procesos y datos hacia la mejora de procesos basada en evidencia

En este trabajo de investigación, Delgado et al. (2021) proponen la metodología PRICED (Process and Data Science for Organizational Improvement) para implementar esfuerzos de minería y análisis de procesos y datos en organizaciones con el objetivo de mejorar sus operaciones basándose en evidencia empírica. La metodología PRICED se centra en la integración de datos de procesos y datos organizacionales en una vista unificada, lo que permite aplicar técnicas de minería de procesos y minería de datos en un mismo conjunto de datos integrados. Esto proporciona a las organizaciones una visión completa y coherente para tomar decisiones fundamentadas. La metodología se organiza en varias disciplinas que abarcan:

- La extracción e integración de datos
- La mejora de la calidad de los datos
- La preparación de datos
- La minería y el análisis de datos
- El cumplimiento de procesos

Cada disciplina define roles específicos y utiliza artefactos concretos para garantizar una implementación estructurada. Además, PRICED se divide en fases principales, como la Fase de Datos y la Fase de Minería/Análisis, cuyo objetivo principal es ofrecer a las organizaciones inteligencia empresarial basada en evidencia para mejorar sus operaciones diarias. Como caso de estudio, la metodología se aplicó al proceso de movilidad estudiantil en la Universidad de la República de Montevideo. Los resultados demostraron que PRICED

es capaz de identificar problemáticas en la implementación de procesos, proporcionando información clave para la mejora de las operaciones.

3.1.2 Modelo para optimizar el proceso de gestión de negocio combinando minería de procesos con inteligencia de negocios desde almacenes de datos

En este trabajo de investigación, Giraldo, Jiménez y Tabares (2017) presentan un modelo de optimización de procesos de negocio basado en una revisión de la literatura y en iteraciones sobre un proceso específico, integrando BPM (Gestión de Procesos de Negocio), minería de procesos y modelado dimensional. El modelo está compuesto por cuatro componentes, los cuales se alinean con el ciclo BPM

- Modelado de la organización: Relacionado con las etapas de Diseño y Modelamiento del ciclo BPM, este componente especifica las características de la organización, sus procesos, actividades y tareas.

- Análisis de procesos: Vinculado a las etapas de Ejecución y Monitoreo, se encarga de monitorear y evaluar las tareas, actividades, procesos e indicadores de desempeño de la organización.

- Origen de datos: Relacionado con la etapa de Ejecución, este componente determina el conjunto de registros compuesto por tareas y variables obtenidas de los sistemas de información de la organización. En esta etapa se aplica minería de procesos.

- Optimización: Asociado con la etapa de Optimización, consiste en analizar los resultados obtenidos y proponer planes de mejora para optimizar los procesos.

El modelo fue aplicado a un caso de estudio, y los resultados demostraron la eficacia de integrar un modelo dimensional para descubrir los registros de eventos de los procesos. Asimismo, se logró identificar el estado de los procesos en función del comportamiento de las actividades dentro del modelo organizacional, facilitando así la toma de decisiones y la propuesta de mejoras.

3.2 Aplicación de minería de datos en procesos para la mejora de indicadores clave de desempeño

3.2.1 Hacia la gestión de indicadores clave de rendimiento para medir el rendimiento de los procesos empresariales

En este trabajo de investigación, Elhadjamor, E. A., y Ghannouchi, S. A. (2017) proponen un enfoque metodológico para la gestión de los Indicadores Clave de Rendimiento (KPI) en el contexto de la medición del rendimiento de los procesos de negocio, con un énfasis particular en el sector de la atención médica. Este enfoque incluye la recopilación de datos cuantitativos y cualitativos relacionados con los procesos de atención médica, la creación de una ontología para representar los KPI y sus relaciones con las actividades del proceso, y la aplicación de técnicas de minería de datos, específicamente el algoritmo Apriori, para identificar relaciones y patrones significativos en los datos.

Para desarrollar este estudio, se emplearon herramientas tecnológicas como jBPM para la gestión de procesos y PostgreSQL como sistema de gestión de bases de datos. El análisis de datos permitió descubrir una relación significativa entre el tiempo de espera antes del tratamiento por parte del personal médico y el tiempo de espera antes del tratamiento por parte del personal paramédico. Estos hallazgos proporcionaron información valiosa que fue utilizada para tomar decisiones informadas, destinadas a mejorar los procesos de atención médica y, en última instancia, aumentar la satisfacción de los pacientes.

3.2.2 Aplicación de técnicas de minería de datos para identificar indicadores clave de rendimiento relevantes

En este trabajo de investigación, Peral, J., Maté, A., y Marco, M. (2017) proponen una metodología para identificar Indicadores Clave de Rendimiento (KPIs) utilizando técnicas de minería de datos. Los KPIs son esenciales para evaluar el desempeño de una organización y tomar decisiones informadas. El artículo se enfoca en cómo identificar KPIs relevantes en contextos donde no existen listas predefinidas, lo cual es común al adoptar nuevas actividades o explorar fuentes de datos desconocidas, como el Big Data**.**
La metodología propuesta consta de seis etapas:

1. Definición y modelado de la estrategia empresarial, incluyendo la identificación de objetivos comerciales y KPIs.
2. Análisis de los requisitos de los usuarios y partes interesadas.
3. Diseño del almacén de datos.

4. Definición y selección de KPIs candidatos, junto con la creación de un modelo multidimensional.
5. Aplicación de técnicas de minería de datos y visualización para analizar los KPIs y sus relaciones.
6. Incorporación de los KPIs en paneles de control para el monitoreo continuo del desempeño y la toma de decisiones informadas.

El artículo presenta un caso de estudio donde se aplican estas etapas a datos de MOOCs (Cursos Masivos Abiertos en Línea) y datos abiertos de la Universidad de Alicante. En el caso de estudio, se analizan datos de estudiantes utilizando técnicas de minería de datos como Support Vector Machines, Random Forest, y redes neuronales, para identificar KPIs relacionados con el rendimiento estudiantil. Los resultados revelan que factores como el número de créditos matriculados y la ubicación geográfica de los estudiantes influyen en los resultados académicos.

3.3 Aplicación de minería de datos en procesos a partir del procesamiento de registros de eventos

3.3.1 Uso de minería de procesos y análisis de registros de eventos para una mejor toma de decisiones sobre estrategias comerciales

En este trabajo de investigación, Midden, J. H. (2021) propone métodos para utilizar la minería de procesos y el análisis de registros de eventos con el fin de asistir a los tomadores de decisiones en el entorno empresarial, reduciendo las opciones consideradas y mejorando el rendimiento general de la empresa. Además, señalan que la toma de decisiones estratégicas en el ámbito de los negocios puede resultar desafiante debido a restricciones de tiempo y recursos. La minería de procesos se emplea comúnmente en el análisis y la gestión de procesos empresariales, aunque suele enfocarse principalmente en asuntos de rendimiento y cumplimiento normativo.

Dentro de los métodos propuestos por los autores se incluye la identificación de indicadores clave de desempeño (KPI) relevantes para la logística empresarial, el análisis de registros de eventos, la aplicación de técnicas de minería de datos orientadas a los procesos comerciales, la evaluación de la conformidad y el rendimiento de los procesos, así como la utilización de métodos de eliminación de escenarios.

En el caso de estudio presentado, se llevaron a cabo tres experimentos para evaluar diferentes escenarios comerciales a lo largo del tiempo. Para ello, se emplearon herramientas como Disco, utilizada para analizar registros de eventos y crear modelos de procesos, y Python,

que permitió el procesamiento de datos y la ejecución de los experimentos. Los datos utilizados provenían de registros de eventos recopilados en una fábrica y las pruebas se desarrollaron en tres etapas, evaluando diversos escenarios comerciales en cada una de ellas. Los resultados de estos experimentos indicaron que los escenarios de mejor rendimiento convergieron hacia patrones similares, lo que demuestra que el análisis de registros de eventos puede ser una herramienta útil para apoyar la toma de decisiones en el ámbito empresarial.

3.3.2 Minería de procesos comerciales a partir de registros web de eventos de comercio electrónico: verificación de conformidad e identificación de cuellos de botella

En este trabajo de investigación, Siek, M., & Mukti, R. M. G. (2021) aplican técnicas de minería de procesos en el ámbito de la industria del comercio electrónico. El objetivo principal del artículo es llevar a cabo un análisis de los datos de registros de eventos de una empresa de comercio electrónico mediante el empleo de algoritmos de minería de procesos, tales como el Inductive Miner y el Fuzzy Miner. A través de este análisis, se generan modelos de procesos específicos para el comercio electrónico y se verifica su conformidad con los procesos estandarizados. El artículo también identifica problemáticas particulares que afectan a la industria del comercio electrónico, como la integración de sistemas, la realización de procesos manuales, la verificación manual de pagos y la cancelación de pedidos, especialmente en el caso de pagos contra reembolso. Para llevar a cabo este análisis y la modelización de los procesos comerciales a partir de los datos de eventos, se emplean herramientas como Disco, Pro-M y RapidMiner Pro-M. En resumen, el artículo subraya la utilidad de la minería de procesos en la resolución de problemas específicos en la industria del comercio electrónico, lo que permite llevar a cabo mejoras continuas en los procesos empresariales. Estas mejoras incluyen la automatización de la auditoría en tiempo real y la reingeniería de procesos.

3.3.3 Análisis en profundidad los procesos empresariales sanitarios y los indicadores clave de rendimiento mediante la minería de procesos.

En este trabajo de investigación, Elhadjamor, E. (2019) documenta el procedimiento para la extracción de conocimiento mediante un software BPMS y la minería de indicadores clave de desempeño, con el objetivo de mejorar el proceso de atención médica en el departamento de emergencias de un hospital. El estudio registra cien eventos de la ejecución del proceso en el software jBPM y utiliza ProM para llevar a cabo la minería de procesos.

El trabajo destaca que existen diversas formas de explotar las herramientas de los softwares para mejorar la obtención e interpretación de la información, optimizando tanto los procesos como los indicadores clave de desempeño. Este análisis permite facilitar la evaluación del desempeño de los procesos al responsable de la toma de decisiones, proporcionando información para tomar acciones orientadas a su mejora. Los resultados de esta investigación muestran que el análisis de los indicadores de desempeño puede mejorar el proceso de negocio desde tres perspectivas principales: organización, comportamiento e información. Por ejemplo, estas mejoras pueden incluir la asignación adecuada de tareas a las personas que las ejecutan, el orden en que deben cumplirse las tareas o los datos de entrada que se tomarán en cuenta para tareas específicas.

3.3.4 Extracción de conocimiento útil de registros de eventos: Un enfoque de minería de conjuntos de elementos frecuentes

En este trabajo de investigación, Djenouri, Belhadi, y Fournier-Viger (2018) presentan un método innovador con un enfoque mejorado de análisis de registro de eventos denominado AllMining. Este método permite construir diversas bases de datos de transacciones a partir de un mismo registro de eventos, aplicando transformaciones en los datos para identificar patrones y correlaciones en los procesos de negocio mediante técnicas de Frequent Itemset Mining (FIM). La herramienta AllMining se compone de tres etapas principales:

1. Preprocesamiento: En esta etapa, el registro de eventos se transforma en una forma adecuada para los análisis, lo cual incluye dos fases:

- Creación de registros categóricos a partir del registro de eventos.
- Generación de bases de datos transaccionales basadas en el registro de eventos.

2. Minería de conjuntos de datos frecuentes: Se aplican técnicas de minería de conjuntos frecuentes sobre las bases de datos generadas, con el objetivo de extraer conjuntos de elementos frecuentes.

3. Descubrimiento de conocimiento: En esta última etapa, se utiliza una estrategia de poda heurística diseñada para maximizar el número de registros cubiertos por un mínimo número de conjuntos de elementos frecuentes. Esto permite preservar una gran cantidad de información relevante sobre los procesos de negocio, al mismo tiempo que reduce significativamente el número de patrones finales.

Los autores concluyen que la herramienta AllMining proporciona un enfoque proactivo para el descubrimiento de conocimiento en registros de eventos. Este enfoque facilita investigaciones adicionales, como la identificación de mejores prácticas o el rediseño de procesos, impulsando así mejoras en los procesos de negocio.

3.3.5 Extracción de actividades de procesos empresariales de los registros de correo electrónico

En este trabajo de investigación, Jlailaty, D., Grigori, D., y Belhajjame, K. (2017) desarrollan un marco de trabajo para extraer información relacionada con procesos comerciales a partir de registros de correos electrónicos, con una intervención mínima del usuario. Los correos electrónicos representan una fuente valiosa de datos sobre procesos comerciales, ya que a menudo reflejan actividades complejas en contextos tanto profesionales como personales.

El estudio propone un enfoque de aprendizaje automático no supervisado, empleando técnicas como clustering jerárquico y Word2vec para agrupar y etiquetar correos electrónicos en función de modelos de procesos comerciales y las actividades relacionadas. En el marco se compararon dos métodos para medir similitud textual: Latent Semantic Analysis (LSA) y Word2vec. Los resultados indicaron que Word2vec es más eficaz en la detección de similitudes textuales. Los correos electrónicos se organizaron según su similitud y se subdividieron en subgrupos que representan diferentes tipos de actividades comerciales. Estos subgrupos fueron etiquetados mediante n-gramas, permitiendo a los usuarios seleccionar las etiquetas más adecuadas para cada actividad. El artículo presenta métricas de calidad, como el índice Rand, precisión, recall y pureza, para evaluar la eficacia del proceso de clustering y sub-clustering. Los resultados mostraron que Word2vec superó a LSA en la mayoría de las métricas, respaldando su utilidad en este contexto.

Como caso de estudio, se aplicó el marco a datos de correos electrónicos de un estudiante de doctorado, identificando procesos comerciales y etiquetando las actividades relacionadas. Finalmente, el artículo plantea posibles direcciones para investigaciones futuras, como la identificación de procesos comerciales completos y la implementación de sistemas de recomendación para evitar repeticiones en los correos electrónicos.

3.4 Conclusiones del estado del arte

Los estudios revisados demuestran cómo la integración de la minería de datos y de procesos en diferentes contextos empresariales puede ofrecer mejoras significativas en la gestión y optimización de procesos, así como en la identificación de indicadores clave de desempeño (KPIs). En general, estas investigaciones han mostrado que el análisis de registros de eventos, la aplicación de modelos predictivos y las técnicas de visualización de datos permiten identificar cuellos de botella, evaluar el cumplimiento de los procesos y descubrir patrones ocultos que afectan el rendimiento organizacional.

Además, el uso de herramientas especializadas como ProM, Disco, jBPM y enfoques como el clustering, el análisis de conjuntos de elementos frecuentes y redes neuronales ha facilitado la extracción de conocimiento útil en dominios diversos como el comercio electrónico, la educación y la atención médica. Los resultados evidencian que estos métodos contribuyen no solo a una toma de decisiones más informada, sino también a la mejora continua de procesos desde múltiples perspectivas: organizacional, comportamental e informativa.

En conjunto, estos trabajos subrayan la relevancia de emplear un enfoque metodológico basado en la evidencia para la gestión de procesos empresariales, destacando el potencial de estas tecnologías para transformar datos complejos en información valiosa para la optimización de estrategias y recursos.

Capítulo 4: Metodología para la mejora de indicadores clave de desempeño mediante minería de datos

Capítulo 4. Metodología para la mejora de indicadores clave de desempeño mediante minería de datos

Este capítulo presenta la metodología propuesta para la mejora de indicadores clave de desempeño. En este se describen las fases, los métodos y tareas que deben aplicarse para llevar a cabo esta metodología.

4.1 Descripción de la metodología

La metodología, mostrada en la Figura 2, proporciona un marco estructurado que guía desde la comprensión del negocio y los datos hasta la interpretación de resultados obtenidos. Este enfoque asegura que los resultados obtenidos sean útiles y aplicables, permitiendo a las organizaciones tomar decisiones fundamentadas y alcanzar mejoras significativas en sus procesos.

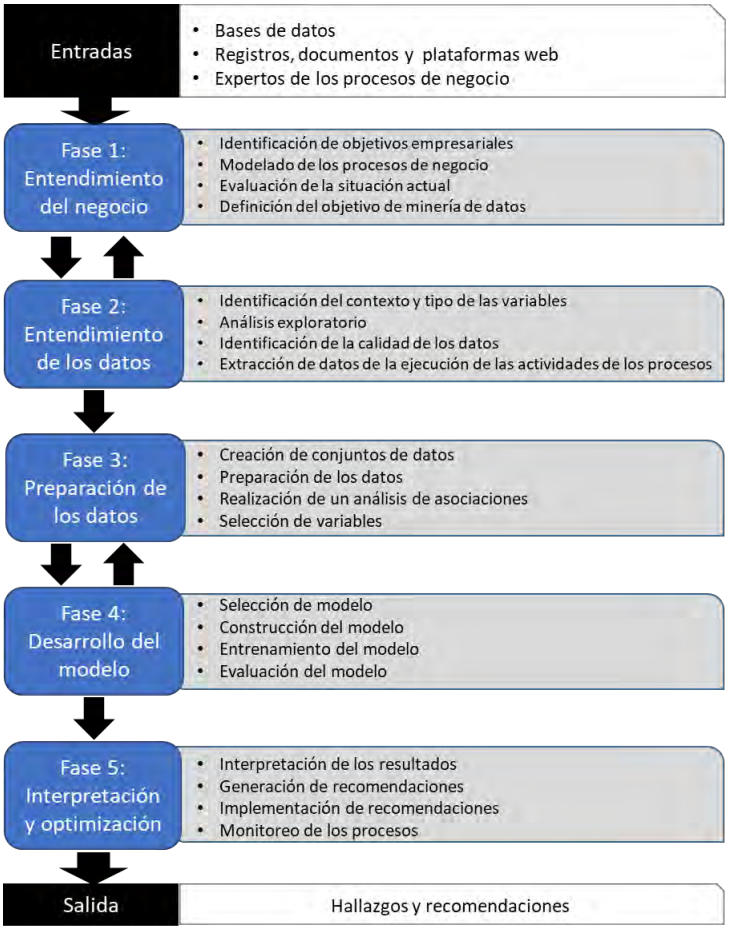


Figura 2 Flujo de la metodología para la mejora de indicadores clave de desempeño mediante minería de datos

Entradas

Las entradas en esta metodología se refieren a la información que se requiere para desarrollar este proceso. Estas entradas pueden ser:

- Bases de datos: Las bases de datos son repositorios de información de los procesos de negocio.

- Registros, documentos y plataformas web: Los registros, documentos y plataformas web son recursos que no almacenan información de los procesos, sino que son medios con los cuales se puede darle un seguimiento a las actividades de los procesos de negocio.

- Expertos en los procesos de negocio: Los expertos en los procesos de negocio se refiere a las personas que tienen un amplio conocimiento sobre los procesos a analizar. Un experto en procesos identifica los roles participantes, las actividades que desempeña cada rol, los recursos que intervienen en los procesos y la información que se recopila durante la ejecución de los procesos de negocio.

Fases

Fase 1. Entendimiento del negocio

La Fase 1 tiene como objetivo comprender el contexto organizacional y establecer una base sólida para el análisis de minería de datos. Esta fase incluye la identificación de objetivos empresariales, los cuales se alinean con los indicadores clave de desempeño a mejorar; el modelado de los procesos de negocio, que permite estructurar actividades, roles, recursos y puntos de decisión críticos; y la evaluación de la situación actual, que analiza los procesos operativos, recursos disponibles y limitaciones. Finalmente, se define el objetivo del análisis de minería de datos, especificando la variable principal a estudiar y delimitando los alcances del proyecto para garantizar su relevancia y viabilidad. Esta fase asegura que el análisis esté alineado con las metas estratégicas de la organización.

Fase 2. Entendimiento de los datos

La Fase 2 se centra en garantizar que los datos utilizados en el análisis sean adecuados y relevantes para los objetivos del negocio. Primero, se identifican y clasifican las variables según su contexto y tipo de dato, facilitando la selección de técnicas de análisis apropiadas. Luego, se realiza un análisis exploratorio para examinar distribuciones y patrones preliminares, plantear hipótesis y determinar la suficiencia de los datos disponibles.

Posteriormente, se evalúa la calidad de los datos, identificando problemas como datos faltantes, inconsistencias y duplicados. Finalmente, se extraen datos relevantes de diversas fuentes, como sistemas informáticos y bases de datos, utilizando herramientas específicas como plataformas web, bases relacionales y daemons, asegurando que toda la información necesaria esté disponible para su análisis. Esta fase asegura que los datos sean precisos, completos y útiles para las etapas posteriores del proceso de minería.

Fase 3. Procesamiento de los datos

La Fase 3 se centra en preparar y estructurar los datos para su uso en modelos de minería de datos. Primero, se crean conjuntos de datos organizados que combinan variables de diversas fuentes. Luego, se realiza la preparación de los datos mediante limpieza, transformación, y balanceo de clases, asegurando calidad y relevancia. La limpieza aborda problemas como datos faltantes o duplicados, mientras que la transformación incluye normalización, codificación y discretización para adaptarse a los requisitos de los algoritmos. Además, se lleva a cabo un análisis de asociaciones, utilizando métodos como la correlación de Pearson y la prueba Chi-cuadrado, para identificar variables significativamente relacionadas con la variable objetivo. Finalmente, se seleccionan las variables más relevantes para optimizar la precisión del modelo, eliminando datos redundantes o irrelevantes. Esta fase garantiza que los datos sean aptos para análisis efectivos y confiables.

Fase 4. Desarrollo del modelo

La Fase 4 se centra en la construcción, entrenamiento y evaluación de un modelo de minería de datos. Primero, se selecciona el tipo de modelo más adecuado según los objetivos del proyecto, como clasificación, predicción o agrupación. Luego, los datos se dividen en conjuntos de entrenamiento y prueba, para optimizar y evaluar el rendimiento del modelo. Durante el entrenamiento, el modelo ajusta sus parámetros para identificar patrones en los datos, permitiendo realizar predicciones o clasificaciones en nuevos escenarios. Finalmente, se evalúa el modelo utilizando métricas como precisión, error y matriz de confusión. Este proceso es iterativo, permitiendo realizar ajustes o regresar a fases anteriores si los resultados no son satisfactorios, asegurando así un desempeño óptimo.

Fase 5 Interpretación y optimización

La Fase 5. se enfoca en transformar los resultados del modelo en acciones prácticas para mejorar los procesos de negocio. Primero, se interpretan los resultados, destacando patrones y relaciones significativas que puedan orientar decisiones estratégicas. Posteriormente, se generan recomendaciones específicas alineadas con los objetivos empresariales, orientadas a optimizar operaciones y resolver ineficiencias. Estas recomendaciones son implementadas en los procesos, mediante ajustes operativos, modificaciones estratégicas o nuevas políticas.

Finalmente, se monitorean los procesos para evaluar el impacto de las mejoras y realizar ajustes adicionales si es necesario, asegurando resultados sostenibles y optimizados.

Salida

Hallazgos y recomendaciones

Los hallazgos encontrados son los patrones, tendencias, relaciones y conocimientos obtenidos después de haber aplicado la metodología propuesta. Estos pueden incluir correlaciones entre variables, agrupaciones de datos similares, asociaciones frecuentes, predicciones, anomalías o cualquier información útil que ayude a comprender mejor los procesos analizados o a tomar decisiones informadas

Por otro lado, las recomendaciones son las sugerencias prácticas derivadas de los hallazgos, orientadas a mejorar procesos, optimizar operaciones o tomar acciones estratégicas. Estas recomendaciones se basan en la interpretación de los patrones detectados y pueden incluir propuestas como ajustar procesos operativos o actividades, enfocarse en áreas específicas de mejora, implementar cambios en estrategias, o monitorear ciertos indicadores clave de desempeño para mejorar los resultados

4.2 Fase 1. Entendimiento del negocio

4.2.1 Identificación de objetivos empresariales

La identificación de objetivos empresariales consiste en definir claramente las metas estratégicas que la empresa busca alcanzar mediante la minería de datos. Estas metas deben estar estrechamente relacionadas con los indicadores clave de desempeño que se quieren mejorar, lo cual es fundamental en la aplicación de metodologías como CRISP-DM, ampliamente utilizadas en diversos contextos empresariales por su enfoque sistemático y replicable (Espinosa Zúñiga, 2020; Schröer, Kruse, & Gómez, 2021).

4.2.2 Modelado de los procesos de negocio

El modelado de los procesos de negocio se lleva a cabo con el objetivo de identificar y estructurar los elementos fundamentales que intervienen en su ejecución. Estos elementos incluyen (Dijkman, Dumas, & Ouyang, 2008; Recker, 2010):

- Actividades clave: Se refiere a las actividades principales dentro de los procesos de negocio, de las cuales es posible extraer datos relevantes para el análisis y evaluación del desempeño.

- Actores: Son las personas o roles que ejecutan las actividades y participan activamente en los procesos de negocio, desempeñando tareas específicas dentro del flujo de trabajo.
- Recursos: Incluyen los elementos necesarios, tales como materiales, herramientas, equipos o información, que se utilizan para llevar a cabo las actividades dentro de los procesos de negocio.
- Puntos de decisión: Son los momentos clave dentro del proceso en los que se toman decisiones importantes, afectando el flujo o la dirección del proceso y determinando las siguientes actividades o pasos a seguir.
- Indicadores clave de desempeño (KPI): Constituyen métricas específicas utilizadas para medir el rendimiento de los procesos y actividades, evaluando si se cumplen los objetivos establecidos y ayudando a identificar áreas de mejora.

Para llevar a cabo el modelado en esta etapa, es posible emplear herramientas de modelado como diagramas de flujo o Notación BPM. Este paso es importante debido a que facilita el siguiente paso, Evaluación de la situación actual.

4.2.3 Evaluación de la situación actual

La evaluación de la situación actual consiste en realizar un análisis detallado de la situación actual de la empresa, revisando los procesos operativos, los recursos disponibles, los datos existentes y la situación actual de los indicadores clave a analizar. Esta evaluación permite identificar qué información es relevante y qué posibles limitaciones o áreas de mejora existen para iniciar el proceso de minería de datos actividades (IBM, 2021). En esta etapa se plantean las siguientes preguntas:

- ¿Quiénes ejecutan las actividades del proceso de negocio?
- ¿Cuál es la situación actual del indicador clave de desempeño?
- ¿En qué entorno se ejecuta el proceso de negocio?
- ¿En qué actividades se almacenan los datos?
- ¿Cómo se almacenan los datos?

4.2.4 Definición el objetivo de la minería de datos

Esta etapa consiste en definir claramente lo que se espera lograr con la minería de datos. Se concreta una meta específica del análisis, definiendo cual es la variable objetivo y delimitando los alcances del proyecto para asegurar su relevancia y viabilidad (IBM, 2021; Ncr et al., 1999).

4.3 Fase 2. Entendimiento de los datos

4.3.1 Identificación del contexto y tipo de las variables

Esta etapa consiste en comprender el significado de cada variable en el conjunto de datos dentro del contexto del negocio. Es fundamental conocer qué representa cada variable y cómo se relaciona con el problema que se busca resolver. Este entendimiento permite determinar la utilidad de cada variable en el análisis y evitar interpretaciones erróneas. Además, se realiza una clasificación de las variables según su tipo de dato, como numérico, categórico o booleano. Esta clasificación es crucial para seleccionar las técnicas de análisis y transformaciones adecuadas, ya que muchos algoritmos y métodos de minería de datos requieren formatos específicos para funcionar correctamente. Para facilitar esta etapa, se recomienda la creación de diccionarios de variables, que incluyan el significado, tipo de dato y posibles valores de cada variable (IBM, 2021; Ncr et al., 1999).

4.3.2 Análisis exploratorio

Esta etapa consiste en examinar los datos de manera preliminar, identificar como está distribuida la base de datos e identificar patrones preliminares, distribuciones y relaciones entre las variables. Dentro de este paso se pueden plantear y dar respuesta a las siguientes preguntas (IBM, 2021):

- ¿Qué hipótesis sobre los datos se han formulado?
- ¿Cuáles son las variables que parecen prometedoras para el análisis?
- ¿Esta exploración ha cambiado su objetivo de minería de datos?
- ¿Se cuenta con los datos necesarios para realizar el análisis?
- ¿Existe manera de extraer más información de las actividades?

4.3.3 Identificación de la calidad de los datos

Esta etapa consiste en identificar datos faltantes, duplicados, inconsistencias o errores en la entrada de información. Si se encuentran problemas, pueden necesitarse procesos de limpieza o corrección de los datos antes de continuar con el análisis. Dentro de los errores más comunes se encuentran (Nuñez-Arcia, Díaz de la Paz, & García-Mendoza, 2016):

Ausencia de datos en un campo no nulo

Ocurre cuando un campo que debe contener información no tiene un valor válido. En algunos casos, se rellenan con valores ficticios o genéricos, como "N/A" o "0", lo que puede afectar el análisis. Ejemplo: En una base de datos de clientes, el campo "Fecha de nacimiento" puede contener "01/01/1900", un valor que no representa una fecha real.

Datos erróneos

Se refiere a datos que son técnicamente válidos en términos de formato, pero no reflejan la entidad o información real que representan. Ejemplo: Un registro que muestra el precio de un producto como "1000" cuando debería ser "100" debido a un error de entrada.

Errores ortográficos

Palabras mal escritas en los campos de las fuentes de datos, lo que puede dificultar la interpretación o el análisis automatizado. Ejemplo: En un campo de "ciudad", se encuentra "Guadajara" en lugar de "Guadalajara".

Valores implícitos

Aparecen cuando existen datos extraños o no estándar en un campo, que no representan valores explícitos o esperados. Ejemplo: En un campo de "estatus de pedido", encontrar un valor como "P" que no tiene un significado claro sin un diccionario de datos.

Valores de campos perdidos

Datos que son almacenados en el campo equivocado, lo que puede generar confusión o errores en los análisis. Ejemplo: En un registro de empleados, el número de teléfono aparece en el campo de "dirección".

Datos ambiguos

Datos que pueden ser interpretados de más de una manera, dependiendo del contexto, lo que genera incertidumbre en el análisis. Ejemplo: En un campo de "categoría", se encuentra el valor "General", que no indica claramente a qué se refiere.

4.3.4 Extracción de datos de la ejecución de las actividades de los procesos

La extracción de datos consiste en recoger los datos que provienen de la operación diaria de la empresa o de sus procesos de negocio. La extracción asegura que la información relevante de cada proceso esté disponible para ser analizada, sin embargo, este paso puede omitirse si se cuenta con la información necesaria para realizar el análisis. Existen distintos métodos

para la extracción de información de los procesos de negocio, como lo son el monitoreo de actividades comerciales, plataformas web, bases de datos relacionales y daemons.

Monitoreo de actividades comerciales

El monitoreo de actividades comerciales (Business Activity Monitoring, BAM) es un concepto que se les otorga a las herramientas de software que realicen la recopilación, organización, análisis y visualización de datos obtenidos en tiempo real, acerca de las actividades ejecutadas en un proceso de negocio. Algunos de ellos son Business Monitor de IBM, Business Activity Monitoring de Oracle y Business Factor de TIBCO Software Inc (Galvis Lista & González Zabala, 2014).

Plataformas web

Las plataformas web son ambientes en los que actualmente se desarrollan los procesos de negocio, proporcionando accesibilidad, escalabilidad y centralización de las operaciones. Estas plataformas suelen estar diseñadas bajo la arquitectura MVC (Modelo-Vista-Controlador), que permite una separación clara entre la lógica de negocio, la interfaz de usuario y el control del flujo de datos (Hernández, Torres, & México, 2019). El Modelo maneja los datos y las reglas de negocio, la Vista se encarga de presentar la información al usuario, y el Controlador actúa como intermediario, gestionando las interacciones y actualizando el modelo o la vista según sea necesario. Este enfoque modular facilita el desarrollo, mantenimiento y escalabilidad de las plataformas, permitiendo adaptarlas rápidamente a cambios en los procesos de negocio. Por ello, las plataformas web, con arquitecturas robustas como MVC, se han consolidado como herramientas clave en la digitalización y automatización empresarial (González & Romero, 2012).

Base de datos relacional

Una base de datos relacional utiliza un grupo de tablas de tablas para representar los datos y relaciones entre las tablas y las variables. Cada tabla está compuesta por varias columnas, y cada columna tiene un nombre único. Se destacan por usar claves primarias para identificar registros únicos y claves foráneas para relacionar tablas, lo que permite establecer relaciones como uno a uno, uno a muchos o muchos a muchos. Estas bases utilizan SQL (Structured Query Language) para gestionar y consultar datos de forma eficiente. Son comunmente utilizadas debido a su estructura clara, flexibilidad para expandirse y capacidad de mantener la integridad de los datos (Silberschatz, A., 2007).

Daemon

Un daemon es un programa que se ejecuta en segundo plano en un sistema operativo. Estos tienen distintos propósitos; sin embargo, pueden programarse para extraer información de las actividades de los procesos de negocio (Orlando & Giovanni, 2008; Daoudagh et al., 2023)

4.4 Fase 3. Procesamiento de los datos.

4.4.1 Creación de conjuntos de datos

Los datos se agrupan y organizan en conjuntos de datos específicos que serán utilizados en el análisis. Esto implica combinar datos y variables de distintas fuentes y asegurar que toda la información necesaria para el análisis esté disponible en un único conjunto.

4.4.2 Preparación de los datos

La preparación de datos abarca técnicas enfocadas en analizar datos en bruto para obtener datos de calidad. Esto incluye principalmente la recolección de datos, la integración de datos, la transformación de datos, la limpieza de datos, la reducción de datos y la discretización de datos. La preparación de datos es importante porque asegura que los datos utilizados en el análisis sean precisos, coherentes y relevantes, lo cual es fundamental para obtener resultados fiables y valiosos en los procesos de minería de datos y aprendizaje automático (Zhang, S., Zhang, C., & Yang, Q. 2003). Dentro de la preparación de los datos, existen tres pasos fundamentales que deben llevarse a cabo para garantizar la calidad y la utilidad del conjunto de datos en el análisis: limpieza de datos, transformación de datos y balanceo de datos.

I. Limpieza de datos

La limpieza de datos consiste en detectar y corregir errores e incoherencias para mejorar la calidad de los datos. Este proceso es crucial para garantizar datos precisos y coherentes eliminando duplicados y consolidando información (Rahm & Do, 2000). Las técnicas de limpieza de datos van a depender de cuales sean los errores que se encuentran en el conjunto de datos y la implementación de estas depende mucho del analista. Las técnicas más comunes son rellenar espacios vacíos, eliminar errores y/o corregir errores.

Rellenar espacios vacíos

La técnica de rellenar se aplica principalmente a los campos que tienen ausencia de datos o valores nulos. Esta técnica se puede implementar utilizando la moda, media, mediana, interpolación o regresión lineal (Dara, Satyanarayana, & Govardhan, 2016).

Eliminación

La eliminación se aplica principalmente a valores incorrectos, duplicados o irrelevantes de la base de datos. Es común que cuando se aplica esta técnica se elimina toda la variable (columna) o toda la instancia (fila) de la tabla (Lup Low, Li Lee, & Wang Ling, 2001).

Corrección

La corrección de los datos implica tratar errores ortográficos, valores implícitos, valores con campos perdidos y ambigüedad en los datos. Estos pueden corregirse manualmente, sin embargo, existen rutinas, códigos y software que facilitan esta tarea (Rengifo, 2007).

II. Transformación de los datos

La transformación de datos es un enfoque que consiste en convertir los datos originales a un formato preferido para la entrada de ciertos algoritmos de aprendizaje antes de su procesamiento (Zhang, S., Zhang, C., & Yang, Q. 2003). Las técnicas de transformación dependen del objetivo de la minería de datos y la calidad en que se encuentren los datos.

Normalización

La normalización de los datos consiste en escalar los valores de las variables para que se sitúen numéricamente en la misma escala (por ejemplo, entre 0 y 1) y, por tanto, tengan la misma importancia. Existen tres técnicas de normalización: Normalización de distribución Z, Normalización Min/Max, y Normalización por escala decimal (Javaheri, Sepehri, & Teimourpour, 2014).

Codificación

La codificación consiste en transformar datos categóricos en valores numéricos. para que los modelos de aprendizaje automático puedan procesarlas, comprenderlas y extraer información útil. Esto se realiza mediante técnicas como el one-hot encoding, que convierte cada categoría de una variable en una representación binaria, especialmente útil cuando no hay un orden natural entre las categorías (Dahouda & Joe, 2021).

Discretización

La discretización consiste en transformar datos numéricos en categorías al dividir un rango continuo de valores en intervalos independientes. Cada valor numérico se asigna a uno de estos intervalos, simplificando los datos al reducir un conjunto amplio de valores numéricos a un grupo más pequeño de categorías para facilitar su análisis (Ramírez-Gallego et al., 2016). En resumen, este método convierte los datos numéricos a categóricos.

III. Balanceo de clases

El balanceo de clases es una tarea que se realiza cuando existe un desequilibrio en un conjunto de datos, cuando una clase tiene muchas más instancias que otra. Esto puede causar que los modelos favorezcan la clase dominante, ignorando la minoritaria. Para solucionarlo, se emplean métodos como el sobremuestreo de la clase minoritaria (SMOTE) (Chawla, Bowyer, Hall, & Kegelmeyer, 2002). El método de Sobremuestreo de Minorías Sintéticas (Synthetic Minority Over-sampling TEchnique, SMOTE) es un enfoque de sobremuestreo que crea

muestras sintéticas de la clase minoritaria. Este método funciona generando nuevas instancias sintéticas a partir de las existentes, mediante la interpolación de las características de los vecinos más cercanos. SMOTE puede ofrecer un rendimiento superior al del sobremuestreo simple, que simplemente duplica las instancias de la clase minoritaria, ya que ayuda a reducir el sobreajuste y mejora la generalización del modelo. Debido a sus beneficios, SMOTE es ampliamente utilizado en situaciones donde los datos están desbalanceados, permitiendo un mejor equilibrio entre las clases y mejorando así la precisión del modelo predictivo (Blagus, R., & Lusa, L. 2013).

4.4.3 Realización de un análisis de asociaciones

En esta etapa se lleva a cabo un análisis de asociaciones para identificar cómo se relacionan las distintas variables del conjunto de datos, con el propósito de determinar cuáles tienen una relación significativa con la variable objetivo y descartar aquellas que podrían generar ruido en el análisis. El método seleccionado depende del tipo de datos: si las variables están representadas como valores numéricos, se utiliza un análisis de correlaciones de Pearson; mientras que, si las variables están representadas como valores categóricos, se aplica la prueba de Chi-cuadrado.

Análisis de correlación de Pearson

El análisis de correlación de Pearson (representado comúnmente por la letra R) es un método estadístico utilizado para determinar si existe una asociación lineal entre dos variables numéricas. Este coeficiente presenta dos aspectos importantes: su magnitud y su signo. La magnitud indica el grado de asociación entre las variables y toma valores entre 0 y 1. Una magnitud cercana a 0 sugiere que las variables no están asociadas, mientras que valores cercanos a 1 indican una fuerte asociación. Por otro lado, el signo del coeficiente señala la dirección de la asociación. Un signo positivo indica una relación proporcional, donde un aumento en los valores de una variable corresponde a un aumento en la otra. En cambio, un signo negativo refleja una relación inversamente proporcional, es decir, cuando los valores de una variable aumentan, los de la otra disminuyen (Camacho-Sandoval, 2008). La Tabla 2 muestra los grados de correlación.

Tabla 2 Interpretación de la magnitud del coeficiente de Pearson (R)

Rango de asociación de |R|	Interpretación
$0 \leq \|R\| < 0.10$	Asociación nula
$0.10 \leq \|R\| < 0.30$	Asociación débil
$0.30 \leq \|R\| < 0.50$	Asociación moderada
$0.50 \leq \|R\| < 1$	Asociación fuerte

Fuente: Lalinde et al., 2018

Prueba del Chi-cuadrado

El Chi-cuadrado es un estadístico que se utiliza para estimar la asociación entre variables categóricas, midiendo la discrepancia entre las frecuencias observadas (datos obtenidos) y las frecuencias esperadas (datos teóricos). Este análisis se realiza utilizando una tabla cruzada, que organiza las frecuencias de las variables categóricas en filas y columnas (Bangdiwala, 2016).

Si el valor de Chi-cuadrado es cercano a 0, se concluye que no existe asociación entre las variables. Sin embargo, un valor mayor indica una discrepancia significativa entre las frecuencias observadas y esperadas, lo que sugiere una posible relación entre las variables (Kearney, 2017). Determinar si esta relación es significativa requiere comparar el valor calculado de Chi-cuadrado con un valor crítico de una tabla estadística, considerando los grados de libertad y el nivel de significancia establecido (generalmente, p–valor ≤ 0.050) (Bangdiwala, 2016). A pesar de su utilidad, el Chi-cuadrado no mide la fuerza de la asociación entre las variables, por lo que este análisis se complementa con el estadístico V de Cramer, que cuantifica el grado de relación y proporciona una medida más clara de la asociación.

La V de Cramer es un estadístico que facilita la interpretación de las asociaciones encontradas con la prueba del Chi-cuadrado. Este estadístico toma valores de 0 a +1, entre más cercano a 0 más débil la asociación, en caso contrario entre más cercano a +1 el grado de asociación es más alto (Kearney, 2017).

Tabla 3 Interpretación del valor V de Cramer

Rango de asociación (V-Cramer)	Interpretación
0.00	No existe asociación
0 < V-Cramer < 0.10	Asociación débil
0.10 ≤ V-Cramer < 0.30	Asociación media
0.30 ≤ V-Cramer < 0.50	Asociación considerable
0.50 ≤ V-Cramer < 1	Asociación muy fuerte

Fuente: Betancourt Velásquez & Caviedes Niño, 2018

4.4.4 Selección de variables

Esta etapa se lleva a cabo después de realizar el análisis de asociaciones, ya que permite seleccionar los variables que están más asociados a la variable objetivo. Al elegir las variables relevantes, se introducen en el modelo, lo que ayuda a mejorar la precisión y la calidad del análisis. Esto es posible porque se eliminan las variables redundantes o irrelevantes, manteniendo únicamente aquellas que aportan valor significativo a los resultados. De esta manera, el modelo se optimiza y se hace más eficiente, utilizando solo la información más importante para el análisis (Mamián, Palta, Cobo, & Castillo, 2019).

4.5 Fase 4. Desarrollo del modelo

4.5.1 Selección de modelo

En este paso, se elige el tipo de modelo que mejor se ajusta al problema y los datos disponibles. Existen varios tipos de modelos, y la selección depende de los objetivos del proyecto (clasificación, predicción, agrupación, etc.) (Espinosa Zúñiga, 2020).

4.5.2 Construcción del modelo

La construcción del modelo comienza con la partición del conjunto de datos. entrenamiento (para que el modelo aprenda) y prueba (para evaluar su rendimiento). La partición de datos en conjuntos de entrenamiento, prueba y validación es una técnica esencial en el aprendizaje automático para evaluar y optimizar el rendimiento de los modelos predictivos). El conjunto de entrenamiento (generalmente 70-80% del total) se utiliza para ajustar los parámetros del modelo. Mientras que el conjunto de prueba, compuesto por datos que el modelo no ha visto previamente, permite evaluar su capacidad de generalización (Bai et al., 2020).

4.5.3 Entrenamiento del modelo

El entrenamiento del modelo es el proceso mediante el cual un algoritmo de aprendizaje automático ajusta sus parámetros utilizando el conjunto de datos de entrenamiento. Su objetivo es permitir que el modelo identifique patrones y relaciones en los datos, para posteriormente realizar predicciones, clasificaciones u otras tareas de manera efectiva con nuevos datos (Espinosa Zúñiga, 2020; Cortina, 2015; Elkabalawy et al., 2024).

4.5.4 Evaluación del modelo

La evaluación del modelo consiste en emplear el conjunto de datos de prueba para calcular métricas relevantes, como la precisión, el error y la matriz de confusión, entre otras, que permitan medir el desempeño del modelo (Espinosa Zúñiga, 2020; Cortina, 2015; Elkabalawy et al., 2024). Este proceso es de naturaleza iterativa (Kannengiesser & Gero, 2023), ya que, en caso de que los resultados obtenidos no sean satisfactorios, es posible regresar a la etapa de construcción del modelo o incluso a la Fase 3. Preparación de los datos, con el objetivo de realizar ajustes y optimizaciones que mejoren el rendimiento del modelo.

4.6 Fase 5. Interpretación y optimización

4.6.1 Interpretación de los resultados

En este paso se analizan y comprenden los resultados obtenidos del modelo, destacando los patrones, relaciones y conocimientos extraídos de los datos. Este análisis permite traducir los hallazgos en información útil para la toma de decisiones, asegurando que las conclusiones sean claras y aplicables al contexto del negocio (Espinosa Zúñiga, 2020; Cortina, 2015; Salcedo Parra et al., 2010).

4.6.2 Generación de recomendaciones

En este paso se formulan recomendaciones específicas basadas en la interpretación de los resultados. Estas pueden estar orientadas a optimizar procesos, mejorar decisiones o corregir ineficiencias. Las recomendaciones deben ser accionables y estar alineadas con los objetivos empresariales (Cortina, 2015; Salcedo Parra et al., 2010).

4.6.3 Implementación de recomendaciones

En esta etapa, las recomendaciones se llevan a la práctica, aplicándolas a los procesos de negocio. Puede implicar realizar ajustes en las operaciones, modificar estrategias o introducir nuevas políticas basadas en los hallazgos encontrados en el análisis de datos (Cortina, 2015; Salcedo Parra et al., 2010).

4.6.4 Monitoreo de los procesos

Después de implementar las recomendaciones, es esencial monitorear los procesos para evaluar el impacto de los cambios realizados. Esto asegura que las mejoras estén funcionando como se esperaba y permite hacer ajustes si es necesario para optimizar aún más los resultados (Cortina, 2015; Salcedo Parra et al., 2010).

Capítulo 5. Plan de pruebas y resultados

Capítulo 5. Plan de pruebas y resultados

5.1 Caso de estudio

La experimentación se llevó a cabo probando la metodología en un caso de estudio, utilizando las bases de datos, la plataforma web y los procesos de negocio del Verano Internacional en Ciencia y Tecnología del Tecnológico Nacional de México.

5.2 Fase 1. Entendimiento del negocio

5.2.1 Identificación de objetivos empresariales

Uno de los objetivos principales de este evento es aumentar la participación estudiantil y, al mismo tiempo, incrementar la tasa de asignación de proyectos, asegurando que todas las instituciones inscritas puedan participar. Por esta razón, se decidió trabajar con el indicador de desempeño: 'Tasa de asignación de proyectos'.

5.2.2 Modelado de los procesos de negocio

El caso de estudio se enfoca en un metaproceso de negocio (Figura 3), compuesto por varios subprocesos interrelacionados. Este metaproceso incluye la participación de roles clave, como el departamento de vinculación, el administrador de la plataforma web, los profesores, los estudiantes y los revisores. La interacción entre estos roles es esencial para coordinar y ejecutar las actividades que permiten alcanzar los objetivos establecidos. Para facilitar la comprensión y visualización de este metaproceso, este se modeló utilizando la notación BPM, lo que permite representar de manera clara y estructurada las interacciones y flujos de trabajo involucrados.

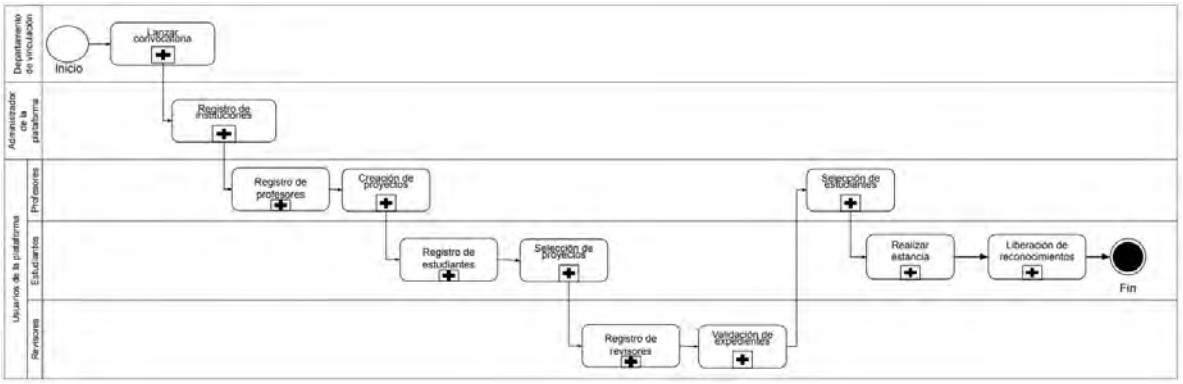


Figura 3 Metaproceso del caso de estudio

Dentro del metaproceso, destaca el subproceso de negocio “Selección de proyectos” (Figura 4), el cual se mide mediante el indicador clave de desempeño "Tasa de asignación de proyectos". Este subproceso también fue modelado utilizando la notación BPM para

proporcionar una representación clara y estructurada. A través de su modelado, se identificaron tanto las actividades donde ya se almacenaban datos relevantes, como aquellas actividades que presentan oportunidades para recolectar información adicional.

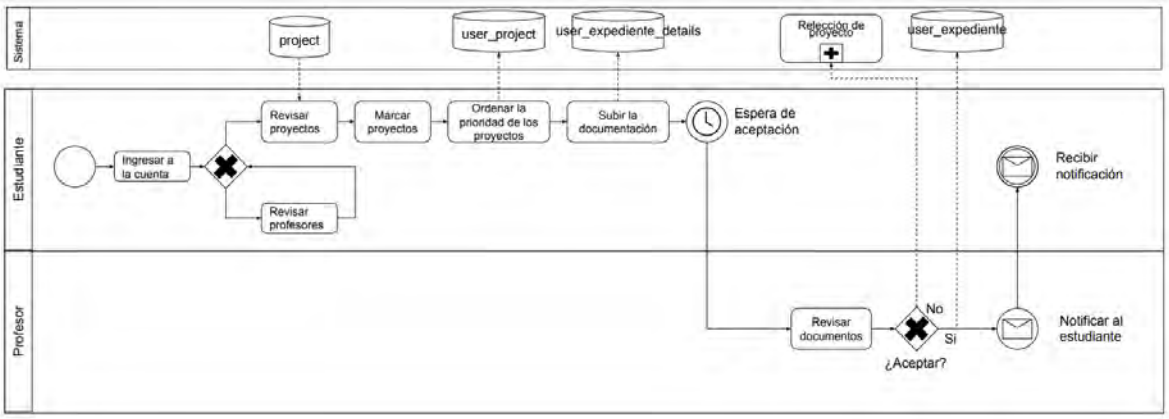


Figura 4 Proceso de negocio en notación BPMN: Selección de proyectos

5.2.3 Evaluación de la situación actual

En esta etapa se plantearon las siguientes preguntas:

- ¿Quiénes ejecutan las actividades del proceso de negocio?

Los profesores y los estudiantes son quienes participan dentro del proceso "Selección del proyecto"

- ¿Cuál es la situación actual del indicador clave de desempeño?

La gráfica de la Figura 5 muestra que, durante el evento de 2023, la Tasa de asignación de proyectos fue del 70%, es decir, 320 de los 457 proyectos fueron seleccionados y desarrollados por los estudiantes.

Figura 5 Tasa de asignación de proyectos del año 2023

La gráfica de la Figura 6 muestra que, durante el evento del año 2024, la Tasa de asignación de proyectos fue del 57%, lo que significa que 333 de los 583 proyectos fueron seleccionados y desarrollados por los estudiantes. En comparación con el año 2023, este resultado refleja un incremento en la participación de los profesores en el registro de proyectos, mientras que la participación de los estudiantes en la selección y desarrollo de los mismos se ha mantenido casi constante.

Figura 6 Tasa de asignación de proyectos del año 2024

- ¿En qué entorno se ejecuta el proceso de negocio?

El proceso de negocio se ejecuta en una plataforma web

- ¿En qué actividades se almacenan datos?

Se almacenan datos en las actividades: Ordenar la prioridad de los proyectos, subir documentación y revisión de documentos.

- ¿Cómo se almacenan los datos?

El proceso almacena sus datos en una base de datos relacional en phpMyadmin

5.2.4 Definición del objetivo de minería de datos

El propósito de este análisis es identificar patrones útiles que contribuyan a mejorar la Tasa de asignación de proyectos en futuras ediciones. Para ello, se planteó la siguiente pregunta:

¿Cuáles son las características que diferencian a los proyectos seleccionados de los no seleccionados?

Con base en esta pregunta, se definió como objetivo del análisis de minería de datos la identificación de las características de los proyectos que influyen directamente en su selección.

5.3 Fase 2. Entendimiento de los datos

5.3.1 Identificación del contexto y tipo de las variables.

En esta etapa se identificaron la función y el tipo de dato de cada variable en la base de datos, con el objetivo de determinar cuáles estaban directamente involucradas en el proceso de Selección de proyectos. Para ello, se elaboró un diccionario de variables que facilita la comprensión y clasificación de cada una. Se encontraron variables de distintos tipos, como cadenas de texto (String, Str), números enteros (Int) y fechas (timestamp). Este diccionario permite una mejor organización de los datos y asegura su correcta preparación para el análisis posterior. Las variables con sus respectivas descripciones y tipos de datos se presentan en el Anexo A. Diccionario de las variables de la base de datos del caso de estudio.

5.3.2 Análisis exploratorio

El análisis exploratorio consistió en revisar la distribución de las bases de datos y en plantearse las siguientes preguntas:

- ¿Qué hipótesis sobre los datos se han formulado?

Se consideró que la modalidad y la temática de los proyectos influye en la Tasa de asignación de proyectos debido a que se encontró que existen preferencias de los estudiantes por temáticas y modalidades. La grafica de la Figura 7 muestra en color verde los proyectos seleccionados y en color rojo los proyectos no seleccionados, se aprecia que la temática "Ingeniería e industria" es la que tiene una mayor cantidad de proyectos registrados, debido a que tiene el mayor número de proyectos seleccionados y no seleccionados. La gráfica también revela que la mayoría de los proyectos se concentra en las temáticas de Biotecnología y ciencias de la tierra, Tecnología 4.0, Ingeniería e industria y Ciencias sociales. Esto sugiere que, dentro de estas temáticas, podrían existir características específicas que ayuden a diferenciar los proyectos seleccionados de los no seleccionados, planteando la necesidad de un análisis detallado para identificar estos factores distintivos.

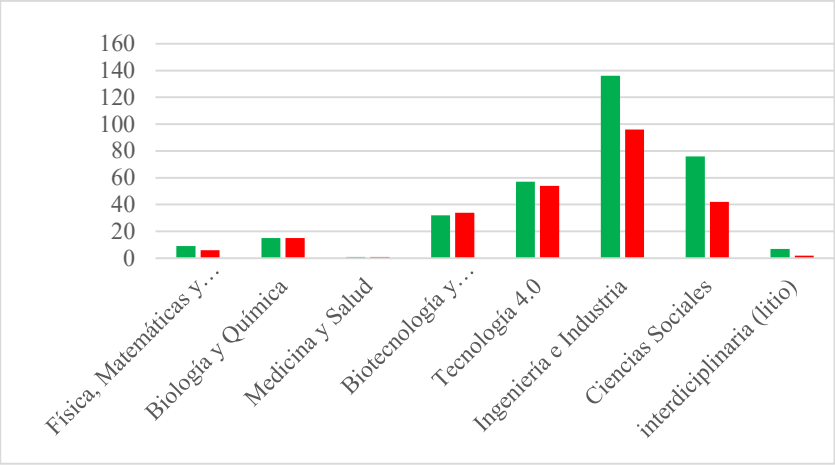


Figura 7 Gráfico de barras de las preferencias de los proyectos por temática

La gráfica de la Figura 8 muestra en color verde los proyectos seleccionados y en color rojo los proyectos no seleccionados, se aprecia que la mayoría de los proyectos fueron registrados en modalidad hibrida, debido a que tiene el mayor número de proyectos seleccionados y no seleccionados.

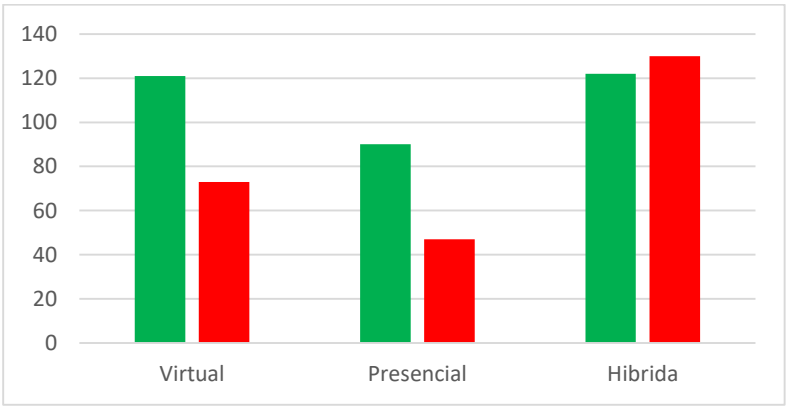


Figura 8 Gráfico de barras de las preferencias de los proyectos por modalidad

También se consideró que la cantidad de texto en la descripción, objetivo y título del proyecto puede influir en la selección de los proyectos. Debido a que existen proyectos que tienen una gran cantidad de texto. También se consideró que la fecha de registro influye en la tasa de asignación de proyectos debido a que entre más temprano se registren los proyectos más tiempo se muestran en la plataforma web.

- ¿Cuáles son las variables que parecen prometedoras para el análisis?

La modalidad, la temática, la fecha de registro del proyecto, el stock, titulo, objetivo y descripción.

- ¿Esta exploración ha cambiado su objetivo de minería de datos?

No, se concluyó que no era necesario cambiar el objetivo de minería de datos

- ¿Se cuenta con los datos necesarios para realizar el análisis?

No, se concluyó que se necesitaba monitorear el comportamiento de los estudiantes cuando ejecutan la actividad *Revisar proyectos* debido a que en la base de datos existente no se contaba con la información necesaria para el análisis.

- ¿Existe manera de extraer más información de las actividades?

Si, el entorno en que se ejecuta el proceso de negocio es mediante una plataforma web.

5.3.3 Identificación de la calidad de los datos

En la base de datos del verano científico se identificaron errores ortográficos que posteriormente fueron corregidos en las variables: Requisitos, Prerrequisitos y keywords.

5.3.4 Extracción de datos de la ejecución de las actividades de los procesos

Debido a que se consideró a que no se contaba la información necesaria para el análisis se realizó este paso. La extracción de datos de la ejecución de las actividades de los procesos se realizó mediante la creación de una tabla llamada *Bitacora_proyectos* (Tabla 4), y la programación de funciones en JavaScript, denominadas en conjunto como "Funcionalidad de eventos". Tanto las funciones como la tabla tuvieron el objetivo de obtener principalmente el número de vistas que tuvo cada proyecto e información sobre el comportamiento de los estudiantes en la plataforma web del verano científico cuando realizan la actividad de "Revisar proyectos" del proceso de negocio *Selección de proyectos*.

La creación de la nueva tabla en la base de datos fue posible gracias a la herramienta "Laragon", que es un software que permite crear un servidor local para utilizar herramientas de desarrollo web, como phpMyAdmin, que permite gestionar las bases de datos de las páginas web. La nueva tabla, llamada "Bitacora_proyectos", fue añadida a la base de datos ya existente del "Verano Científico". La tabla contiene los siguientes atributos:

Tabla 4 Atributos de la tabla "Bitacora_proyectos"

Bitacora_proyectos		
Nombre de la columna	**Tipo de dato**	**Atributos de la base de datos de los roles**
id	int	Identificador de la instancia
ip_usuario	varchar(20)	Dirección IP del usuario que entra a ver los proyectos de la página del verano científico
id_usuario	int	Identificador del usuario
id_proyecto	int	Identificador del proyecto
fecha_entrada	date	Fecha en que el usuario comienza a revisar el proyecto

hora_entrada	time	Hora en que el usuario comienza a revisar el proyecto
fecha_salida	date	Fecha en que el usuario termina de revisar el proyecto
hora_salida	time	Hora en que el usuario termina de revisar el proyecto

La Funcionalidad de eventos fue desarrollada utilizando la arquitectura MVC (Modelo Vista Controlador). Esta funcionalidad se encarga de registrar el comportamiento de los estudiantes en la tabla "Bitacora_proyectos" mientras navegan por el catálogo de proyectos en la plataforma. La funcionalidad incluye dos algoritmos (A y B), cuyos códigos se encuentran en el Anexo B. Código de la funcionalidad de eventos. El algoritmo A registra cuando los estudiantes comienzan a revisar los detalles de los proyectos, este está compuesto de la siguiente manera:

Tabla 5 Estructura MVC del algoritmo A

Componente MVC	Nombre de la función	Descripción
Vista	onMuestraDetalle	Muestra la información de los proyectos cuando un estudiante da clic en el botón “Ver Detalle” y manda a llamar la función registroBitacora.
	registroBitacora	1. Obtiene id_proyecto, fecha_entrada y hora_entrada. 2. Envía los datos a InsertBitacora.
Controlador	InsertBitacora	1. Recibe los datos obtenidos de registroBitacora. 2. Obtiene ip_usuario e id_usuario 3. Construye un arreglo con los datos obtenidos. 4. Envía el arreglo al modelo IBitacora
Modelo	IBitacora	1. Recibe el arreglo del controlador InsertBitacora 2. Asigna los datos del arreglo a cada una de las columnas correspondientes de la tabla “Bitacora_proyectos”

El algoritmo B registra cuando los estudiantes terminan de revisar los detalles de los proyectos. Además, el algoritmo B depende del algoritmo A, ya que, si el algoritmo A no se ejecuta primero, el algoritmo B no puede ejecutarse. Los componentes del algoritmo B se muestran en la Tabla 6.

Tabla 6 Estructura MVC del algoritmo B

Componente MVC	Nombre de la función	Descripción
Vista	onRegresarProyectos	Devuelve al estudiante al catálogo de proyectos cuando da un clic en el botón “Regresar”.
	registroBitacora_out	1. Obtiene id_proyecto, fecha_salida y hora_salida. 2. Envía los datos a ActBitacora
Controlador	ActBitacora	1. Recibe los datos obtenidos de registroBitacora_out. 2. Obtiene ip_usuario e id_usuario 3. Construye un arreglo con los datos obtenidos. 4. Envía el arreglo al modelo ABitacora
Modelo	ABitacora	1. Recibe el arreglo del controlador ActBitacora 2. Busca coincidencias con los datos obtenidos con los datos registrados en la tabla “Bitacora_proyectos”, cuando encuentra que el id_proyecto y el id_usuario, o el id_proyecto y el ip_usuario, coinciden entones inserta la fecha_salida y hora_salida a la tabla.

Al agregar esta nueva tabla a la base de datos y crear nuevas funciones para registrar la entrada de datos, fue necesario modificar el proceso de negocio Selección de proyectos de la Figura 4. Estos cambios se reflejaron en el modelo del proceso utilizando notación BPMN, actualizando las actividades y puntos de decisión involucrados. La Figura 9 presenta el proceso modelado con la “Funcionalidad de eventos” implementada. La flecha punteada de color azul representa la “Funcionalidad de eventos”, mientras que el símbolo de base de datos de color azul, representa la tabla agregada a la base de datos existente.

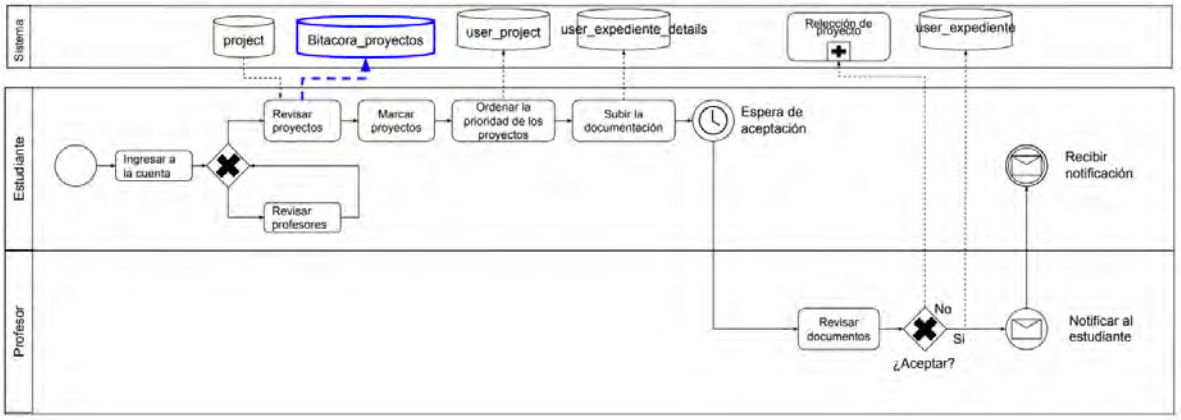


Figura 9 Proceso de “Selección de proyectos” con la funcionalidad de eventos

5.4 Fase 3. Procesamiento de los datos.

5.4.1 Creación de un conjunto de datos para el caso de estudio

La creación del conjunto de datos implicó la selección y organización de las variables más relevantes de los proyectos para el análisis. En total, se recopilaron 583 instancias que pertenecen a los proyectos registrados del año 2024, y se incluyeron las variables mostradas en la Tabla 7 como base para el análisis.

Tabla 7 Variables del primer conjunto de datos

Nombre	Tipo	Descripción
Id_proyecto	Int	Número identificador del proyecto
Vistas	Int	Número de vistas que tuvo cada proyecto
Modalidad	Int	Modalidad en que el alumno desarrollará el proyecto: 1 – Virtual, 2 – Presencial y 3 – Hibrida
Stock	Int	Número de estudiantes permitidos en el proyecto
Id_Institución	Int	Número identificador de la institución que propone el proyecto
Id_temática	Int	Número identificador de la temática a la que pertenece el proyecto
Fecha de registro	Timestamp	Fecha en que se registró el proyecto en formato DD-MM-AAAA HH:MM:SS
Titulo	Str	Título del proyecto
Objetivo	Str	Objetivo del proyecto
Descripción	Str	Descripción del proyecto
Requisitos	Str	Requerimientos de habilidades y conocimientos
Prerrequisitos	Str	Requerimientos de hardware y software
Keywords	Str	Palabras clave de los proyectos

5.4.2 Preparación de los datos

La preparación de los datos implicó transformarlos en un conjunto utilizable para su procesamiento por los modelos, evitando así el ruido y las inconsistencias que pudieran afectar los resultados.

I. Limpieza de datos

En este caso de estudio, se hicieron correcciones ortográficas a las variables: Titulo, Objetivo, Descripción, Requisitos, Prerrequisitos y Keywords. Este paso se realizó manualmente.

II. Transformación de datos

En este caso de estudio se identificaron las características de las variables del conjunto de datos. La variable *Vistas* tiene valores numéricos continuos (int); la variable *Modalidad* se representa como numérica categórica (int); y la variable *Fecha de registro* toma valores de tipo timestamp en formato DD-MM-AAAA HH:MM:SS. Por otro lado, las variables *Título, Objetivo, Descripción, Requisitos, Prerrequisitos* y *Keywords* contienen datos en formato de texto (Str).

Para homogenizar los datos y garantizar que todas las variables fueran de tipo numérico categórico, se aplicaron diferentes procesos de transformación según las características de cada variable, facilitando así su preparación para el análisis posterior.

Variable: *Vistas*

El número de vistas es una variable con datos numéricos continuos, por lo que se utilizó un enfoque de discretización. Para ello, se creó la variable "Rvistas" y se definieron cuatro rangos de vistas como lo muestra la Tabla 8, permitiendo una mejor categorización y análisis de esta variable en el contexto del estudio:

Tabla 8 Rangos de vistas

Rvistas	Rangos de vistas
0	0
1	1 ≤Vistas≤ 9
2	9 <Vistas≤ 18
3	18 <Vistas≤ 27

Variable: *Modalidad*

La variable Modalidad, originalmente una variable numérica categórica, fue transformada para crear nuevas variables categóricas binarias. Para ello, se extrajeron los valores de la variable original y se generó una columna independiente para cada modalidad: Virtual, Presencial e Híbrida. Cada una de estas nuevas variables toma valores binarios, indicando con un "1" la presencia de la modalidad correspondiente y con un "0" su ausencia.

Variable: *Fecha de registro*

La variable fecha de registro fue transformada a una nueva variable denominada como DRC (*Días transcurridos desde el inicio de la convocatoria hasta que se registró el proyecto*), esa variable se obtuvo con la siguiente formula en Excel:

$$DCR = max\ (Fecha\ de\ registro - Fecha\ de\ salida\ de\ la\ convocatoria, 0)$$

Donde:

- Fecha de registro: Fecha en que se registró el proyecto
- Fecha de salida de la convocatoria: Fecha en que los proyectos podían empezar a registrarse

DCR es una variable numérica continua por lo que fue discretizada en dos nuevas variables Quincena de registro (Tabla 9) y Mes de registro (Tabla 10).

Tabla 9 Rangos de la variable "Quincena de registro"

Quincena de registro	Rango de días
1	0 ≤ Días ≤ 15
2	15 < Días ≤ 30
3	30 < Días ≤ 45
4	45 < Días ≤ 60
5	60 < Días ≤ 75
6	75 < Días ≤ 90

Tabla 10 Rangos de la variable "Mes de registro"

Mes de registro	Rango de días
1	0 ≤ Días ≤ 30
2	30 < Días ≤ 60
3	60 < Días ≤ 90

Variable: *Requisitos y prerrequisitos*

Los requisitos y prerrequisitos de los proyectos se categorizaron para poder incluirlos en el análisis, de esta manera se puede identificar si existe una asociación que afecte su selección. Para poder categorizarlos se utilizó una nube de palabras del software Orange, la cual permite identificar cuáles son las palabras más recurrentes.

Figura 10 Nube de palabras de requisitos y prerrequisitos frecuentes

Al identificar los requisitos más recurrentes de la nube de palabras como lo muestra la Figura 10, se propusieron las siguientes categorías**:** Hardware, Software y Conocimientos. Los requisitos de Hardware se agruparon en una solo variable, ya que los proyectos suelen solicitar computadoras con características específicas. Por otro lado, los requisitos de Software y Conocimientos se reflejaron en variables separados, permitiendo una mejor organización y análisis de los proyectos en función de sus necesidades.

A continuación, se enlistan algunos de los requisitos de software (empiezan con "R_") y los de conocimientos (empiezan con "C_"):

- R_Python
- R_HeramientasGoogle
- C_Sensores
- C_Lenguaje natural

Variable: *Keywords*

Las keywords se utilizaron para crear grupos, a los cuales se les denominó como impactos y consisten en identificar si existe una correlación entre las temáticas de los proyectos con otras áreas. Las áreas que se definieron fueron:

a) Desarrollo digital
b) Sistemas de control e instrumentación
c) Energía y medio ambiente
d) Redes y comunicación
e) Educación y capacitación
f) Ciencia e investigación

g) Manufactura, diseño y producción
h) Administración y mercadotécnica
i) Salud y seguridad
j) Agricultura y alimentación
k) Social y comunitario

Estas áreas fueron definidas por las palabras clave más frecuentes que tienen los proyectos, para esto, se utilizó el software Orange, que tiene herramientas como la nube de palabras que permiten visualizar las variables de tipo string (texto), la nube de palabras resultante se muestra en la Figura 11.

Figura 11 Nube de palabras de las palabras clave de los proyectos

Por lo que se asignaron estas palabras clave a cada impacto correspondiente como lo muestra la Tabla 11.

Tabla 11 Impactos encontrados en la base de datos

Impacto	Palabras
Desarrollo digital	Datos, IA, inteligencia artificial, IoT, Big data, Deep learning, web, industria 4.0, análisis de datos, minería de datos, Python, programación, software, machine learning, computadora, computación, api, plataforma, ciencia de datos, data science, mySqL, lenguaje, algoritmos, lógica, backend, fronted, javascript, json, realidad virtual, reconocimiento de patrones, servidores, tecnología digital, código, PHP, blazor.
Sistemas de control e instrumentación	Automatización, control, instrumentación, robotica, sistemas de control, PLC, sensores, actuadores, electrónica, electricidad, circuitos, microcontroladores, hardware, componentes eléctricos, sistemas embebidos, arduino.

Energía y medio ambiente	Energía, sustentabilidad, sostenibilidad, desarrollo sostenible, desarrollo sustentable, renovable, solar, fotovoltaico, agua, medio ambiente, ambiente, residuos, manejo de residuos, ecología, ecológico, baterías, hidráulica, ciclo, conservación.
Redes y comunicación	Red, redes, protocolos de red, comunicación, Wi-Fi, Bluetooth, redes de área amplia, infraestructura de redes, celular, conectividad, satelital.
Educación y capacitación	Educación, capacitación, formación, académico, académica, aprendizaje, enseñanza, docente, docencia, curso, seminario, universidad
Ciencia e investigación	Investigación, análisis, diagnostico, científico, experimentación, tecnologías, métodos, experimentación, experimental, laboratorio, prueba, química, biotecnología, genética, biorreactores, bioestimulantes,
Manufactura, diseño y producción	Manufactura, producción, industria, industrial, logística, procesos, maquinaria, calidad, planificación, diseño, cadena de suministros, fabricación, SMART, solidworks, autocad, materiales, 3D, fatiga.
Administración y mercadotécnica	Negocios, administración, mercado, comercialización, marketing, mercadotécnica, ventas, financiero, finanzas, comercio, estrategia, cliente, precios, negociación, gestión, planificación, empresa, competitividad, empresarial, PYMES.
Salud y seguridad	Salud, seguridad, prevención, riesgo, bienestar, medicina, medicina, sanitarios.
Agricultura y alimentación	Agricultura, cultivos, alimentos, agrodiversidad, fertilización, semillas, hortalizas, producción agrícola, nutrición, agroindustria, granja, coco, fruta, vegetal, plagas, maíz.
Social y comunitario	Social, sociedad, comunidades, comunidad, comunitario, cultural, discapacidad, inclusión, justicia, derechos, diversidad, pobreza, lgbtq, inclusivo, inclusiva, solidaridad.

Los impactos forman once columnas nuevas en la base datos, que toman valores binarios, 1 si se encuentra alguna de las palabras en cada instancia, de no ser así se toma como valor un 0. Los proyectos pueden tener más de un impacto, sin embargo, se trata de encontrar si es que existe una asociación entre temáticas e impacto de los proyectos que afecte su selección.

III. Balanceo de clases

El balanceo de clases se llevó a cabo sobre la variable objetivo, denominada "Selección", para evitar que el modelo se sesgue hacia una de las clases debido a un desequilibrio en la distribución de las instancias. En este caso de estudio, del total de 583 instancias, 333 pertenecen a la clase 1 (proyectos seleccionados) y 250 a la clase 0 (proyectos no seleccionados). Este desequilibrio podría limitar la capacidad del modelo para identificar correctamente los proyectos no seleccionados.

Para abordar este problema, se utilizó el nodo de balanceo SMOTE (Synthetic Minority Oversampling Technique) en KNIME, el cual generó 83 instancias sintéticas de la clase de proyectos no seleccionados. Como resultado, se obtuvo un conjunto de datos equilibrado con 333 instancias en ambas clases, garantizando una distribución equitativa.

5.4.3 Realización del análisis de asociaciones

El análisis de asociaciones consistió en descartar las variables que no presentaban relevancia para el análisis. El objetivo fue identificar las variables más significativas y relacionadas a la variable objetivo "Selección" del conjunto de datos mostrado en la Tabla 12, garantizando que solo se incluyan en el modelo aquellos que aporten valor al análisis.

Tabla 12 Variables procesadas del conjunto de datos

Nombre	Tipo	Descripción
Id_proyecto	Int	Número identificador del proyecto
Rvistas	Int	Rangos de numero de vistas
Modalidad	Int	Número identificador de modalidad
Virtual	Int	1, Sí el proyecto será desarrollado en modalidad virtual, 0 si no es así.
Presencial	Int	1, Sí el proyecto será desarrollado en modalidad presencial, 0 si no es así.
Hibrida	Int	1, Sí el proyecto será desarrollado en modalidad hibrida, 0 si no es así.
Stock	Int	Número de estudiantes permitidos en el proyecto
Id_Institucion	Int	Número identificador de la institución
Física, Matemáticas y Ciencias de la Tierra	Int	1, Si el proyecto está registrado en la temática 1 0, Sí el proyecto no está registrado en la temática 1
Biología y Química	Int	1, Si el proyecto está registrado en la temática 2 0, Sí el proyecto no está registrado en la temática 2
Medicina y Salud	Int	1, Si el proyecto está registrado en la temática 3 0, Sí el proyecto no está registrado en la temática 3
Biotecnología y Ciencias Agropecuarias	Int	1, Si el proyecto está registrado en la temática 4 0, Sí el proyecto no está registrado en la temática 4
Tecnología 4.0	Int	1, Si el proyecto está registrado en la temática 5 0, Sí el proyecto no está registrado en la temática 5
Ingeniería e Industria	Int	1, Si el proyecto está registrado en la temática 6 0, Sí el proyecto no está registrado en la temática 6
Ciencias Sociales	Int	1, Si el proyecto está registrado en la temática 7 0, Sí el proyecto no está registrado en la temática 7
interdiciplinaria (litio)	Int	1, Si el proyecto está registrado en la temática 8 0, Sí el proyecto no está registrado en la temática 8
Caracteres Titulo	Int	Número de caracteres que tiene el título de un proyecto

Caracteres Descripción	Int	Numero de caracteres que tiene la descripción del proyecto
Días transcurridos desde que salió la convocatoria	Int	Número de días transcurridos desde que salió la convocatoria hasta que se registró el proyecto
Quincena de registro	Int	Toma valores del 1 al 6 como lo muestra la Tabla 9
Mes de registro	Int	Toma valores del 1 al 3 como lo muestra la Tabla 10
Hardware	Int	1, Si el proyecto tiene requisitos de hardware 0, Sí el proyecto no tiene requisitos de hardware
Software	Int	1, Si el proyecto tiene requisitos de software 0, Sí el proyecto no tiene requisitos de software
Conocimientos	Int	1, Si el proyecto tiene requisitos de conocimientos 0, Sí el proyecto no tiene requisitos de conocimientos
R_Python	Int	1, Si el proyecto tiene como requisito Python 0, Sí el proyecto no tiene como requisito Python
R_Herramientas de google	Int	1, Si el proyecto tiene como requisito Herramientas de Google 0, Sí el proyecto no tiene como requisito Herramientas de Google
R_Microsoft office	Int	1, Si el proyecto tiene como requisito Microsoft Office 0, Sí el proyecto no tiene como requisito Microsoft Office
R_Solidworks	Int	1, Si el proyecto tiene como requisito Solidworks 0, Sí el proyecto no tiene como requisito Solidworks
R_Arduino	Int	1, Si el proyecto tiene como requisito Arduino 0, Sí el proyecto no tiene como requisito Arduino
R_Autocad	Int	1, Si el proyecto tiene como requisito Autocad 0, Sí el proyecto no tiene como requisito Autocad
R_Matlab	Int	1, Si el proyecto tiene como requisito Matlab 0, Sí el proyecto no tiene como requisito Matlab
R_Conocimiento sensores	Int	1, Si el proyecto tiene como requisito Arduino 0, Sí el proyecto no tiene como requisito Arduino
C_Lenguaje Natural	Int	1, Si el proyecto requiere de conocimientos de lenguaje natural 0, Sí el proyecto no requiere de conocimientos de lenguaje natural
C_Inteligencia artificial	Int	1, Si el proyecto requiere de conocimientos de inteligencia artificial 0, Sí el proyecto no requiere de conocimientos de inteligencia artificial
C_Electronica	Int	1, Si el proyecto requiere de conocimientos de electrónica

		0, Sí el proyecto no requiere de conocimientos de electrónica
C_Instrumentación	Int	1, Si el proyecto requiere de conocimientos de instrumentación 0, Sí el proyecto no requiere de conocimientos de instrumentación
C_Quimica	Int	1, Si el proyecto requiere de conocimientos de química 0, Sí el proyecto no requiere de conocimientos de química
C_Mecánica	Int	1, Si el proyecto requiere de conocimientos de mecánica 0, Sí el proyecto no requiere de conocimientos de mecánica
C_Termodinamica	Int	1, Si el proyecto requiere de conocimientos de termodinámica 0, Sí el proyecto no requiere de conocimientos de termodinámica
C_Estadistica	Int	1, Si el proyecto requiere de conocimientos de estadística 0, Sí el proyecto no requiere de conocimientos de estadística
C_Biología	Int	1, Si el proyecto requiere de conocimientos de biología 0, Sí el proyecto no requiere de conocimientos de biología
C_Agropecuaria	Int	1, Si el proyecto requiere de conocimientos de agropecuaria 0, Sí el proyecto no requiere de conocimientos de agropecuaria
C_Logistica	Int	1, Si el proyecto requiere de conocimientos de logística 0, Sí el proyecto no requiere de conocimientos de logística
C_Electricidad	Int	1, Si el proyecto requiere de conocimientos de electricidad y magnetismo 0, Sí el proyecto no requiere de conocimientos de electricidad y magnetismo
C_Redes sociales	Int	1, Si el proyecto requiere de conocimientos en redes sociales 0, Sí el proyecto no requiere de conocimientos en redes sociales
C_Android	Int	1, Si el proyecto requiere de conocimientos de Android 0, Sí el proyecto no requiere de conocimientos de Android

Desarrollo digital	Int	1, Si el proyecto tiene un impacto orientado al desarrollo digital 0, Sí el proyecto no tiene un impacto orientado al desarrollo digital
Sistemas de control e instrumentación	Int	1, Si el proyecto tiene un impacto orientado a los sistemas de control e instrumentación 0, Sí el proyecto no tiene un impacto orientado a los sistemas de control e instrumentación
Energía y medio ambiente	Int	1, Si el proyecto tiene un impacto orientado a la energía y el medio ambiente 0, Sí el proyecto no tiene un impacto orientado a la energía y el medio ambiente
Redes y comunicación	Int	1, Si el proyecto tiene un impacto orientado a las redes y comunicación 0, Sí el proyecto no tiene un impacto orientado a las redes y comunicación
Educación y capacitación	Int	1, Si el proyecto tiene un impacto orientado a la educación y capacitación 0, Sí el proyecto no tiene un impacto orientado a la educación y capacitación
Ciencia e investigación	Int	1, Si el proyecto tiene un impacto orientado a la ciencia e investigación 0, Sí el proyecto no tiene un impacto orientado a la ciencia e investigación
Manufactura, diseño y producción	Int	1, Si el proyecto tiene un impacto orientado a la manufactura, diseño y producción 0, Sí el proyecto no tiene un impacto orientado a la manufactura, diseño y producción
Administración y mercadotécnica	Int	1, Si el proyecto tiene un impacto orientado a la administración y mercadotecnia 0, Sí el proyecto no tiene un impacto orientado a la administración y mercadotecnia
Salud y seguridad	Int	1, Si el proyecto tiene un impacto orientado a la salud y seguridad 0, Sí el proyecto no tiene un impacto orientado a la salud y seguridad
Agricultura y alimentación	Int	1, Si el proyecto tiene un impacto orientado a la agricultura y agropecuaria 0, Sí el proyecto no tiene un impacto orientado a la agricultura y agropecuaria
Social y comunitario	Int	1, Si el proyecto tiene un impacto orientado a lo social y comunitario 0, Sí el proyecto no tiene un impacto orientado a lo social y comunitario
Selección	Int	1, Si el proyecto fue seleccionado 0, Sí el proyecto no fue seleccionado

El análisis de asociaciones se llevó a cabo utilizando la prueba de Chi-cuadrado, dado que este enfoque fue el más adecuado para identificar asociaciones significativas entre variables categóricas, sin requerir que los datos sean numéricos lineales. Es importante destacar que, cuando el valor de Chi-cuadrado es elevado, el p-valor será pequeño, lo que indica la existencia de una asociación significativa entre las variables. Para que esta asociación sea considerada significativa, el p-valor debe ser menor a 0.05 (p-valor<0.05).

Además, para cuantificar el grado de asociación entre las variables, se utilizó la V de Cramer, la cual pondera la fuerza de la asociación en variables categóricas, la Tabla 3 muestra los rangos de relación que se utilizaron para este análisis. El análisis se realizó en Python utilizando las librerías pandas y scipy, lo que permitió un procesamiento automático y eficiente de los datos. La Tabla 13 muestra en la primera columna la variable objetivo, en la segunda columna denominada *Variable 2* se muestran los nombre de las variables que están más asociadas a la variable objetivo, la tercera columna muestra los valores obtenidos del Chi-Cuadrado, la cuarta columna muestra los valores obtenidos del p-valor, y la última columna muestra los valores obtenidos de la V de Cramer.

Tabla 13 Asociaciones encontradas de la variable "Selección".

Variable 1	*Variable 2*	*Chi-Cuadrado*	*p-valor*	*V-Cramer*
Selección	Rvistas	278.14	$5.3328x10^{-60}$	0.6907
Selección	id_institucion	245.51	$3.207x10^{-18}$	0.6489
Selección	Id_temática	8.56	0.03409812	0.1212
Selección	Modalidad	14.095	0.00086921	0.1554
Selección	Mes de registro	14.42	0.00073642	0.1573
Selección	Quincena de registro	17.02	0.00445186	0.1708

A continuación, se interpretan las asociaciones encontradas.

Asociación: Selección y Rvistas:

El *p-valor* es de $5.3328X10^{-60}$, lo que es significativamente menor a 0.05, indicando que existe una asociación entre las variables. Además, el valor de V de Cramer, de 0.6907, refleja un grado de asociación muy fuerte entre el número de vistas y la selección de los proyectos. En conclusión, el número de vistas que recibe un proyecto influye de manera significativa en su selección.

Asociación: Selección e id_institucion:

El *p-valor* obtenido es de 3.21×10^{-18}, lo cual es significativamente menor a 0.05, lo que indica que existe una asociación entre las variables. Además, el valor de V de Cramer de

0.6489 refleja un grado de asociación muy fuerte entre las instituciones y la selección de los proyectos.

Asociación: Selección e id_tematica:

El *p-valor* es de 0.03409812, lo que es significativamente menor a 0.05, indicando que existe una asociación entre las variables. Además, el valor de V de Cramer, de 0.1212, refleja un grado de asociación medio entre el id_temática y la selección de los proyectos. En conclusión, la temática es un factor que influye en la selección de los proyectos.

Asociación: Selección y modalidad:

El *p-valor* es de 0.00086921, lo que es significativamente menor a 0.05, indicando que existe una asociación entre las variables. Además, el valor de V de Cramer, de 0.1554, refleja un grado de asociación medio entre la modalidad y la selección de los proyectos. En conclusión, la modalidad es un factor que influye en la selección de los proyectos.

Asociación: Selección, Mes de registro y Quincena de registro:

Los valores del *p-valor* obtenidos son de 0.00073642 para el mes de registro y de 0.00445186 para la quincena de registro, los cuales son significativamente menores a 0.05, lo que indica que existe una asociación entre las variables. Además, los valores de V de Cramer son de 0.1573 para el mes de registro y de 0.1708 para la quincena de registro, estos reflejan, un grado de asociación medio entre el tiempo y la selección de los proyectos. El tiempo de que tarda en registrarse un proyecto desde la fecha de salida de la convocatoria, es un factor que influye en la selección de los de proyecto.

5.4.4 Selección de variables

El análisis previo de asociaciones permitió identificar las variables significativas que podrían contribuir al buen desempeño de un modelo de minería de datos. La Tabla 14 muestra las variables seleccionadas para realizar el análisis.

Tabla 14 Variables que están asociadas directamente con la selección de los proyectos

Nombre	Tipo	Descripción
Rvistas	Int	Rangos de numero de vistas
Id_Institucion	Int	Número identificador de la institución
Modalidad	Int	Número identificador de modalidad
Temática	Int	Número identificador de la temática
Quincena de registro	Int	Variable categórica que indica cuántas quincenas han pasado desde la salida de la convocatoria hasta que el proyecto fue registrado.

5.5 Fase 4. Desarrollo del modelo

5.5.1 Selección del modelo

El caso de estudio se adecuó a un árbol de decisión debido a que las variables fueron transformados a datos cualitativos. Con el árbol de decisión C4.5 se validaron las asociaciones encontradas en el análisis de asociaciones.

5.5.2 Construcción del modelo de minería de datos

La construcción del modelo se realizó mediante la partición de los datos en dos conjuntos. El 70% de las instancias se utilizó para entrenar el modelo, mientras que el 30% restante se destinó a su evaluación. Este procedimiento se llevó a cabo utilizando un flujo de trabajo en KNIME, representado en la Figura 12.

Tabla 15 Simbología del flujo de trabajo 1.

Nombre	Imagen	Descripción
Excel Reader		Este nodo se encarga de leer el archivos de Excel
SMOTE		Este nodo se encarga de balancear los datos
Column Filter		Este nodo se encarga de eliminar columnas de las tablas, en los casos de estudio, las columnas representan variables
Partitioning		Este nodo se encarga de dividir los datos en dos conjuntos. En ambos casos, se utilizó el 70% de los datos para el aprendizaje del modelo, mientras que el 30% restante se utilizó para pruebas.

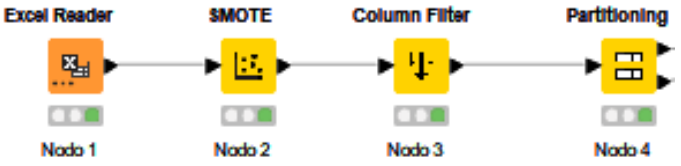
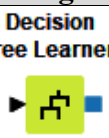


Figura 12 Flujo de trabajo 1

5.5.3 Entrenamiento del modelo

El entrenamiento del modelo se llevó a cabo incorporando un nodo denominado "Decision Tree Learner", este nuevo flujo de trabajo se muestra en la Figura 13. Este nodo es responsable de entrenar el modelo, permitiéndole identificar patrones en los datos y ajustar sus parámetros.

Tabla 16 Simbología del nodo Decision Tree Learner del flujo de trabajo 2

Nombre	Imagen	Descripción
Decision Tree Learner		Este nodo utiliza el modelo de árbol de decisión C4.5 y recibe el 70% de los datos para entrenarlo.

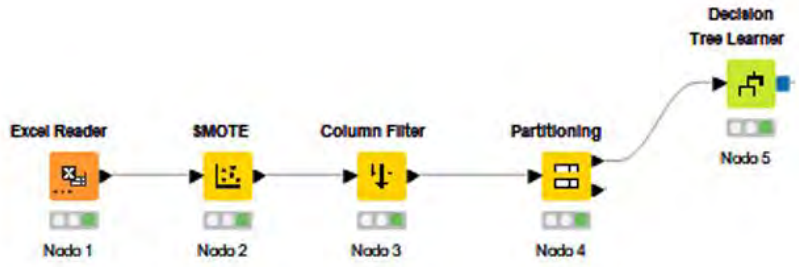


Figura 13 Flujo de trabajo 2

5.5.4 Evaluación del modelo

La evaluación del modelo se llevó a cabo incorporando dos nuevos nodos, este nuevo flujo de trabajo se muestra en la Figura 14. Estos nodos facilitan la evaluación del modelo, la Tabla 17 muestra la simbología de los nodos.

Tabla 17 Simbología de los nodos del flujo de trabajo 3

Nombre	Imagen	Descripción
Decision Tree Predictor		Este nodo toma el árbol de decisión generado por el "Decision Tree Learner" y utiliza sus reglas y estructura para hacer predicciones sobre las nuevas muestras de datos. El resultado es una columna adicional en el conjunto de datos con las predicciones realizadas por el modelo, permitiendo evaluar cómo se desempeña el modelo en datos no vistos. Además, este nodo trabaja con datos de prueba (30% de datos).
Scorer		Este nodo evalúa el rendimiento de los modelos y compara las predicciones realizadas por el modelo con las etiquetas reales de los datos y proporciona varias métricas de evaluación

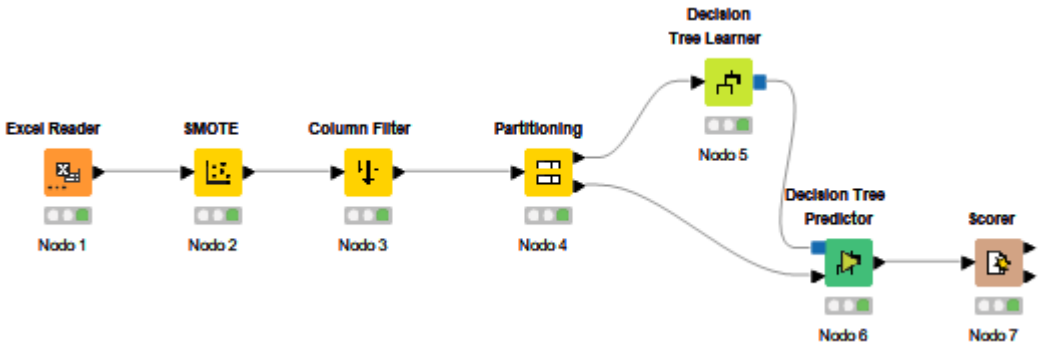


Figura 14 Flujo de trabajo 3

De acuerdo a la Figura 15 , el modelo mostró un buen rendimiento en la clasificación de proyectos seleccionados y no seleccionados, alcanzando una precisión aproximada del 83% y una tasa de error aproximada del 17%. Estos resultados indican que el modelo es adecuado para predecir la selección de proyectos. Además, los resultados demuestran que Las variables identificados en el análisis de correlaciones tienen una fuerte influencia en la selección de los proyectos.

selecciona...	1	0
1	92	5
0	28	72

Correct classified: 164 Wrong classified: 33

Accuracy: 83.249% Error: 16.751%

Figura 15 Métricas del modelo enfocado a las correlaciones de la variable: Selección

5.6 Fase 5. Interpretación y optimización

5.6.1 Interpretación de resultados

La interpretación de resultados se basa en los hallazgos encontrados y validados por el análisis de asociaciones y el desarrollo del modelo. A continuación, se muestra la interpretación de los patrones encontrados.

Patrón: Número de vistas y selección del proyecto

Los resultados obtenidos por el análisis de asociaciones y por el árbol de decisión muestran que la cantidad de vistas tiene una fuerte influencia en la selección de proyectos. La grafica de la Figura 16 muestra el número de vistas que tuvo cada proyecto seleccionado, mientras que la gráfica de la Figura 17 muestra el número de vistas que tuvo cada proyecto que no fue seleccionado, en promedio 4 vistas por proyecto mientras que los proyectos no seleccionados tienen menos vistas, en promedio 1 vista por proyecto, estos valores se muestran en la Tabla 18. En resumen, entre más vistas tenga un proyecto tiene más probabilidad de ser seleccionado.

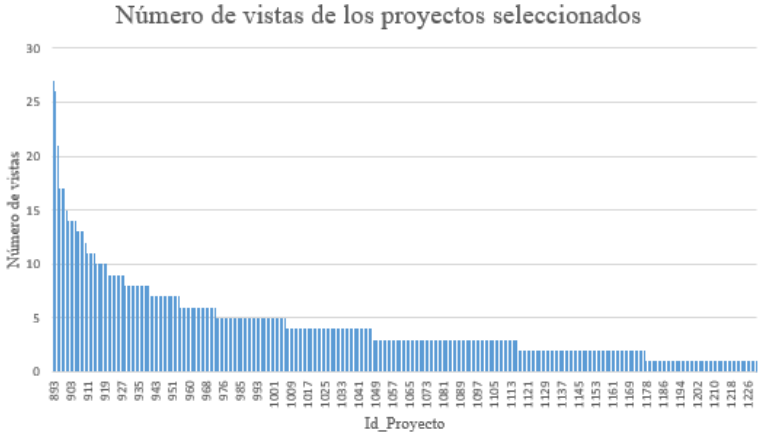


Figura 16 Número de vistas de proyectos seleccionados

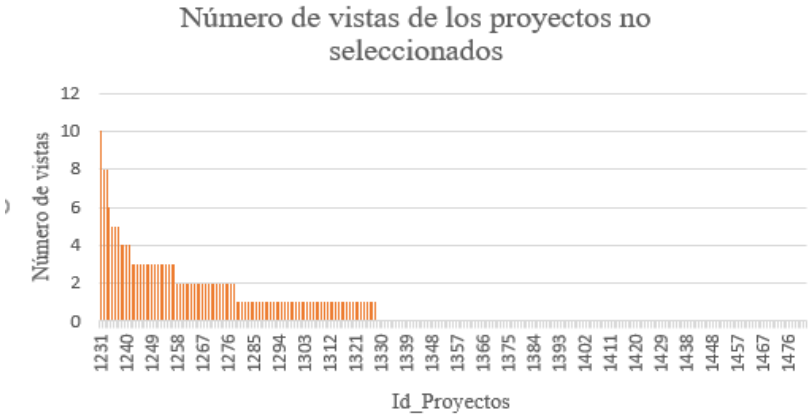


Figura 17 Número de vistas de proyectos no seleccionados

Tabla 18 Medidas estadísticas de las vistas de los proyectos seleccionados y no seleccionados

	valor mínimo	*valor máximo*	*promedio*	*moda*
Proyectos Seleccionados	1	27	4.32732733	3
Proyectos No Seleccionados	0	10	0.812	0

Patrón: Número de vistas y temática

El análisis de las vistas también fue contrastado con las temáticas de los proyectos, ya que los estudiantes pueden filtrar los proyectos por temática en la plataforma web del Verano Científico. La Figura 18 muestra el número total de vistas que tuvo cada temática, y se observa que las temáticas Biotecnología y ciencias agropecuarias, Tecnología 4.0, Ciencias Sociales e Ingeniería e Industria concentran la mayor cantidad de vistas. Esto refleja el interés predominante de los estudiantes por estas áreas temáticas, lo que sugiere que ciertos campos de estudio son más atractivos o relevantes para los participantes.

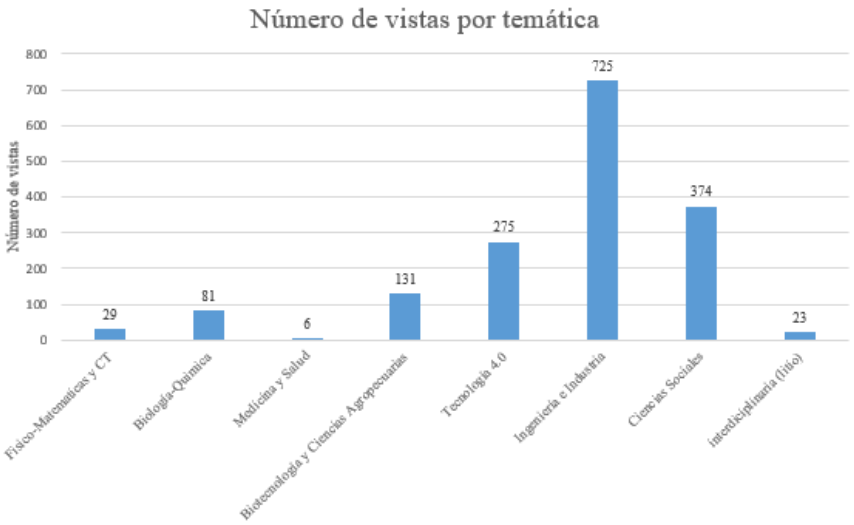


Figura 18 Número total de vistas por temática.

Patrón: Número de vistas e Institución de los proyectos

El análisis de las vistas también fue contrastado con la institución de los proyectos, la Figura 19 muestra que el 44% de las vistas corresponden a vistas de estudiantes que revisaron proyectos de su misma institución (CIE), mientras que el 45% corresponden a vistas de estudiantes que no revisaron proyectos de su misma institución (NCIE), el 10% de vistas corresponden a usuarios no registrados en la plataforma (NR), y finalmente el 1% corresponde a estudiantes que se registraron pero no está registrada su institución (INR).

Figura 19 Porcentaje de vistas realizadas por institución

Patrón: Tiempo transcurrido y la Selección de los proyectos

El análisis de correlaciones realizado reveló que existe una relación significativa entre el tiempo transcurrido desde que los proyectos fueron subidos a la plataforma y su selección. En la Figura 20, se observa que la mayor cantidad de proyectos seleccionados y no seleccionados se concentra en la primera quincena desde su registro. Sin embargo, los proyectos seleccionados tienden a destacar en este periodo, lo que indica una tendencia favorable para los proyectos que se registran tempranamente.

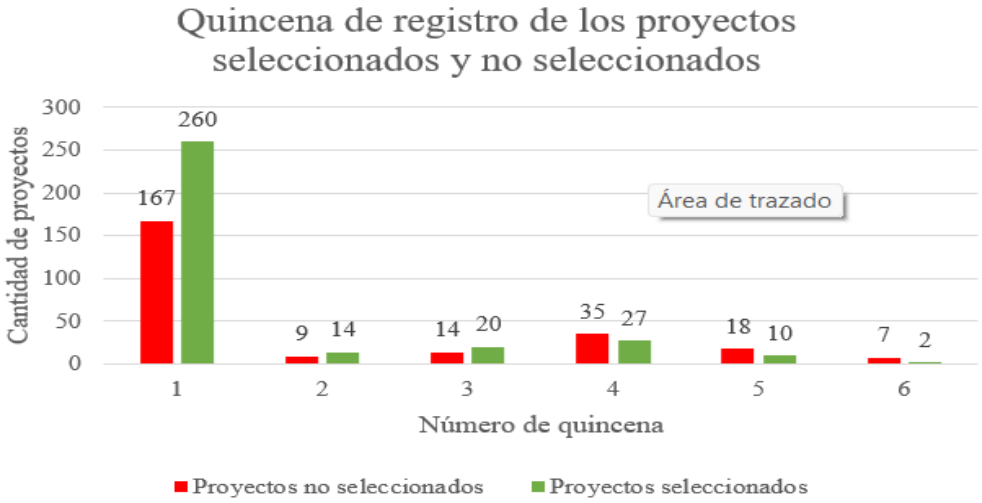


Figura 20 Gráfica de barras de la selección de proyectos registrados en cada quincena desde que salió la convocatoria

Al analizar los datos estadísticos, se encontró que los proyectos seleccionados tienen un promedio de 11 días transcurridos desde su registro, mientras que los proyectos no

seleccionados presentan un promedio significativamente mayor, con 19 días transcurridos. Estos valores se presentan de manera detallada en la Tabla 19. Estos hallazgos sugieren que existe una tendencia favorable a los proyectos que se registren en la primera quincena transcurrida desde la salida de la convocatoria.

Tabla 19 Medidas estadísticas de las vistas de los proyectos seleccionados y no seleccionados

	Valor Mínimo	*Valor Máximo*	*Promedio*	*Moda*
Proyectos Seleccionados	0	76	11	0
Proyectos No Seleccionados	0	86	19	0

Patrón: Días transcurridos y las Vistas

La relación entre los días transcurridos desde el registro y las vistas demuestra que cuanto más temprano se registra un proyecto, más tiempo permanece visible en la plataforma web, lo que influye en su selección. La gráfica de la Figura 21 muestra los proyectos seleccionados con puntos de color verde, mientras que los proyectos no seleccionados son puntos de color rojo, se observa que los proyectos no seleccionados tienden a concentrarse en el lado izquierdo y en la parte superior de la gráfica, mientras que los proyectos seleccionados están más distribuidos a lo largo de la misma. Esta dispersión sugiere que los proyectos tienden a ser seleccionados cuando transcurren menos días en sus registros y obtienen más vistas, esto se relaciona que pasan más tiempo expuestos en la plataforma web.

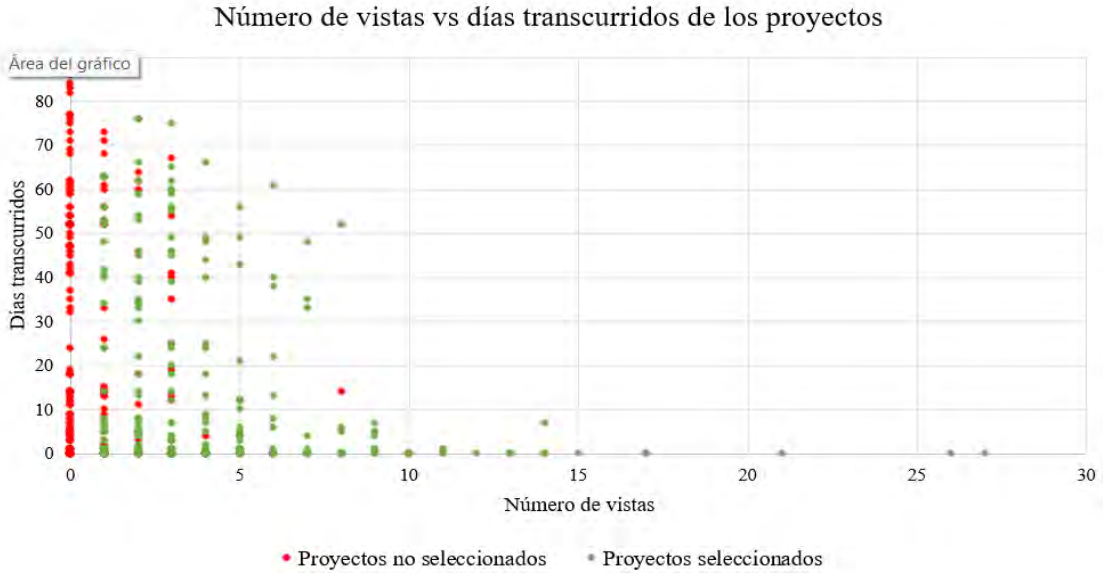


Figura 21 Gráfica de dispersión: Días transcurridos desde que salió la convocatoria vs la cantidad de vistas

Patrón: Institución de los proyectos y Selección de los proyectos

Un análisis de las variables relacionadas con la institución en la fueron propuestos los proyectos seleccionados muestra una fuerte preferencia de los alumnos por los proyectos que pertenecen a su misma institución. La gráfica de la Figura 22 muestra que 249 proyectos, lo que representa el 74.77%, fueron seleccionados por estudiantes que pertenecen a la misma institución del proyecto, mientras que solo 84 proyectos, equivalentes al 25.33%, fueron elegidos por estudiantes de otras instituciones.

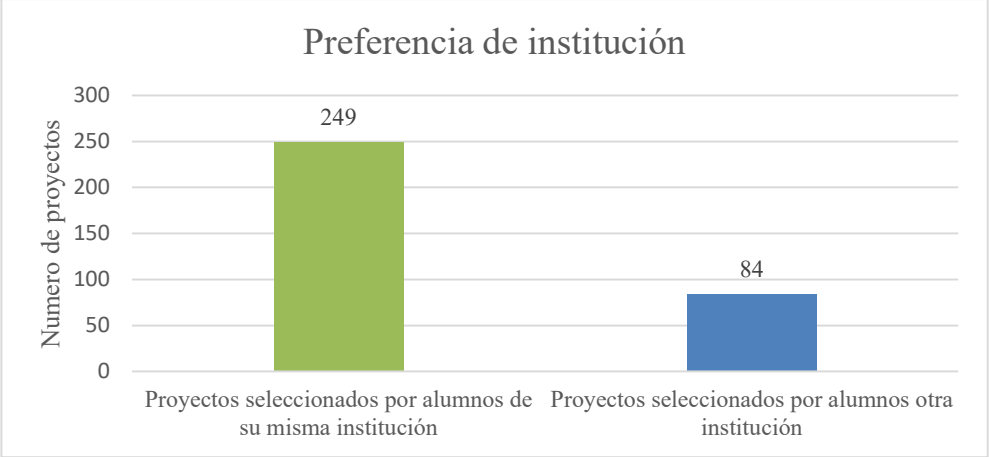


Figura 22 Gráfica de proyectos seleccionados por preferencia de institución

Este patrón está estrechamente relacionado con la modalidad en que se cursan los proyectos. La Tabla 20 muestra en porcentajes cómo se distribuyen los proyectos seleccionados por modalidad y preferencias de institución. Los proyectos seleccionados por estudiantes que pertenecen a la misma institución del proyecto tienden a ser elegidos preferentemente en la modalidad híbrida, representando el 31.83% del total de proyectos seleccionados. En contraste, los estudiantes de otras instituciones muestran una mayor preferencia por la modalidad virtual, con un 17.42% del total de proyectos seleccionados.

Tabla 20 Porcentajes de selección de proyectos por institución y modalidad

Modalidad	*Proyecto y estudiante de la misma institución*	*Proyecto seleccionado por estudiantes de otra institución*
Virtual	18.92 %	17.42 %
Presencial	24.02 %	3.00 %
Hibrida	31.83 %	4.81 %
Totales	**74.77 %**	**25.33 %**

La gráfica de mostrada en la Figura 23 muestra en cantidades la distribución de los proyectos seleccionados por temática y preferencia de institución.

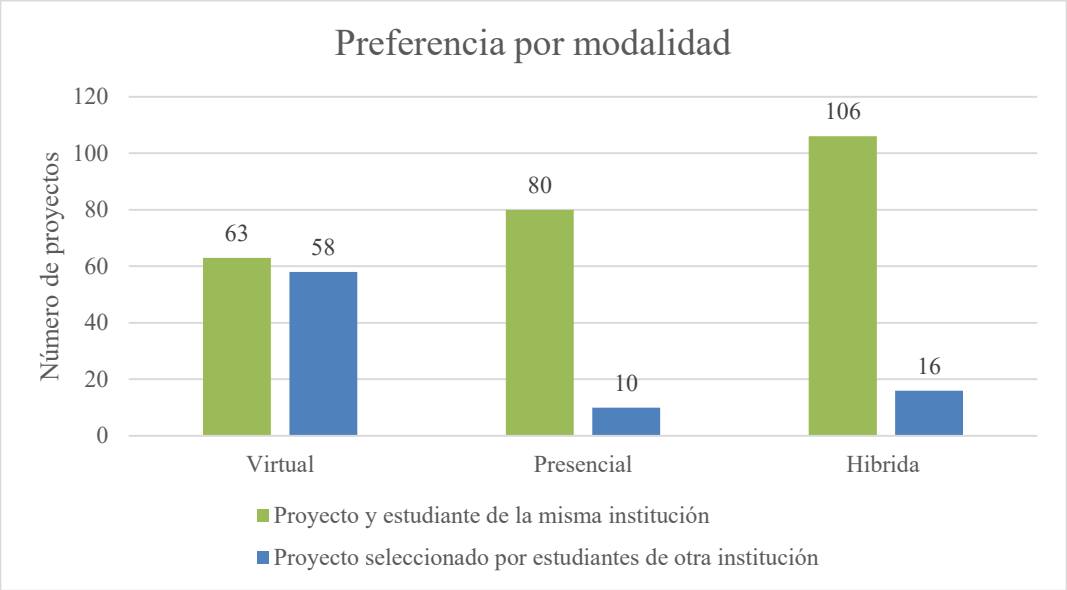


Figura 23 Preferencia de proyectos por modalidad

5.6.2 Generación de recomendaciones

I. Optimizar la visibilidad de los proyectos

Recomendación: Incrementar la visibilidad de los proyectos mediante estrategias como mejorar la interfaz de la plataforma, destacar proyectos con pocas vistas, también se sugiere proponer nuevas temáticas o nuevos filtros que faciliten la navegación dentro de la plataforma. También se sugiere programar en la plataforma un daemon que retire los proyectos de la plataforma cuando rebasan las 10 vistas y el stock del proyecto ya está lleno, es decir si un proyecto ya fue seleccionado, el número de estudiantes permitidos ya está completo y ya tiene más de 10 vistas, entonces el proyecto queda seleccionado y se retira de la plataforma, no es necesario que genere más vistas.

Justificación: Se encontró un grado de asociación considerable entre el número de vistas y la selección de proyectos (V de Cramer = 0.69), por lo que aumentar las vistas incrementará la probabilidad de que sean seleccionados.

II. Fomentar la participación en otras instituciones

Recomendación: Incentivar la participación de estudiantes en otras instituciones mediante campañas internas o modificar la plataforma, para crear un sistema de recomendación de proyectos en base a la actividad que han tenido los estudiantes, crear un hipervínculo en el

nombre del profesor, para mostrar una lista de los proyectos propuestos y un correo electrónico para que le estudiante pueda contactarlo.

Justificación: La mayoría de los proyectos fueron seleccionados por estudiantes que pertenecen a la misma institución lo que sugiere que una mayor promoción de la participación interinstitucional podría aumentar la tasa de asignación en otras instituciones.

III. Promover la creación temprana de proyectos

Recomendación: Notificar a los profesores que entre más pronto creen su proyecto, más tiempo estará visible en la plataforma del verano científico, por lo que tendrá más probabilidad de tener más vistas y de ser seleccionado.

Justificación: Los proyectos recientes tienden a ser seleccionados más rápidamente. El promedio de días transcurridos para los proyectos seleccionados es menor (11 días) comparado con los no seleccionados (19 días), lo que indica que los proyectos recientes captan más atención.

IV. Desarrollo de proyectos orientados a las áreas de interés más populares

Recomendación: Crear nuevas temáticas o subtemáticas basadas en los requisitos de conocimientos más frecuentes, de esta manera el estudiante podrá identificar cuáles son los proyectos que van más acorde a su perfil.

Justificación: La temática influye en la selección de los proyectos, sin embargo, los resultados muestran que la temática "Ingeniería e industria" acapara el mayor número de vistas y proyectos seleccionados. Dentro de esta temática no se encontraron correlaciones relevantes que puedan ayudar a clasificar si un proyecto es seleccionado o no, sin embargo, se encontró que esta temática tiene diversos proyectos, incluyendo algunos que corresponden a otras temáticas.

V. Facilitar la selección de proyectos por modalidad

Recomendación: Implementar modalidades híbridas y virtuales que faciliten la participación de estudiantes de otras instituciones. También se puede incentivar la modalidad presencial mostrando las instalaciones y equipo de laboratorio que ofrece la institución, finalmente se propone poner un filtro de modalidad en la plataforma del verano científico, esto le facilitará al estudiante la revisión de los proyectos de acuerdo a sus posibilidades.

Justificación: Se observó que el 31.83% del total de los proyectos seleccionados por estudiantes de la misma institución fueron en modalidad híbrida, mientras que los estudiantes de otras instituciones prefirieron la modalidad virtual (17.42%).

Capítulo 6. Conclusiones

Capítulo 6. Conclusiones

6.1 Conclusiones generales

En este trabajo de investigación se logró cumplir el objetivo general planteado, desarrollando una metodología que permite realizar el análisis de datos provenientes de la ejecución de procesos de negocio de manera sistemática. Esta metodología propuesta, denominada como Metodología para la mejora de indicadores clave de desempeño mediante minería de datos, fue diseñada con el propósito de identificar asociaciones significativas entre los diferentes elementos que conforman dichos procesos, como roles, áreas relacionadas e indicadores clave de desempeño, utilizando técnicas de minería de datos. Con este enfoque, se logró establecer un marco estructurado que no solo facilita la comprensión de los datos, sino que también permite detectar patrones y relaciones que podrían pasar desapercibidos, contribuyendo así a la optimización de procesos y la mejora de los indicadores clave de desempeño.

Este trabajo de investigación fue desarrollado utilizando información del evento denominado como "Verano de Ciencia y Tecnología" del CENIDET. Este caso de estudio fue fundamental para cumplir con los objetivos específicos planteados, ya que proporcionó un contexto realista y datos relevantes que facilitaron la aplicación de la metodología propuesta.

El primer objetivo específico planteado fue crear una base de datos procesada que sirviera como entrada para el análisis de datos en la ejecución de procesos de negocio. Este objetivo se logró mediante la implementación las siguientes fases de la metodología propuesta: Fase 2. Entendimiento de los datos y Fase 3. Procesamiento de los datos. En la Fase 2 se identificaron las variables disponibles y se recopiló información adicional en una nueva tabla denominada como *Bitacora_proyectos*, en la base de datos existente de la plataforma web, diseñada para registrar la actividad de los estudiantes durante su navegación. En la Fase 3 se recopilaron, filtraron y prepararon las variables para identificar las variables relevantes y con ellas crear un nuevo conjunto de datos que sirviera para el análisis, siendo este conjunto el producto resultante del primer objetivo específico. El conjunto de datos contenía información de los proyectos como la institución a la que pertenecían (variable nombrada como Id_Institución), el tiempo transcurrido desde la salida de la convocatoria hasta sus registros (variable nombrada como Quincena de registro), la cantidad de vistas que tuvo cada proyecto (variable nombrada como Rvistas), la temática de cada proyecto (variable nombrada como Temática) y la modalidad en que fueron desarrollados (variable nombrada como Modalidad).

El segundo objetivo planteado consistió en desarrollar un modelo de análisis de datos que permitiera encontrar correlaciones entre las variables relevantes. Este objetivo se logró en la Fase 4 de la metodología, correspondiente al desarrollo del modelo, utilizando un enfoque

que permitió clasificar si un proyecto sería seleccionado o no. El modelo fue implementado en el software KNIME y alcanzó una precisión del 83.24%, lo que demostró un buen desempeño con las variables seleccionadas y que los valores de las variables no son mera casualidad, sino que estas se relacionan entre si y tienen patrones que influyen en la selección de los proyectos. Entre los patrones identificados se encuentra que los proyectos seleccionados suelen caracterizarse por tener un mayor número de vistas, pertenecer mayoritariamente a la temática de Ingeniería e Industria, ser preferidos por estudiantes de la misma institución, presentarse con mayor frecuencia en modalidad híbrida y ser registrados tempranamente.

Por otro lado, el tercer objetivo específico que consiste en desarrollar un plan de pruebas se logró de manera simultánea con el primer y segundo objetivo específico. Este plan fue detalladamente documentado en el Capítulo 5: Plan de pruebas y requirió de una investigación previa, la cual se describe en el Capítulo 4: Metodología para la mejora de indicadores clave de desempeño mediante minería de datos. La preparación del plan implicó la realización de diversas pruebas utilizando distintos métodos para garantizar la obtención de los mejores resultados. El plan de pruebas siguió un enfoque sistemático, aplicando la información recopilada del caso de estudio con el propósito de validar los hallazgos obtenidos y asegurar la solidez de la investigación.

El último objetivo que consistió en documentar los pasos que componen la metodología para que pueda ser replicada fue completada al documentar la investigación realizada en el Capítulo 4: Metodología para la mejora de indicadores clave de desempeño mediante minería de datos y fue implementada simultáneamente con un caso de estudio.

La metodología propuesta presenta varias ventajas, destacándose por su naturaleza sistemática. Al estar dividida en fases claras y estructuradas, permite que cada fase, desde el entendimiento del negocio hasta la fase de interpretación y optimización, se lleve a cabo de manera ordenada. La metodología también se caracteriza por ser un proceso iterativo, lo que permite refinar y mejorar los resultados a lo largo de varias fases. Otra de las características que debe resaltarse de esta metodología es el análisis de correlaciones o asociaciones, debido a que realizar esta tarea permite obtener resultados preliminares y detectar patrones significativos. En este caso de estudio, se utilizaron métodos específicos como el análisis de Chi-cuadrado y los árboles de decisión, seleccionados en función de los datos disponibles, pero la metodología no se limita a solo a este enfoque, sino que también sugiere la aplicación del análisis de correlaciones de Pearson, lo que la convierte en una herramienta de análisis flexible.

Una de las principales fortalezas de la metodología es su fase de adquisición y preparación de datos, ya que es fundamental evaluar si se cuenta con la información necesaria para

realizar el análisis. De no ser así, se pueden realizar ajustes en los datos para garantizar su adecuación.

6.2 Trabajos futuros

Las futuras investigaciones podrían explorar las siguientes direcciones identificadas durante el estudio:

1) Automatización del proceso de extracción de los datos, preparación de datos y desarrollo del modelo

La creación de un sistema que automatice la Fase 3 y la Fase 4 de esta metodología representa un desafío interesante para futuras investigaciones. En este contexto, sería necesario explorar si es más conveniente adaptar el sistema a los procesos de negocio existentes o, por el contrario, ajustar los procesos de negocio para que se alineen con las capacidades del sistema. Esta investigación podría ofrecer una mejor comprensión de los beneficios y limitaciones de ambos enfoques, ayudando a determinar cuál genera mayor eficiencia y flexibilidad en la implementación de la metodología.

2) Refinación de la metodología

La replicación de la metodología en diferentes contextos y entornos organizacionales contribuiría significativamente al refinamiento de esta investigación, ya que permitiría ajustar y adaptar el enfoque a las necesidades y particularidades de otras organizaciones. Esta implementación en diversos escenarios ayudaría a identificar posibles áreas de mejora y a optimizar el proceso, aumentando su efectividad y aplicabilidad en distintos tipos de procesos de negocio.

6.3 Logros obtenidos

Artículo publicado

- Rebolledo, E., Estrada, H., & Martinez, A. (2023). Review of the state of the art of the integration of data mining techniques in business process analysis. TECNOLOGÍA Y CIENCIA APLICADA, 6(2), 83–88.

Ponencia

- Presentación presencial del artículo “Review of the state of the art of the integration of data mining techniques in business process analysis” en el primer Congreso Internacional de Tecnología y Ciencia Aplicada 2023.

Referencias

Bai, Y., Chen, M., Zhou, P., Zhao, T., Lee, J. D., Kakade, S., Wang, H., & Xiong, C. (2020). How Important is the Train-Validation Split in Meta-Learning? En arXiv [cs.LG]. http://arxiv.org/abs/2010.05843

Bangdiwala, S. I. (2016). Chi-squared statistics of association and homogeneity. International Journal of Injury Control and Safety Promotion, 23(4), 444–446. https://doi.org/10.1080/17457300.2016.1228144

Betancourt Velásquez, A. C., & Caviedes Niño, I. L. (2018). Metodología de Correlación estadística de un sistema integrado de gestión de la calidad en el sector salud. SIGNOS - Investigación en sistemas de gestión, 10(2), 119–139. https://doi.org/10.15332/s2145-1389.2018.0002.07

Blagus, R., & Lusa, L. (2013). SMOTE for high-dimensional class-imbalanced data. BMC Bioinformatics, 14(1). https://doi.org/10.1186/1471-2105-14-106

Bose, I., & Mahapatra, R. (2001). Business data mining - A machine learning perspective. ELSEVIER, 211-225.

Bradley, P. S., Bennett, K. P., & Demiriz, A. (2000). Constrained k-means clustering. Microsoft Research, Redmond, 20(0).

Camacho-Sandoval, J. (2008). Asociación entre variables cuantitativas: análisis de correlación.http://creativecommons.org/licenses/by-nc-nd/3.0/. http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S0001600220080002000005&lng=en&tlng=es.

Camacho-Sandoval, J. (2008). Asociación entre variables cuantitativas: análisis de correlación.http://creativecommons.org/licenses/by-nc-nd/3.0/. http://www.scielo.sa.cr/scielo.php?script=sci_arttext&pid=S0001600220080002000005&lng=en&tlng=es.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. The journal of artificial intelligence research, 16, 321–357. https://doi.org/10.1613/jair.953

Coll, A. M., Rodríguez, A. M., Gutiérrez, D. M., Moreno, A. S. P., & Llanes-Santiago, O. (2022). Propuesta de metodología para el diagnóstico de fallos basado en árboles de decisión y lógica difusa. Ingeniería Electrónica, Automática y Comunicaciones, 43(2), 1-16.

Contreras, L. E., Tarazona Bermúdez, G. M., & Alemán Cardona, A. P. (2023). Machine Learning aplicado al rendimiento académico en educación superior: factores, variables y herramientas. Editorial Universidad Distrital Francisco José de Caldas.

Cortina, V. G. (2015). Aplicación de la metodología crisp-dm a un proyecto de minería de datos en el entorno universitario. Trabajo de Fin de Grado (Universidad Carlos III de Madrid).

Dahouda, M. K., & Joe, I. (2021). A deep-learned embedding technique for categorical features encoding. IEEE access: practical innovations, open solutions, 9, 114381–114391. https://doi.org/10.1109/access.2021.3104357

Daoudagh, S., Marchetti, E., Calabrò, A., Ferrada, F., Oliveira, A. I., Barata, J., Peres, R., & Marques, F. (2023). DAEMON: A domain-based monitoring ontology for IoT systems. SN Computer Science, 4(5). https://doi.org/10.1007/s42979-023-01975-y

Dara, R., Research Scholar, Department of Computer Science and Engineering Jawaharlal Nehru Technological University Kakinada, Andhra Pradesh, India, Satyanarayana, C., & Govardhan, A. (2016). A novel approach for data cleaning by selecting the optimal data to fill the missing values for maintaining reliable data warehouse. International journal of modern education and computer science, 8(5), 64–70. https://doi.org/10.5815/ijmecs.2016.05.08

Delgado, A., Calegari, D., Marotta, A., González, L., & Tansini, L. (2021). A methodology for integrated process and data mining and analysis towards evidence-based process improvement. Proceedings of the 16th International Conference on Software Technologies.

Dijkman, R. M., Dumas, M., & Ouyang, C. (2008). Semantics and analysis of business process models in BPMN. Information and Software Technology, 50(12), 1281–1294. https://doi.org/10.1016/j.infsof.2008.02.006

Djenouri , Y., Belhadi, A., & Fournier-Viger, P. (2018). Extracting Useful Knowledge from Event Logs: A Frequent Itemset Mining Approach. ELSEVIER, 132-148.

Doumas, M., La Rosa, M., Mendling , L., & Reijers, H. (2013). Fundamentals of Business Process Management. Springer. doi:10.1007/978-3-642-33143-5

Elhadjamor, E. (2019). Analyze in depth health care business process and key performance indicators using process minig. ELSEVIER, 610-617.

Elhadjamor, E. A., & Ghannouchi, S. A. (2017). Towards managing key performance indicators for measuring business process performance. En Information Systems (pp. 579–591). Springer International Publishing. environmental science, 729(1), 012133. https://doi.org/10.1088/1755-1315/729/1/012133

Elkabalawy, M., Al-Sakkaf, A., Mohammed Abdelkader, E., & Alfalah, G. (2024). CRISP-DM-based data-driven approach for building energy prediction utilizing indoor and environmental factors. Sustainability, 16(17), 7249. https://doi.org/10.3390/su16177249

Espinosa Zúñiga, J. J. (2020). Aplicación de metodología CRISP-DM para segmentación geográfica de una base de datos pública. Ingeniería Investigación y Tecnología, 21(1), 1–13. https://doi.org/10.22201/fi.25940732e.2020.21n1.008

Galvis Lista, E. A., & González Zabala, M. P. (2014). Herramientas para la gestión de procesos de negocio y su relación con el ciclo de vida de los procesos de negocio: una revisión de literatura. Ciencia e ingeniería neogranadina, 24(2), 37. https://doi.org/10.18359/rcin.392

Géron, A. (2022). Hands-on machine learning with scikit-learn, keras, and tensorflow (3a ed.). O'Reilly Media.

Giraldo, J., Jiménez, J., & Montoya, D. (2015). Integración de BPM y Minería de Datos para la optimización de Indicadores Clave de Proceso KPI. doi:10.13140/RG.2.1.3174.6088

Giraldo, J., Jimenez, J., & Tabares, M. (2017). Modelo para optimizar el proceso de gestión de negocio combinando minería de procesoscon inteligencia de negocios desde almacenes de datos. ESPACIOS, 38, 9-19.

González, Y. D., & Romero, Y. F. (2012). Patrón Modelo-Vista-Controlador. Telemática, 11(1), 47-57.

Gras, J. (2019). Modelando el negocio con BPM en el mundo real. BPMteca.com.
Hernández, A. B., Torres, F. V., & México, C. D. M. X. (2019). Monitoreo de aplicaciones web. En Emprendiendo Innovaciones con Tecnologías Exponenciales.

Hitpass, B. (2017). BPM: Business Process Management Fundamentos y Conceptos de Implementación (Cuarta ed.). Santiago Chile: BHH-Ltda.

Huber, S., Wiemer H, Shneider, D., & Ihlenfeldt, S. (2019). DMME: Data mining methodology for engineering applications – a holistic extension to the CRISP-DM model. ELSEVIER, 403-408.

IBM. (2021). Guía de CRISP-DM de IBM SPSS Modeler. IBM Corporation.

Javaheri, S. H., Sepehri, M. M., & Teimourpour, B. (2014). Response Modeling in Direct Marketing. En Data Mining Applications with R (pp. 153–180). Elsevier.

Jlailaty, D., Grigori, D., & Belhajjame, K. (2017). Mining business process activities from email logs. 2017 IEEE International Conference on Cognitive Computing (ICCC).

Kannengiesser, U., & Gero, J. S. (2023). Modelling the design of models: An example using crisp-dm. Proceedings of the Design Society, 3, 2705–2714. https://doi.org/10.1017/pds.2023.271

Kearney, M. W. (2017). Cramér's V. The SAGE encyclopedia of communication research methods. Sage Publications.

Lalinde, J. D. H., Castro, F. E., Rodríguez, J. E., Rangel, J. G. C., Sierra, C. A. T., Torrado, M. K. A., & Pirela. (2018). Sobre el uso adecuado del coeficiente de correlación de Pearson: definición, propiedades y suposiciones. Archivos venezolanos de Farmacología y Terapéutica, 37, 587–595.

Lup Low, W., Li Lee, M., & Wang Ling, T. (2001). A knowledge-based approach for duplicate elimination in data cleaning. Information Systems, 26(8), 585–606. https://doi.org/10.1016/s0306-4379(01)00041-2

Mamián, Y. L. Z., Palta, K. F. M., Cobo, N. I. M., & Castillo, S. (2019). APLICACIÓN DE MÉTODOS DE SELECCIÓN DE ATRIBUTOS PARA DETERMINAR FACTORES RELEVANTES EN LA CLASIFICACIÓN DE ACTIVIDADES FÍSICAS COTIDIANAS. Encuentro Internacional de Educación en Ingeniería.

Midden, J. H. (2021). Using process mining and event log analysis for better business strategy decision-making (Bachelor's thesis).

Moine, J. M., Haedo, A. S., & Gordillo, S. E. (2011). Estudio comparativo de metodologías para minería de datos. XIII Workshop de Investigadores en Ciencias de la Computación.

Ncr, P. C., Clinton, J., Ncr, R. K., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (1999). CRISP-DM 1.0.

Nuñez-Arcia, Yaisel, Díaz-de-la-Paz, Lisandra, & García-Mendoza, Juan Luis. (2016). Algorithm to correct instance level anomalies in large volumes of data using MapReduce. Revista Cubana de Ciencias Informáticas, 10(3), 105-118. Recuperado en 01 de diciembre de 2024, de http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S2227-18992016000300008&lng=es&tlng=en.

Orlando, D. P., & Giovanni, F. (2008). An integrated system, with natural language management, for the monitoring activities in e-learning environments. 2008 International Conference on Complex, Intelligent and Software Intensive Systems.

Rahm, E., & Do, H. H. (2000). Data cleaning: Problems and current approaches. IEEE Data Eng. Bull, 23(4), 3–13.

Ramírez-Gallego, S., García, S., Mouriño-Talín, H., Martínez-Rego, D., Bolón-Canedo, V., Alonso-Betanzos, A., Benítez, J. M., & Herrera, F. (2016). Data discretization: taxonomy and big data challenge: Data discretization. Wiley Interdisciplinary Reviews. Data Mining and Knowledge Discovery, 6(1), 5–21. https://doi.org/10.1002/widm.1173

Recker, J. (2010). Opportunities and constraints: the current struggle with BPMN. Business Process Management Journal, 16(1), 181–201. https://doi.org/10.1108/14637151011018001

Rengifo, H. F. C. (2007). Estrategias para la detección y corrección automática de errores en fuentes de datos. https://doi.org/10.13140/RG.2.2.21896.62723

Riquelme Santos, J. C., Ruiz, R., & Gilbert, K. (2006). Minería de datos: Conceptos y tendencias. Inteligencia Artificial: Revista Iberoamericana de Inteligencia Artificial, 10 (29), 11-18.

Ruggieri, S. (2002). Efficient C4.5 [classification algorithm]. IEEE transactions on knowledge and data engineering, 14(2), 438–444. https://doi.org/10.1109/69.991727

Salcedo Parra, O. J., Galeano, R. M., & Rodriguez, L. G. (2010). Metodología crisp para la implementación Data Warehouse. Tecnura, 14(26), 35–48.

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181, 526–534. https://doi.org/10.1016/j.procs.2021.01.199

Siek, M., & Mukti, R. M. G. (2021). Business process mining from e-commerce event web logs: Conformance checking and bottleneck identification. IOP conference series. Earth and

Silberschatz, A. (2007). Fundamentos de Bases de Datos. McGraw-Hill Interamericana.

SYDLE. (2022). *Guía completa del modelado a la automatización de procesos.*

Vanrell, J. A., Bertone, R., & Martínez, R. (2010). Un modelo de procesos de explotación de la información. XII Workshop de Investigadores en Ciencias de la Computación, 167–171. http://sedici.unlp.edu.ar/handle/10915/19462

Wegener, D., & Rüping , S. (2010). On Integrating Data Mining into Business. BUSINESS INFORMATION SYSTEMS, 47, 183-194.

White, S. A., & Miers, D. (2009). Guía de referencia y modelado BPMN. Future Strategies.

Zhang, S., Zhang, C., & Yang, Q. (2003). Data preparation for data mining. Applied Artificial Intelligence: AAI, 17(5–6), 375–381. https://doi.org/10.1080/713827180

Anexos

Anexo A. Diccionario de las variables de la base de datos del caso de estudio.

user		
Nombre la columna	**Tipo de dato**	**Atributos de la base de datos**
id	bigint unsigned	Enumeración de los usuarios registrados en el verano científico
name	varchar(255)	Nombre del usuario
falastname	varchar(255)	Apellido paterno del usuario
molastname	varchar(255)	Apellido materno del usuario
email	varchar(255)	Correo electrónico
password	varchar(255)	Contraseña de la cuenta en el verano científico
profile_photo_path	text	Foto de perfil
token	varchar(100)	Token identificador
status	int	Cuenta activa o inactiva
id_role	int	Rol del usuario:
id_institution	int	ID de la institución a la que pertenece el usuario
curp	varchar(50)	CURP
emailalt	varchar(100)	email secundario
movil	varchar(50)	Número de teléfono celular
telefono	varchar(50)	Número de teléfono secundario
rfc	varchar(50)	RFC
nivel_sni	varchar(50)	Nivel de SNI
grado_maximo	varchar(50)	Grado académico
area_conocimiento	varchar(100)	Temática
sector_conocimiento	varchar(100)	Subtemática
area_prioritaria	varchar(200)	Área
puesto	varchar(200)	Puesto de trabajo
id_thematic	int	ID de la tematica
id_subtematica	int	ID de la subtematica
created_at	timestamp	Marca de tiempo en la que se registró
updated_at	timestamp	Marca de tiempo en que se actualizó
deleted_at	timestamp	Marca de tiempo de eliminación de usuario

Role		
Nombre de la columna	**Tipo de dato**	**Atributos de la base de datos de los roles**
id_rol	bigint unsigned	ID del rol
name	varchar(255)	Nombre del rol

status	int	Estado activo e inactivo
created_at	timestamp	Marca de tiempo de creación del rol
updated_at	timestamp	Marca de tiempo actualizada
deleted_at	timestamp	Marca de tiempo de rol eliminado

project		
Nombre de la columna	**Tipo de dato**	**Comment**
id	bigint unsigned	Número del proyecto registrado
name	varchar(255)	Nombre del proyecto
description	text	Descripción del proyecto
image_path	varchar(100)	Imagen adjuntada del proyecto
objective	text	Objetivo del proyecto
pre_requisites	text	Prerequisitos que debe cumplir el alumno
requisites	text	Requisitos que debe cumplir el alumno
modalidad	text	Modalidad en la que se va a trabajar el proyecto
status	int	Estado activo o inactivo
stock_alumnos	int	Números de alumnos permitidos para trabajar en los proyectos
id_institution	int	ID de la institución
id_thematic	int	ID de la temática
id_subtematica	int	ID de la subtemática
id_user	int	ID del usuario
keyword	text	Palabras clave
created_at	timestamp	Marca de tiempo de la creación del poyecto
updated_at	timestamp	Marca de tiempo de actualizado
deleted_at	timestamp	Marca de tiempo de la eliminación del proyecto

user_expediente		
Nombre de la columna	**Tipo de dato**	**Comment**
id	bigint unsigned	Numeración en la base de datos
id_user	int	ID del usuario
id_expediente	int	ID de los documentos del expediente: 1-Solicitudes de ingreso/2-Constancia de cursamiento/3-Carta motivos/4-Fotogafía/5-CVU esudiante
status	int	Estado del alumno: 5,Aceptado/4,Alumno rechazado/3,Documentación enviada y aceptada/2,Expediente enviado a evluación/1,Cambio de proyecto
observaciones	text	Indica el estado del expediente en letra
created_at	timestamp	Marca de tiempo de cuando fue evaluado
updated_at	timestamp	Marca de tiempo en la que fue actualizado el expediente

deleted_at	timestamp	Marca de tiempo en la que fue borrado el expediente
id_evaluador	int	ID del evaluador

user_expediente_details		
Nombre de la columna	**Tipo de dato**	**Comment**
id	int	Numeración
id_user_expediente	int	ID del expediente del usuario
id_expediente	int	ID de los documentos del expediente: 1-Solicitudes de ingreso/2-Constancia de cursamiento/3-Carta motivos/4-Fotogafía/5-CVU esudiante
path_exp	varchar(100)	Ruta del expediente
name_exp	varchar(100)	Nombre del documento subido
observaciones	varchar(100)	Observaciones
status	varchar(45)	Estado activo
created_at	timestamp	Marca de tiempo en que fueron subidos los documentos
updated_at	timestamp	Marca de tiempo de actualización de los documentos
deleted_at	timestamp	Marca de eliminación de los documentos

user_project		
Nombre de la columna	**Tipo de dato**	**Comment**
id	int	Numeración de la tabla
id_user	int	ID del usuario
id_project	int	ID del proyecto
sequence	int	Secuencia de prioridad de los proyectos
status	int	1 Rechazado o 5 aceptado
prioridad	int	Prioridad de los proyectos seleccionados
created_at	timestamp	Marca de tiempo en que los proyectos fueron seleccionados
updated_at	timestamp	Marca de tiempo en que los proyectos fueron seleccionados

thematic		
Nombre de la columna	**Tipo de dato**	**Comment**
id	bigint unsigned	ID de la temática
name	varchar(255)	Nombre de la temática

id	Temática
1	Física, Matemáticas y Ciencias de la Tierra
2	Biología y Química
3	Medicina y Salud
4	Biotecnología y Ciencias Agropecuarias
5	Tecnología 4.0
6	Ingeniería e Industria
7	Ciencias Sociales
8	interdiciplinaria (litio)

subthematic		
Nombre de la columna	**Tipo de dato**	**Comment**
id	int	ID de la subtemática
name	varchar(100)	Nombre de la subtemática
id_thematic	int	ID de la temática

Subtemáticas		
id	**Subtematica**	**id_tematica**
1	Astrofísica	1
2	Astronomía	1
3	Ciencias de la Tierra y del espacio	1
4	Física	1
5	Lógica	1
6	Matemáticas	1
7	Prospectiva	1
8	Ciencias del mar	1
9	Ciencias de la vida	2
10	Prospectiva	2
11	Ciencias biomédicas	2
12	Biología	2
13	Química	2
14	Ciencias Ambientales	2
15	Ciencias médicas	3
16	Medicina	3
17	Enfermería	3
18	Ingeniería biomédica	3
19	Especialidad médica	3
20	Odontología	3
21	Investigación médica	3
22	Biotecnología	4
23	Ciencias agrarias	4
24	Biotecnología agrícola	4
25	Salud y Producción Animal	4

26	Pesca	4
27	Robótica	5
28	Simulaciones	5
29	Materiales avanzados	5
30	Realidad virtual	5
31	Realidad aumentada	5
32	Big Data	5
33	Inteligencia Artificial	5
34	Software como servicio	5
35	Manufactura aditiva	5
36	Ciencias tecnológicas	6
37	Ingeniería	6
38	Prospectiva	6
39	Ciencia política	7
40	Ciencias de la educación	7
41	Ciencias económicas	7
42	Ciencias jurídicas y derecho	7
43	Demografía	7
44	Formación docente	7
45	Geografía	7
46	Historia	7
47	Sociología	7
48	Prospectiva	7
49	Administración y negocios	7
50	Medios de comunicación y comunicaciones	7
51	Comunicación Científica	7
52	Gestión	7
53	Estudios de Género	7
54	Evaluación de reservas de litio	8
55	Innovación en la explotación y transformación del litio	8
56	Almacenamiento de energía	8
57	Aplicaciones	8
58	Comercialización, logística y transporte del litio	8

Anexo B. Código de la funcionalidad de eventos

Vistas

Algoritmo A

```
function registroBitacora(id) {
let id_proyecto = id;
localStorage.setItem('id_proyecto', id_proyecto);
// Obtener la fecha y hora actual
var fechaHora = new Date();
var fecha_in = fechaHora.toISOString().slice(0, 10);
var hora_in = fechaHora.toLocaleTimeString();

// Construir los datos a enviar
let url = '/Project/InsertBitacora';
let formData = new FormData();
formData.append("id_project", id_proyecto);
formData.append("fecha_in", fecha_in);
formData.append("hora_in", hora_in);
invokeAjaxCall(url, formData, "POST", cbInsertBitacora);
}
```

Algoritmo B

```
function registroBitacora_out() {
var id_proyecto = localStorage.getItem('id_proyecto');
// Obtener la fecha y hora actual
var fechaHora = new Date();
var fecha_out = fechaHora.toISOString().slice(0, 10);
var hora_out = fechaHora.toLocaleTimeString();
// Construir los datos a enviar
let url = '/Project/UpdateBitacora';
let formData = new FormData();
formData.append("id_project", id_proyecto);
formData.append("fecha_out", fecha_out);
formData.append("hora_out", hora_out);
invokeAjaxCall(url, formData, "POST", cbUpdateBitacora);

localStorage.removeItem('id_proyecto');
}
function cbInsertBitacora(info){
alert(info);
}
function cbUpdateBitacora(info){
alert(info);
}
```

Controladores

Algoritmo A

```
public function InsertBitacora()
{
$ip_user = $_SERVER['REMOTE_ADDR'];
$id_project = intval($_POST["id_project"]);
$fecha = $_POST["fecha_in"];
$hora = $_POST["hora_in"];
$id_user = $_SESSION['id_User'];
$arrData = $this->model->IBitacora($ip_user, $id_user,
$id_project, $fecha, $hora);
if ($arrData) {
$arrResponse = array(
'status' => true,
'datos' => json_encode($arrData, JSON_UNESCAPED_UNICODE)
);
} else {
$arrResponse = array(
'status' => false,
'datos' => json_encode(array(), JSON_UNESCAPED_UNICODE)
);
}
echo json_encode($arrResponse, JSON_UNESCAPED_UNICODE);
die();
}
```

Algoritmo B

```
public function ActBitacora()
{
$ip_user = $_SERVER['REMOTE_ADDR'];
$id_project = intval($_POST["id_project"]);
$fecha_out = $_POST["fecha_out"];
$hora_out = $_POST["hora_out"];

$id_user = $_SESSION['id_User'];
$arrData = $this->model->ABitacora($id_project, $ip_user, $id_user,
$fecha_out, $hora_out);
if ($arrData) {
$arrResponse = array(
'status' => true,
'datos' => json_encode($arrData, JSON_UNESCAPED_UNICODE)
);
} else {
$arrResponse = array(
'status' => false,
'datos' => json_encode(array(), JSON_UNESCAPED_UNICODE)
);
}
```

```
echo json_encode($arrResponse, JSON_UNESCAPED_UNICODE);
die();
}
```

Modelo

Algoritmo A

```
public function IBitacora($ip_user, $id_user, $id_project, $fecha,
$hora)
{
$status = 1;
$arrData = array($ip_user, $id_user, $id_project, $fecha, $hora);
$query = "INSERT INTO Bitacora(ip_user, id_user,
id_project, Fecha_in, Hora_in) VAlUES (?,?,?,?,?)";
$request_insert = $this->insert($query, $arrData);
return $request_insert;}
```

Algoritmo B

```
public function ABitacora($id_project, $ip_user, $id_user, $fecha_out,
$hora_out)
{$query = "UPDATE bitacora_proyectos
SET
Fecha_out = ?,
Hora_out = ?
WHERE
id_user = ? AND id_project = ?";
$request_update = $this->update($query, array($fecha_out, $hora_out,
id_user, $id_project));
return $request_update;}
```