



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN**

TEMA:

Desarrollo de un prototipo de sistema web basado en machine learning para la identificación de tendencias y análisis de incidentes de la empresa de seguridad privada PROTEMAXI.

AUTORES:

**Pazos Celi, Andrés Sebastián
Vega Doylet, Priscilla Rebeca**

**Trabajo de integración curricular previo a la obtención del título de
INGENIERO EN CIENCIAS DE LA COMPUTACIÓN**

TUTOR:

Ing. Yong Yong, Byron Severo

**Guayaquil, Ecuador
Día 17 de febrero de 2025**



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

CERTIFICACIÓN

Certificamos que el presente trabajo de integración curricular fue realizado en su totalidad por el Sr. **Pazos Celi, Andrés Sebastián** y la Srta. **Vega Doylet, Priscilla Rebeca** como requerimiento para la obtención del título de **Ingeniero en Ciencias de la Computación**.

TUTOR

f. _____

Ing. Yong Yong, Byron Severo

Guayaquil, a los 17 días del mes de febrero del año 2025



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

DECLARACIÓN DE RESPONSABILIDAD

Nosotros, **Pazos Celi, Andrés Sebastián**
Vega Doylet, Priscilla Rebeca

DECLARAMOS QUE:

El Trabajo de Integración curricular, “**Desarrollo de un prototipo de sistema web basado en machine learning para la identificación de tendencias y análisis de incidentes de la empresa de seguridad privada PROTEMAXI**”, previo a la obtención del título de **Ingeniero en Ciencias de la Computación**, ha sido desarrollado respetando derechos intelectuales de terceros conforme las citas que constan en el documento, cuyas fuentes se incorporan en las referencias o bibliografías. Consecuentemente este trabajo es de nuestra total autoría.

En virtud de esta declaración, nos responsabilizamos del contenido, veracidad y alcance del Trabajo de Integración Curricular referido.

Guayaquil, a los 17 días del mes de febrero del año 2025

EL AUTOR

EL AUTOR

f.

Pazos Celi, Andrés Sebastián

f.

Vega Doylet, Priscilla Rebeca



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

AUTORIZACIÓN

Nosotros, **Pazos Celi, Andrés Sebastián**
Vega Doylet, Priscilla Rebeca

Autorizamos a la Universidad Católica de Santiago de Guayaquil a la **publicación** en la biblioteca de la institución el Trabajo del Trabajo de Integración Curricular, “**Desarrollo de un prototipo de sistema web basado en machine learning para la identificación de tendencias y análisis de incidentes de la empresa de seguridad privada PROTEMAXI**”, cuyo contenido, ideas y criterios son de nuestra exclusiva responsabilidad y total autoría.

Guayaquil, a los 17 días del mes de febrero del año 2025

EL AUTOR

EL AUTOR

f.

Pazos Celi, Andrés Sebastián

f.

Vega Doylet, Priscilla Rebeca



UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN

REPORTE COMPILATIO



CERTIFICADO DE ANÁLISIS
magister

RS_Trabajo de Titulación ICC
2025-01-30

< 1%
Textos
sospechosos

Nombre del documento: RS_Trabajo de Titulación ICC 2025-01-30.docx
ID del documento: 28efaed6885cbf9d251f7c7c4378b90d53d23a73
Tamaño del documento original: 5,94 MB
Autores: []

Depositante: Byron Severo Yong Yong
Fecha de depósito: 1/2/2025
Tipo de carga: interface
fecha de fin de análisis: 1/2/2025

TUTOR

Byron Yong Yong

鄺勇

f.

Ing. Yong Yong Byron Severo

Firmado digitalmente por BYRON SEVERO
YONG YONG.
Nombre de reconocimiento (DN): c=EC,
ou=BANCO CENTRAL DEL ECUADOR,
ou=ENTIDAD DE CERTIFICACION DE
INFORMACION-ECBICE, l=QUITO,
serialNumber=1000750033, cn=BYRON
SEVERO YONG YONG.
Fecha: 2025.02.01 13:37:28 -05'00'

AGRADECIMIENTO

Agradezco a Dios y a mi Familia, porque con ellos he corrido una vida entera llena de altos y bajos por mi camino Universitario. A Dios por haberme dado bienestar y salud en los momentos que más lo necesite para poder llegar hasta este punto de mi vida. A mi Padre por siempre ser esa persona que me ha apoyado económicamente y ser una imagen de superación a lo largo de mi etapa universitaria. A mi madre y a mi hermana por enseñarme a ser un mejor hijo, hermano, compañero y persona a través de sus consejos y lecciones de vida.

Agradezco a todos mis compañeros y al personal de la facultad de ingeniería, por brindarme su apoyo y conocimiento a lo largo de cada año. A mis compañeros por haberme brindado su ayuda y por las experiencias que hemos compartido en cada momento de la carrera. Al personal de la facultad por haberme mostrado las oportunidades y conocimientos que esta carrera puede ofrecer; por haberme enseñado y guiado a esta última etapa de mi carrera universitaria. Que todas las enseñanzas y experiencias vividas sean de utilidad para lograr nuestras metas y objetivos en nuestro campo profesional y personal.

Sebastián Pazos

Ante todo, siempre agradecerle a Dios, por haberme guiado y dado fuerzas para seguir adelante durante todo el camino que he recorrido. A mi familia, que siempre me han dado un apoyo incondicional y siempre llenándome de mucho amor. A mis amigos por haberme dado la mano en los momentos difíciles que he tenido que atravesar, eternamente agradecida.

Priscilla Vega

DEDICATORIA

Dedico este trabajo a mis padres y mis abuelos, ya que han sido pilares fundamentales en toda mi vida tanto en lo bueno como en lo malo. A mis padres, por su gran apoyo emocional y económico y por reflejarme esfuerzo, dedicación y superación. La dedicatoria también va a mis abuelos que con mucho esfuerzo han construido una gran familia que es la que tengo hoy demostrándome que muchas veces solo se necesita tener una actitud y valores para lograr cosas inimaginables.

Este trabajo va dedicado a todas las personas que han formado parte de mi vida en diferentes etapas durante toda mi etapa universitaria; amigos, compañeros, docentes y todos aquellos con los que haya recibido una clase o conocido ya que con su apoyo y confianza me han hecho ver que la etapa universitaria es una etapa de mi vida muy de admirar.

Sebastián Pazos

Para ti mamá. A pesar de que no estas físicamente conmigo, sé que siempre me estas cuidando. Gracias por siempre creer en mí, por todo el amor y apoyo incondicional que me diste. Eres mi mayor inspiración para todo lo que hago y todos mis logros te los dedicaré a ti.

Priscilla Vega



**UNIVERSIDAD CATÓLICA
DE SANTIAGO DE GUAYAQUIL
FACULTAD DE INGENIERÍA
CARRERA DE INGENIERÍA EN CIENCIAS DE LA COMPUTACIÓN**

TRIBUNAL DE SUSTENTACIÓN

f. _____

**Ing. Camacho Coronel, Ana Mgs
DIRECTORA DE CARRERA**

f. _____

**Ing. Cornejo Gómez, Galo Enrique, Mgs
DOCENTE DE LA CARRERA**

f. _____

**Ing. Erazo Ayón, José Miguel, Mgs
OPONENTE**

ÍNDICE GENERAL

Resumen.....	XV
Abstract.....	XVI
Introducción.....	2
Capítulo I: El Problema	4
ANÁLISIS DEL PROBLEMA.	4
Ubicación del Problema en un Contexto	4
Delimitación del Problema	6
Formulación del Problema	6
OBJETIVOS	7
Objetivo General.....	7
Objetivos Específicos.....	7
ALCANCES DEL PROBLEMA	8
JUSTIFICACION E IMPORTANCIA	9
HIPÓTESIS O PREGUNTA DE INVESTIGACIÓN	11
VARIABLES DE LA INVESTIGACIÓN	11
Capítulo II: Marco Teórico	13
2.1 Sistema Web	13
2.2 Automatización Inteligente	15
2.3 Software	18
2.3.1 Machine Learning	18
2.3.2 Lenguajes de Programación	24
2.3.3 Lenguajes de programación: Python	25
2.3.4 Framework: Laravel.....	26
2.3.5 Blade	27
2.3.6 Bootstrap	27
2.3.7 Librerías: FPDF	28
2.3.8 Frontend y Backend	29
2.3.9 Base de Datos	29
2.3.10 Base de Datos: MYSQL	30
2.3.11 API.....	30
2.3.12 Entorno de Desarrollo Integrado	30

2.3.13 Plataforma de control de versiones.....	34
Capítulo III: Metodología De La Investigación.....	36
Tipo Investigación y enfoque metodológico.....	36
Población y muestra	37
Instrumentos de recolección de datos	39
Metodología de desarrollo de prototipo	40
Capítulo IV: Propuesta Tecnológica.....	43
Levantamiento de información	43
Definición del prototipo	45
Características y funcionalidades	45
Requisitos funcionales y técnicos	45
Herramientas para el desarrollo.....	46
Arquitectura del sistema	46
Modelo entidad relación	48
Proceso de extracción y carga de datos a la base	48
Creación de proyecto de Laravel.....	50
Configuración del archivo.env	53
Creación de Modelos MVC.....	54
Migrations.....	54
Uso del XAMP	56
Creación de Controlador	58
Carpeta de Imágenes	58
Vista con Blade	58
Instalación de la librería FPDF	59
Construcción del Reporte en el Controlador.....	60
Ejecución del proyecto	60
Desarrollo del modelo de Machine Learning	61
Comparación y elección del tipo de modelo	61
Learning Seleccionado: Gradient Boosting Regressor	68
Estructura y explicación del modelo.....	70
Configuración de KeyWords y niveles de severidad	73
Creación de la tabla recomendaciones.....	74
Obtención de Datos	75
Detección de keywords.....	76

Extracción de características	77
Preprocesamiento de los Datos	78
Clasificar severidad y frecuencia	79
Disponibilidad de Resultados	84
Diagrama de flujo del prototipo.....	86
Resultados	87
Informe generado.....	87
Precisión del modelo de Machine Learning	92
Tiempo de generación de informe	96
Recomendaciones.....	99
Referencias Bibliográficas	100
Anexos	105
Anexo 1. Preguntas de la entrevista de levantamiento de información	105
Anexo 2. Levantamiento de información	105
Anexo 3. Archivo Informe generado	105

INDICE DE IMÁGENES

FIGURA 1	ÁRBOL DEL PROBLEMA	5
FIGURA 2	ESQUEMA DE INTERACCIÓN PROTOTIPO.....	8
FIGURA 3	LÍNEA DE TIEMPO HISTORIA Y EVOLUCIÓN MACHINE LEARNING	21
FIGURA 4	FASES DE IMPLEMENTACIÓN ML	23
FIGURA 5	ETIQUETA EN HTML.....	26
FIGURA 6	FRONTEND VS BACKEND	29
FIGURA 7	FUNCIONAMIENTO DE API	30
FIGURA 8	METODOLOGÍA CASCADA	41
FIGURA 9	ARQUITECTURA DEL PROTOTIPO	47
FIGURA 10	METODOLOGÍA CASCADA	48
FIGURA 11	REGISTRO DE INCIDENCIAS	49
FIGURA 12	EJEMPLO DE INFORME	50
FIGURA 13	COMANDO DE CREACIÓN.....	51
FIGURA 14	EJECUCIÓN.....	51
FIGURA 15	INSTALACIÓN DE PAQUETES Y RECURSOS	52
FIGURA 16	CREACIÓN DE ARCHIVOS Y DIRECTORIOS.....	52
FIGURA 17	FINALIZACIÓN DE CREACIÓN DEL PROYECTO Y MIGRACIÓN DE TABLAS A LA BD	53
FIGURA 18	CONFIGURACIÓN DEL ARCHIVO.ENV	53
FIGURA 19	CREACIÓN DE MODELO DETALLE	54
FIGURA 20	CREACIÓN DE MODELO RECOMENDACIÓN.....	54
FIGURA 21	CONTENIDO DE MIGRATIONS	55
FIGURA 22	SCHEMA TABLA DETALLES	55
FIGURA 23	SCHEMA TABLA RECOMENDACIONES	56
FIGURA 24	EJECUCIÓN MIGRATE	56
FIGURA 25	PANEL DE CONTROL XAMPP	57
FIGURA 26	BASE DE DATOS.....	57
FIGURA 27	EJECUCIÓN DE COMANDO.....	58
FIGURA 28	DIRECTORIO IMAGES.....	58
FIGURA 29	ARCHIVO BLADE.....	59
FIGURA 30	EJECUCIÓN DE COMANDO.....	59
FIGURA 31	CONTROLADOR REPORTE	60
FIGURA 32	COMANDO EJECUCIÓN DEL PROYECTO.....	60

FIGURA 33 PAGINA DE GENERACIÓN DE REPORTES	61
FIGURA 34 RANDOM FOREST REGRESSOR	63
FIGURA 35 SUPPORT VECTOR MACHINE	65
FIGURA 36 GRADIENT BOOSTING REGRESSOR	66
FIGURA 37 ESTRUCTURA DEL MODELO	70
FIGURA 38	72
FIGURA 39 CONFIGURACIÓN INICIAL	73
FIGURA 40 KEYWORDS	74
FIGURA 41 TABLA RECOMENDACIONES	75
FIGURA 42 OBTENCIÓN DE DATOS	76
FIGURA 43 OBTENCIÓN DE DATOS	76
FIGURA 44 DETECCIÓN DE KEYWORDS	77
FIGURA 45 EXTRACCIÓN CARACTERÍSTICAS	78
FIGURA 46 FUNCIÓN PROCESAR DATOS	79
FIGURA 47 SEVERIDAD Y FRECUENCIA	80
FIGURA 48 SEVERIDAD Y FRECUENCIA	80
FIGURA 49 GENERACIÓN DE RECOMENDACIONES	81
FIGURA 50 ENTRENAMIENTO DEL MODELO	82
FIGURA 51 PIPELINE	83
FIGURA 52 PREDICCIÓN Y GUARDAR DATOS	84
FIGURA 53 CÓDIGO GUARDAR RECOMENDACIÓN	85
FIGURA 54 CÓDIGO ACTUALIZAR RECOMENDACIONES	85
FIGURA 55 DIAGRAMA DE FLUJO DEL PROTOTIPO	86
FIGURA 56 PÁGINA DE GENERACIÓN DE INFORMES	87
FIGURA 57 PÁGINAS DEL INFORME	88
FIGURA 58 RECOMENDACIONES CLIENTE 1	93
FIGURA 59 RECOMENDACIONES CLIENTE 4	93
FIGURA 60 RECOMENDACIONES CLIENTE 2	94
FIGURA 61 RECOMENDACIONES CLIENTE 3	95

ÍNDICE DE TABLAS

TABLA 1 CRITERIOS DE LIMITACIÓN DE PROBLEMA.....	6
TABLA 2 CRITERIOS DE EVALUACIÓN DEL PROBLEMA	6
TABLA 3 COMPARATIVA ENTRE FRONTEND Y BACKEND	29
TABLA 4: CLASIFICACIÓN DE SISTEMAS DE CONTROL DE VERSIONES	34
TABLA 5 CLASIFICACIÓN DE SISTEMAS DE CONTROL DE VERSIONES	46
TABLA 6 CUADRO COMPARATIVO	67
TABLA 7 COMPARACIÓN FRECUENCIA	95

Resumen

Este trabajo es la construcción e investigación de un prototipo de sistema web basado en Machine Learning, la automatización inteligente es una nueva forma de potenciar la automatización de procesos tradicionales con la Inteligencia artificial, PROTEMAXI una empresa que brinda servicio de seguridad a sus clientes dentro de la ciudad de Guayaquil tiene la necesidad de agilizar el proceso de generación de informes de las novedades reportadas mensualmente a la vez que la analiza para obtener información valiosa, comúnmente este proceso se hace de forma manual pero es poco preciso y puede contener errores; por este motivo el presente proyecto se basa en la construcción del prototipo del sistema que genera reportes automatizados de acuerdo al cliente y la fecha es decir el periodo integrando la tecnología del aprendizaje automático; La construcción del sistema fue hecha con los requerimientos tecnológicos que la empresa solicito como el Framework de Laravel para la interfaz web y para el modelo de Machine Learning Python. Desde las diferentes reuniones de levantamiento de información hasta el desarrollo del modelo de Machine Learning se utiliza una metodología de software en cascada que permite trabajar de manera más directa y sin petición a cambios no establecidos. Los objetivos del proyecto se basan en la implementación del empleo de técnicas de Machine Learning para generar resultados en base a los patrones encontrados, la elaboración del sistema web, la creación de una base de datos y el uso de las tecnologías requeridas por la empresa por termino de compatibilidad. Para finalizar, el proyecto concluye con la generación de informes de acuerdo con el cliente y al mes seleccionado donde se muestran datos precargados y los resultados del análisis del modelo de Machine Learning, gracias a esta implementación se logra reducir el tiempo de generación del informe y se obtienen patrones de frecuencia sobre las novedades ocurridas y que recomendaciones se pueden tomar al respecto.

Palabras clave: *Aprendizaje Automático, Automatización inteligente, inteligencia artificial, prototipo, Seguridad, Python, Laravel, Patrones.*

Abstract

This work is the construction and research of a prototype web system based on Machine Learning, intelligent automation is a new way to enhance the automation of traditional processes with artificial intelligence, PROTEMAXI a company that provides security service to its customers within the city of Guayaquil has the need to streamline the process of generating reports of developments reported monthly while analyzing it for valuable information, commonly this process is done manually but is inaccurate and may contain errors; For this reason the present project is based on the construction of the prototype of the system that generates automated reports according to the client and the date that is to say the period integrating the technology of the automatic learning; The construction of the system was made with the technological requirements that the company requested as the Laravel Framework for the web interface and for the Machine Learning Python model. From the different information gathering meetings to the development of the Machine Learning model, a cascade software methodology was used, which allows working in a more direct way and without requesting unestablished changes. The objectives of the project are based on the implementation of the use of Machine Learning techniques to generate results based on the patterns found, the development of the web system, the creation of a database and the use of the technologies required by the company in terms of compatibility. Finally, the project concludes with the generation of reports according to the client and the selected month where preloaded data and the results of the analysis of the Machine Learning model are shown, thanks to this implementation it is possible to reduce the time of generation of the report and frequency patterns are obtained about the new developments that occurred and recommendations that can be taken in this regard.

Keywords: *Machine Learning, Intelligent Automation, artificial intelligence, prototype, Security, Python, Laravel, Patterns.*

Introducción

En el entorno empresarial moderno, la intervención de la tecnología para mejorar procesos manuales o ineficientes puede marcar la diferencia entre una empresa que sabe cómo reducir riesgos y evitar limitaciones operativas por error humano que podrían tener un grave impacto a nivel competitivo; debido a esto, muchas empresas buscan soluciones tecnológicas automatizar sus procesos y también hagan un análisis e identifiquen patrones y tendencias para mejorar la toma de decisiones estratégicas empleando herramientas en auge como la inteligencia artificial.

A este tipo de soluciones se las conoce como Automatización inteligente a comúnmente llamada automatización cognitiva, donde se usa tecnologías de automatización (inteligencia artificial [IA], gestión de procesos de negocios (BPM¹) y automatización robótica de procesos [RPA²]) para mejorar la rapidez y calidad de la toma de decisiones en todas las organizaciones.(IBM³, 2021, p. 1) El Machine Learning una rama destacada de la inteligencia artificial es atractiva debido a su capacidad de mejorar sus resultados sin la necesidad de que un humano le diga cómo, produciendo resultados con mayor precisión y rapidez que el humano.(Serrahima, 2022) Esto permite a las organizaciones implementar mejoras tecnológicas mediante el uso de Machine Learning en la automatización de procesos repetitivos y manuales en una gran variedad de sectores debido a la capacidad de aplicación que puede ser perceptiva y cognitiva.

El presente trabajo, aborda la implementación de una mejora tecnológica mediante el desarrollo de un prototipo de sistema web basado en Machine Learning para la identificación de tendencias y análisis de incidentes de la empresa de PROTEMAXI; empresa privada con más de 22 años de trayectoria que brinda diferentes servicios de seguridad sus clientes, la empresa mantiene un registro de incidencias reportadas en el sistema por su personal en las instalaciones de cada cliente, esta información permite a la empresa de seguridad optimizar la toma de

¹ Business Process Management

² Robotic Process Automation

³ International Business Machines

decisiones, tomar medidas preventivas y mejorar los protocolos de seguridad. Sin embargo, mensualmente se realiza un informe de manera manual por cliente que contiene todas las incidencias registradas de acuerdo con cada cliente, este proceso representa una limitación ya que no solo demanda tiempo y recursos, además de aumentar el riesgo de errores y la complejidad para identificar patrones o tendencias con eficacia.

Con los años los avances tecnológicos han proporcionado mejoras innovadoras que producen la automatización de varios procesos que dan ventajas competitivas como lo son la reducción de costos, ahorro de tiempo e incremento de productividad dentro de las empresas (J. R. A. Barrera et al., 2024, p. 4). Así, se decide presentar el siguiente trabajo como una oportunidad de mejorar un proceso importante de PROTEMAXI mediante el uso de tecnologías que más allá de efectivizar tiempo, muestren un análisis propicio y más preciso en base a datos que pueda proporcionar recomendaciones para el manejo de las incidencias; lo que permitiría a la empresa optimizar sus operaciones, mejorar la seguridad proactiva y brindar un servicio más eficiente y efectivo a sus clientes.

La estructura del siguiente trabajo consta de 4 capítulos, el primer capítulo abordar la problemática de nuestro trabajo partiendo desde su análisis hasta la definición de las variables de investigación. El segundo capítulo se desarrolla el marco lógico que sirve como un medio de planificación y evaluación. El tercer capítulo se encuentra la metodología que se utilizara para llevar a cabo el trabajo. Y en el cuarto capítulo, se detallada y explica que se implementar para resolver la problemática.

Capítulo I: El Problema

En este capítulo se describe la problemática que tiene la empresa en su gestión de datos y generación de informes de incidencias debido al manejo actual con el que se lleva a cabo estos procesos,

ANÁLISIS DEL PROBLEMA.

Para mejorar la gestión de datos de los incidentes de seguridad registrados por cada uno de los clientes de PROTEMAXI se plantea un desafío en la gestión eficiente de sus datos; los cuales tratan sobre incidentes que la empresa recopila en los que se incluyen robos, vandalismo, intrusiones a propiedad privada y otros eventos similares, que generan un repositorio de datos tanto estructurados como no estructurados, informes escritos e imágenes.

Ubicación del Problema en un Contexto

Debido a la gestión manual de los datos por parte de PROTEMAXI y a la ausencia de herramientas especializadas para el procesamiento y análisis de datos, este proceso de gestión resulta ser una carga de trabajo que puede llegar a generar algunas limitaciones operativas que se describen a continuación:

En el Proceso Manual

- Riesgo de error humano en el procesado de datos y generación de informes.
- Tiempo prolongado en la digitación y transcripción de informes de incidentes para los informes.
- Identificación de patrones inexistente por la falta de sistematización.

Escasa Capacidad Analítica

- Dificultad para identificar tendencias y patrones de comportamiento en sus datos.
- Imposibilidad de generar análisis predictivos eficaces a partir de la data de las incidencias reportadas.

Generación de Informes

- Carencia de análisis profundo de los datos en los informes.
- Inexistencia de recomendaciones basadas en evidencia histórica.
- Demoras en la generación de informes.

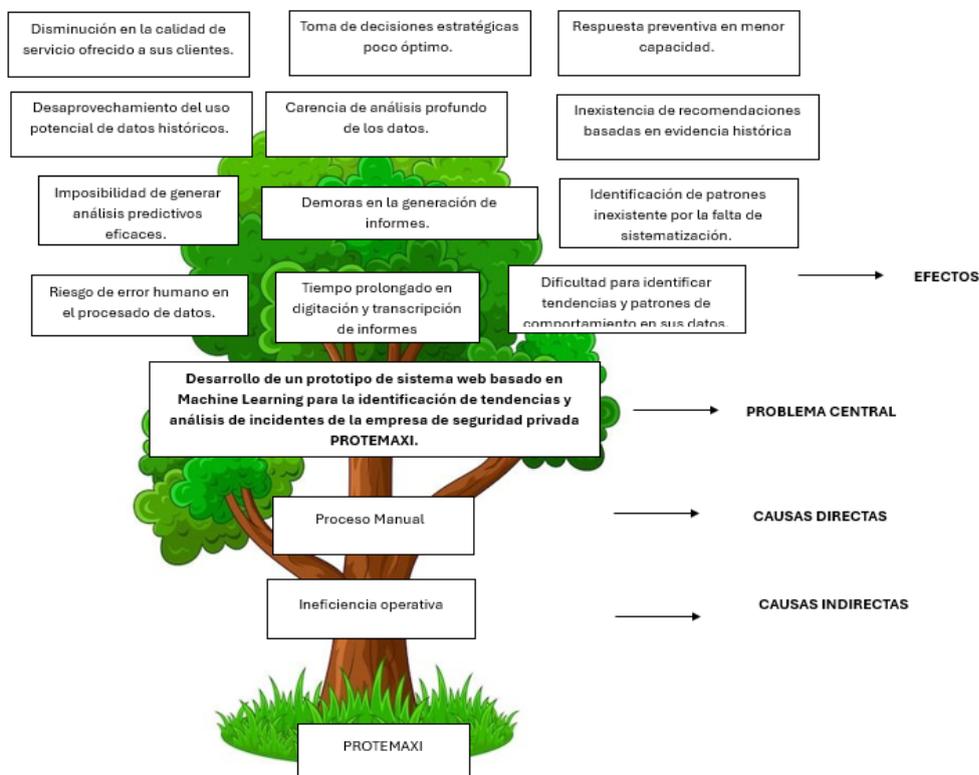
Gestión Empresarial

- Respuesta preventiva en menor capacidad.
- Toma de decisiones estratégicas basado en un análisis de datos manual.
- Desaprovechamiento del uso potencial de datos históricos.
- Disminución en la calidad de servicio ofrecido a sus clientes.

Causas y Consecuencias del Problema

A continuación, se especifica gráficamente la problemática identificada en la empresa respecto a la operación de los datos y generación de informes:

Figura 1
Árbol del Problema



Nota: Análisis de la problemática de la empresa privada mediante un árbol del problema, identificando las causas principales y los efectos derivados del problema central.

Delimitación del Problema

Para la delimitación del problema se escogieron 4 criterios, los que se indican en la tabla siguiente:

Tabla 1
Criterios de Limitación de Problema

#	Criterios	Detalles
1	Campo	Tecnología
2	Área	Análisis de datos y Desarrollo web.
3	Aspecto	Machine Learning, Identificación de tendencias, Toma de decisiones estratégicas, Reportería
4	Propuesta	Desarrollo de un prototipo de sistema web basado en Machine Learning para la identificación de tendencias, análisis de incidentes y generación automatizada de informes para la empresa de seguridad privada PROTEMAXI.

Nota: Criterios que delimitan el alcance y enfoque de la investigación.

Formulación del Problema

Ante la problemática surge como propuesta el desarrollo de un prototipo de sistema web basado en Machine Learning para la identificación de tendencias y análisis de incidentes de la empresa de seguridad privada PROTEMAXI; este optimizará la eficiencia en la gestión de los incidentes registrados, la calidad de los servicios brindados a sus clientes y la capacidad de la empresa a la hora de toma decisiones estratégicas informadas; de este modo, se automatizará la gestión de los datos y la generación de informes.

Evaluación del Problema

De los diez criterios definidos para la evaluación del problema, se identificaron siete que aplican a nuestro proyecto, mismos que se explican en la tabla siguiente:

Tabla 2
Criterios de Evaluación del Problema

#	Criterios	Contexto
1	Delimitado	El problema se reduce a la falta de automatización en la gestión, análisis y procesamiento de datos que se ha estado realizando de forma manual sobre las incidencias reportadas por los clientes de la empresa.
2	Claro	La definición del problema permite comprender el impacto del proceso manual en gestión, análisis y procesado de datos de la empresa.

#	Criterios	Contexto
3	Relevante	Para resolver el problema se requiere del desarrollo de un prototipo de sistema web basado en Machine Learning para la identificación de tendencias y análisis de incidentes.
4	Evidente	Como se indica en la ilustración # 1 las causas que se indican son claras, no requieren mayor detalle para poder definir el problema.
5	Original	La empresa no cuenta con otra herramienta que utilice Machine Learning implementada dentro de sus soluciones.
6	Factible	La empresa dispone de la tecnología para implementar el prototipo de sistema que solventa el problema, además se estableció un tiempo aceptable para su desarrollo.
7	Identifica productos esperados	El desarrollo de un prototipo de sistema web que automatice la gestión de datos y la generación de informes de los reportes de incidencias con recomendaciones para una mejor toma de decisiones estratégicas para la empresa.

Nota: Criterios establecidos para la evaluación del problema, proporcionando un marco para analizar y medir la problemática.

OBJETIVOS

Objetivo General

Desarrollar un prototipo de sistema web basado en Machine Learning para la identificación de tendencias y análisis de incidentes de una empresa de seguridad privada localizada en la ciudad de Guayaquil (PROTEMAXI).

Objetivos Específicos

- Emplear técnicas de aprendizaje automático que identifiquen patrones importantes y que puedan extraer información esencial para optimizar el manejo de novedades
- Elaborar un sistema web que maneje la generación de informes estandarizados y que consolide incidencias de acuerdo con el cliente que tenga la empresa de seguridad con la inclusión del análisis de Machine Learning
- Diseñar y construir la base de datos que gestione la carga de la data enviada por PROTEMAXI
- Usar herramientas de desarrollo que hayan sido requeridas por la empresa, así como Laravel para el desarrollo de la interfaz web y sus funcionalidad y Python para la construcción del modelo de Machine Learning

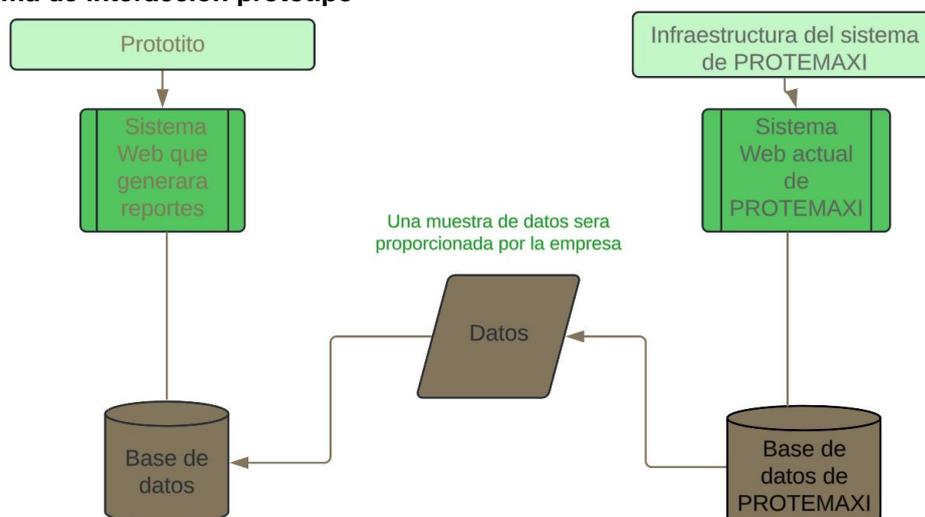
ALCANCES DEL PROBLEMA

El enfoque de este proyecto es crear un prototipo de plataforma web, usando técnicas de Aprendizaje Automático. El prototipo funcionará de forma autónoma, esto quiere decir que será un sistema totalmente independiente del que la empresa utiliza para el registro de incidencias; de esta manera se mantiene la confidencialidad, confiabilidad, la seguridad y la privacidad de la información crítica de PROTEMAXI, cumpliendo con los requerimientos de protección de datos exigidos por la empresa.

La infraestructura se ubicará conjuntamente en la nube, facilitando un desarrollo y acceso del prototipo seguro, veloz y escalable, manteniendo una separación entre el sistema de PROTEMAXI y el prototipo a construir.

Para el procesamiento y análisis de datos que requiere el prototipo se utilizará información proporcionada por la empresa de seguridad; armando un conjunto de datos preseleccionados, para garantizar el cumplimiento de las normas de privacidad y protección de datos vigentes, con lo que aseguramos que haya una línea marcada entre ambas infraestructuras y se mantienen seguros los servidores, conexiones, se cumplen los protocolos de seguridad y la información que existe se mantiene de manera confidencial sin haber sido tratada lo que mantienen toda la seguridad de la infraestructura tecnológica de PROTEMAXI. A continuación, se ilustra la interacción entre el prototipo a construir y el sistema web de la empresa:

Figura 2
Esquema de interacción prototipo



Nota: Interacción de datos entre el sistema de PROTEMAXI y el prototipo

Las características que incluye el sistema son la generación informes estandarizados que consolidan las novedades registradas de manera mensual por cliente de formato automatizada, el proceso de análisis de patrones de las novedades o incidencias de seguridad que son reportadas y la interfaz web donde se visualiza los clientes con su periodo para generar y visualizar los informes. El uso de Machine Learning tendrá importancia en el uso de algoritmos de procesamiento de lenguaje natural que puedan analizar detalles tipo texto o imágenes para un análisis gráfico de novedades.

Los datos que se utilizaran para trabajo y análisis serán solamente con el grupo o data set que haya proporciona PROTEMAXI, cumpliendo estrictamente con la Ley de Protección de Datos de los Clientes (Barrezueta, 2021) con los protocolos de confidencialidad y privacidad establecidos por la normativa vigente: esto asegura que los datos sensibles sean tratados de manera segura y conforme a las regulaciones aplicables.

El análisis se limitará a campos previamente acordados, como fecha, descripción y material gráfico, a partir de este análisis, el sistema generará insights valiosos, produciendo recomendaciones personalizadas y optimizando la toma de decisiones en materia de seguridad para los clientes de PROTEMAXI.

Se debe mencionar que existen limitaciones y punto claves a tener en cuenta dentro del proyecto, el no tener acceso a la base de datos o tener un recurso limitado de datos así como se restringe el análisis a ciertos tipos de datos sensibles para la empresa, estas restricciones sirven para mantener un protocolo que priorice la protección y confidencialidad de la data que no ha sido proporcionada; el prototipo del sistema tiene su función en mejorar de manera eficiente la gestión de las novedades o incidentes reportados a PROTEMAXI que se traduce en la generación de un informe que consolide todo esa información de manera automatizada y con la inclusión de los resultados del Modelo de Machine Learning.

JUSTIFICACION E IMPORTANCIA

El presente trabajo de integración curricular se ha llevado a cabo por la relevancia que tiene sobre PROTEMAXI al incorporar una mejora tecnológica que sirve para automatizar y mejorar un proceso manual utilizando algoritmos de Machine

Learning; esto debido a las limitaciones o capacidades humanas que existen al momento de realizar un análisis más rápido y preciso de las incidencias de los diferentes clientes que provocan limitaciones operativas o riesgo de errores al analizar los datos dentro de la empresa.

La competitividad y toma de decisiones a la que la empresa puede favorecerse, teniendo en cuenta que una organización es bien valorada por la calidad de servicios o bienes que es capaz de llegar a ofrecer; esto mejoraría la calidad del servicio de la empresa tanto en la entrega de informes como en la información que contiene el mismo y por ende puede sobrepasar o estar a la par de su competencia.

Acorde a lo indicado se detallan los siguientes justificativos:

1. La toma de decisiones estratégicas radica en la capacidad de identificar patrones o tendencias que junto con las recomendaciones se pueden extraer del informe para que la empresa pueda tomar medidas preventivas, cambiar sus protocolos de seguridad con cada cliente y tomar mejores decisiones en cómo opera a sus empleados.
2. La mejorar el proceso actual con la implementación del prototipo beneficiaria a la empresa, al personal administrativo o privado encargado de la generación de incidencias y a los analistas que generan los informes; sus clientes también contribuirían indirectamente a la sociedad fortaleciendo las instalaciones o establecimientos donde sus empleados y compradores asisten, esto promueve la seguridad en la ciudad teniendo en cuenta las problemáticas que enfrenta la ciudad de Guayaquil con respecto a la inseguridad y delincuencia.
3. Automatizar el proceso optimiza el uso de recursos de la empresa, el personal invertiría menos tiempo en labores repetitivas que haya dentro del proceso de análisis de las incidencias registradas y en la creación del informe; concentrándose en brindar propuestas de mejora en la seguridad los clientes de PROTEMAXI con la información generada por el informe.

Acorde con las justificaciones descritas el desarrollo de este prototipo es útil al optimizar y automatizar procesos internos de la empresa con relación al registro de incidencias, mejorar la toma de decisiones estratégicas de la empresa y brindar un mejor servicio siendo más eficientes que su competencia.

En cuanto a sus clientes, reciben un análisis en menos tiempo y más preciso lo cual puede elevar la satisfacción y el valor que le tienen a la empresa; lo que representa un ejemplo de innovación y una base referencial para todas aquellas empresas o personas que deseen aplicar la tecnología en sus procesos manuales.

El conjunto de estos justificativos beneficia a la empresa de seguridad en su competitividad, servicios, decisiones y sirve de ejemplo para la automatización de procesos utilizando Machine Learning.

HIPÓTESIS O PREGUNTA DE INVESTIGACIÓN

¿El prototipo de sistema web basado en Machine Learning podrá generar eficientemente informes de calidad para la empresa de seguridad física PROTEMAXI?

VARIABLES DE LA INVESTIGACIÓN

Precisión de análisis: El prototipo debe presentar un análisis donde se haya identificado e interpretado con éxito los patrones en base a los datos cargados mostrando un análisis preciso. Esto se medirá obteniendo el porcentaje, utilizando el número patrones que hayan sido correctos sobre los patrones reportados por el modelo de Machine Learning en cada informe, el número de patrones correctos se determinará con relación a la data proporcionada. Estas mediciones se harán cada vez que se realice una prueba sobre el prototipo para evaluarlo con el paso del tiempo.

Generación de Informes: Los informes se deben crearse de manera automatizada y eficientemente; La forma en la que se generan los informes debe tener la capacidad de poder diferenciarse ante el proceso manual de generación de informes; Esto se medirá mediante el tiempo que se emplea en la generación de un informe tomando en consideración el volumen de datos que se vaya a analizar, comparándolo con el tiempo que toma el proceso manual bajo las mismas

condiciones; Esta medición constará de varias pruebas para ver bajo qué condiciones el tiempo puede verse afecto y obtener una diferencia para determinar si es mayor o menor.

Relevancia de recomendaciones: Las recomendaciones generadas en cada informe deben ser útiles para la toma de decisiones de la empresa ya que podrían mejorar sus protocolos y reforzar su seguridad. Por esta razón se medirá el puntaje de relevancia realizando una encuesta donde personas designadas por la empresa en base a su conocimiento previo pueda dar una calificación sobre la relevancia en las recomendaciones. Esta se manejará de una escala del 1 al 5; luego se calculará el promedio que servirá como una referencia sobre el puntaje general de la relevancia que tienen las recomendaciones.

Capítulo II: Marco Teórico

Dentro de este capítulo se definen términos asociados a los sistemas web y a la automatización inteligente, mencionando sus características, importancia, aplicaciones, ventajas y desventajas, dicha información se pone bajo referencias bibliográficas para garantizar la validez de estos; estableciendo bases conceptuales que sirvan para el desarrollo del prototipo de sistema web basado en Machine Learning y para el entendimiento del lector.

2.1 Sistema Web

Se considera un sistema web a la agrupación de varios elementos de software y hardware que se integran y trabajan de forma simultánea dándole una mejor experiencia al usuario conforme a sus requerimientos, componentes como la interfaz gráfica, bases de datos y servicios o protocolos web son elementos principales que conforman a esta clase de sistemas (Maheswari et al., 2023).

Los usuarios usan de manera indiferente las palabras “sistema web” y “aplicación web” ya que ellos solo interactúan con la parte visual del sistema, a nivel técnico existen diferencias puntuales que denominan al sistema web como el conjunto de varios componentes de software y hardware que trabajan entre si con el fin de que el usuario pueda ver el resultado de aquello, en otras palabras está conformado por todo lo que el usuario puede y no puede ver a simple vista mientras que una aplicación web es una parte de todo ese conjunto cuyo objetivo es brindar interacción y funcionalidades que sean intuitivas para los usuarios, por ejemplo una plataforma de e-commerce creada para que los usuarios puedan comercializar a través de internet es todo un sistema web mientras que el apartado del carrito de compras del e-commerce sería la aplicación web ya que proporciona al usuario la posibilidad de llevar una cesta de los productos para estar al tanto de todo lo que vaya a comprar en línea. En el lenguaje cotidiano las personas usan ambas palabras indistintamente como sinónimos sin embargo a nivel técnico una es un componente que va dentro de la otra.

Un sistema web puede tener un efecto positivo en un negocio, entidad o empresa cuando se identifica la necesidad u oportunidad de mejora que se desee

cubrir. El estudio académico de Alvarez-Intriago & Torres-Samaniego (2019), analiza el impacto que tendría desarrollar un sistema web para optimizar insumos en un local de comida, para obtener los resultados se definió el producto, diseñó la solución y desarrolló el sistema web. El artículo concluye que la implementación de dicho sistema permite un control de stock y rangos de precios en cada sucursal del restaurante, obteniendo incremento de utilidad en ventas, minimización de pérdidas de insumos y logrando una mejor toma de decisiones sobre la gestión del inventario.

Berilu Castañeda (2023) realiza una automatización de la gestión de procesos presupuestarios de una compañía de arquitectura e ingeniería instaurando un sistema web, así automatiza las tareas manuales o de validación, esto disminuye el tiempo administrativo al agilizar el registro y verificación de datos presupuestarios, evita la duplicidad de montos y la desaparición de información; La empresa reflejó estos beneficios teniendo un aumento positivo en su eficacia presupuestaria del 23% y un 54% en la productividad del nivel de servicio. El autor describe que este sistema permitió a la organización tener un mejor dominio y entendimiento sobre sus finanzas y al mismo tiempo mejoró el nivel de servicio.

La importancia y el desarrollo de los sistemas web ha ido evolucionando, volviéndolos más dinámicos y atractivos para los usuarios debido a las nuevas formas de desarrollarlos dejando a un lado a aplicativos webs que solo mostraban información estática o básica, las personas valoran lo visual e intuitivo que puede llegar a ser porque así ellos se sienten más cómodos usándolos y esto incrementa la aceptación y reputación del negocio; así más organizaciones se ven envueltas en un constante cambio debido a que cada vez surgen más tecnologías y tienen que decidir si adaptarse o quedarse atrás.

Dentro del proyecto, los usuarios accederán a un sistema web que representa el sitio donde ellos generan los informes de acuerdo a la información que ellos deseen generar; la aplicación web funcionará como un medio por el cual ellos podrán interactuar y consultar la información que necesiten para generar el informe; esto la convierte en un componente imprescindible en el proyecto ya que sin ella no se puede realizar el proceso de la automatización de informes, es decir el aplicativo web se tiene la función de agilizar la generación de informes y servir como un sistema de generación de informes.

2.2 Automatización Inteligente

Según Amazon Web Services, Inc. (2023) la automatización inteligente (AI) es el uso de la inteligencia artificial aplicada a las automatizaciones tradicionales con el fin de convertirlas en automatizaciones avanzadas que puedan ejecutar procesos simples o complejos teniendo un aprendizaje para adaptarse a cualquier cambio.

Estas se caracterizan por utilizar tecnologías y técnicas relacionadas a la inteligencia artificial como lo es el aprendizaje automático para ejecutar tareas complejas, la detección de patrones que beneficia al aprendizaje y entendimiento del lenguaje natural para analizar texto; dentro de las empresas que desarrollan o implementan esto en sus automatizaciones tienen el beneficio que ahorrarse tiempo y dinero debido a que no están reclutando personal o invirtiendo recursos para llevar a cabo una tarea compleja y repetitiva la cual puede ser reemplazada por el nuevo sistema.

La automatización inteligente puede llegar a cambiar los procesos tradicionales mediante su optimización, donde las tareas o actividades a realizar sean completadas a tiempo maximizando los recursos requeridos. De acuerdo con IBM (2021) un proceso se define como el conjunto de actividades que se realizan de manera secuencial o también paralela, estas suelen estar asociadas una de otras de manera dependiente o independiente; las actividades o tareas son necesarias para alcanzar un objetivo o cumplir con el proceso.

IBM (2021) menciona las ventajas que puede tener la automatización inteligente en las diversas industrias debido a la capacidad de esta para utilizar considerables volúmenes de datos, cálculos certeros y ejecutar operaciones comerciales (gestión de procesos, atención al cliente y análisis de datos). A continuación, se describen los aspectos positivos:

Reducción de costos: la automatización de tareas repetitivas con el análisis de datos influye en la precisión que obtiene la empresa para realizar tareas sin necesidad de reclutar nuevo personal logrando así optimizar su tiempo y recursos; gracias a esto la empresa obtiene como resultado una reducción de costos operativos lo que provoca un aumento en la producción permitiendo escalar a la organización de manera inmediata sin poner en riesgo su calidad de procesos y tampoco fatigando a

su fuerza laboral.

Mejora de precisión en calidad de procesos y decisiones: como se mencionó anteriormente la precisión es una variable que se potencia cuando se realizan automatizaciones, sin embargo, esto alcanza un nivel más elevado cuando se implementa la inteligencia artificial siendo capaz de adaptarse a cambios y aprender de un proceso cuando se vuelva más complejo con el paso del tiempo, teniendo un enfoque más coherente donde pueda generar resultados más confiables; esto impulsa a una mejor toma de decisiones por parte de la organización al tener resultados más precisos y coherentes.

Mejora competitiva: la aplicación de esta automatización puede utilizarse para mejorar el bien o servicio que la empresa ofrece, esto provoca que su producto sea de alta calidad y más confiable al dar soluciones rápidas a peticiones o procesos que se tornaban complejos para los usuarios, esto incrementa el nivel de competitividad de la organización.

Adaptación al cumplimiento de normas y regulaciones: las empresas suelen regirse bajo estrictas políticas o normativas que deben seguirse de manera continua y obligatoria ya sea por entidades que las rigen con el fin de regularizarlas. Al aplicar la automatización inteligente se asegura que se cumplan las reglas impuestas con el fin de presentar un cumplimiento como puede ser la generación de informes o seguimiento regulatorios.

La automatización inteligente es un componente de alto valor porque al igual que el sistema web también interviene en el proceso de generación de informes; cuando se requiere generar un informe, la información proveniente de la base datos que se mostrara en el informe necesita ser analizada aplicado tecnologías como la inteligencia artificial y algoritmos de Machine Learning que puedan interpretar la información.

2.2.1 Automatización Inteligente de Procesos

De acuerdo con Zhang (2019) la automatización inteligente es un software en específico que facilita la realización de actividades o procedimientos con la más mínimo interacción con el individuo, esto se logra gracias a las tecnologías que se

integran en este tipo de automatización como la robótica, inteligencia artificial y el análisis de datos; las 3 pueden combinarse o interactuar por separado dependiendo del proceso a realizar. El autor resalta las 3 subcategorías más sobresalientes dentro de la automatización inteligente de procesos:

1. La automatización de procesos robótica es usada en procedimientos mecánicos que se caracterizan por tener un flujo de trabajo extenso donde se tienen que seguir reglas o condiciones que lleven a dar un resultado único, al tener una sola respuesta se necesita conocer y seguir de manera estricta todo el proceso; este tipo de automatización también se aplica en transacciones o actividades donde se manejen datos organizados y formateados porque facilita el reconocimiento y entendimiento de la máquina que esté utilizando la información.
2. La automatización de procesos con inteligencia artificial utiliza dicho campo de la tecnología donde las máquinas adquieren las habilidades de aprender y analizar a través de automatizar procesos que requieran habilidades cognitivas, aprendizaje automático y de razonamiento; este tipo de automatización se encarga de funcionar como un soporte de decisión cuando se requiere obtener información de cualquier tipo y examinar riesgos, por ejemplo al aplicar esta automatización puede proporcionarnos datos que sean relevantes cuando necesitemos analizar un riesgo o generar conclusiones, también es capaz de generar predicciones y realizar análisis; es común que dentro de procesos como la auditoria de contratos se utilice esta subcategoría.
3. La automatización de procesos con análisis de datos se comprende del campo que se encarga de identificar patrones o información que sea relevante basándose de un grupo de datos; dando un ejemplo transaccional en una auditoria, el análisis puede identificar irregularidades y valores atípicos que después serán examinados para comprobar que las operaciones que se realicen en la compañía se hagan de manera correcta, por lo general esto se realiza para evitar fraudes, prevenir pérdida de activos y validar la información financiera y operativa de la empresa; Al aplicar esta automatización los auditores dejan de realizar el proceso

manual donde ellos tienen que realizar pruebas sobre un conjunto aleatorio de datos para después corroborarlo con información de respaldo con el fin de validar que todo se lleve de manera correcta y en vez de eso utilizan este tipo de implementación que para realizar pruebas sobre una muestra de datos que puedan representar un riesgo para la entidad es decir en análisis proporcionara datos relevantes.

La automatización inteligente de procesos con inteligencia artificial se aplica en este proyecto mediante la generación de informes al momento de utilizar Machine Learning o aprendizaje autónomo para realizar un análisis sobre los datos de los incidentes con el fin de generar recomendaciones en el informe que actúen como soporte en la toma de decisiones por parte del personal de la empresa; la utilización de esta combinación de tecnologías logra facilitar el trabajo manual que tomaría realizar un análisis de todos los incidentes para emitir recomendaciones cuando se realice un informe, disminuye posibles errores y genera información de valor al analizar datos y aprender mientras más se vaya entrenando al modelo, esta clase de automatización diferencia al sistema de cualquier automatización común o mecánica donde solo se replican pasos como puede ser la robótica, esta se compone un análisis, recomendaciones y aprendizaje.

2.3 Software

2.3.1 Machine Learning

De acuerdo con Oracle (2024), el aprendizaje automático (ML) es un campo que se deriva de la inteligencia artificial cuyo propósito es mejorar el aprendizaje, rendimiento de los sistemas tradicionales en relación a los datos que procesa. Los términos inteligencia artificial y aprendizaje automático se usan comúnmente creyendo que tienen el mismo significado, sin embargo, es totalmente lo contrario debido a que si bien todo aprendizaje automático es inteligencia artificial no toda inteligencia artificial está hecha bajo dicha ramificación.

La inteligencia artificial es todo aquello que tiene una intelectualidad o razonamiento autónomo y propio, diferenciándose al no es natural es decir es fabricado o creado por alguien, también es una parte de la informática que es capaz

de integrarse en diferentes sistemas para recrear tareas o procesos que requieren de inteligencia humana, aquí se incluye tareas que pueden ir desde lo mecánico hasta lo complejo como resolución de problemas, generación de recomendaciones, predicciones y toma de decisiones.

Todos los campos de la inteligencia artificial son importantes ya que cada uno se enfoca en imitar una parte del sistema cognitivo del ser humano y al mismo tiempo de sus comportamientos; el Machine Learning está orientado al reconocimiento de patrones y del aprendizaje a partir de datos que pueden ser de muchos tipos, esto beneficia a los sistemas tradicionales ya que les permite mejorar de manera autónoma sin necesidad de requerir una programación manual para cada tarea.(Cuadrado & Angulo, s/f)

IBM (2021), el Machine Learning puede descomponerse en 3 partes esenciales que demuestran el funcionamiento de su sistema de aprendizaje, la primera establece que todos los algoritmos de Machine Learning tiene el uso de clasificar o predecir algo en relación a lo que se proporcione. Los datos que se provean son esenciales porque permiten al algoritmo encontrar tendencias o producir estimaciones en relación con los patrones encontrados. La segunda parte evalúa los posibles errores de predicción que puedan existir en los resultados que proporciona el modelo a través de una función de error, esto puede ser escalable cuando conocemos ejemplos confiables sobre lo que se quiere obtener. Por último, el proceso de optimización permite al modelo generar mejores respuestas que puedan tener una mayor precisión; al momento en que se evalúa y compara con respecto a un ejemplo.

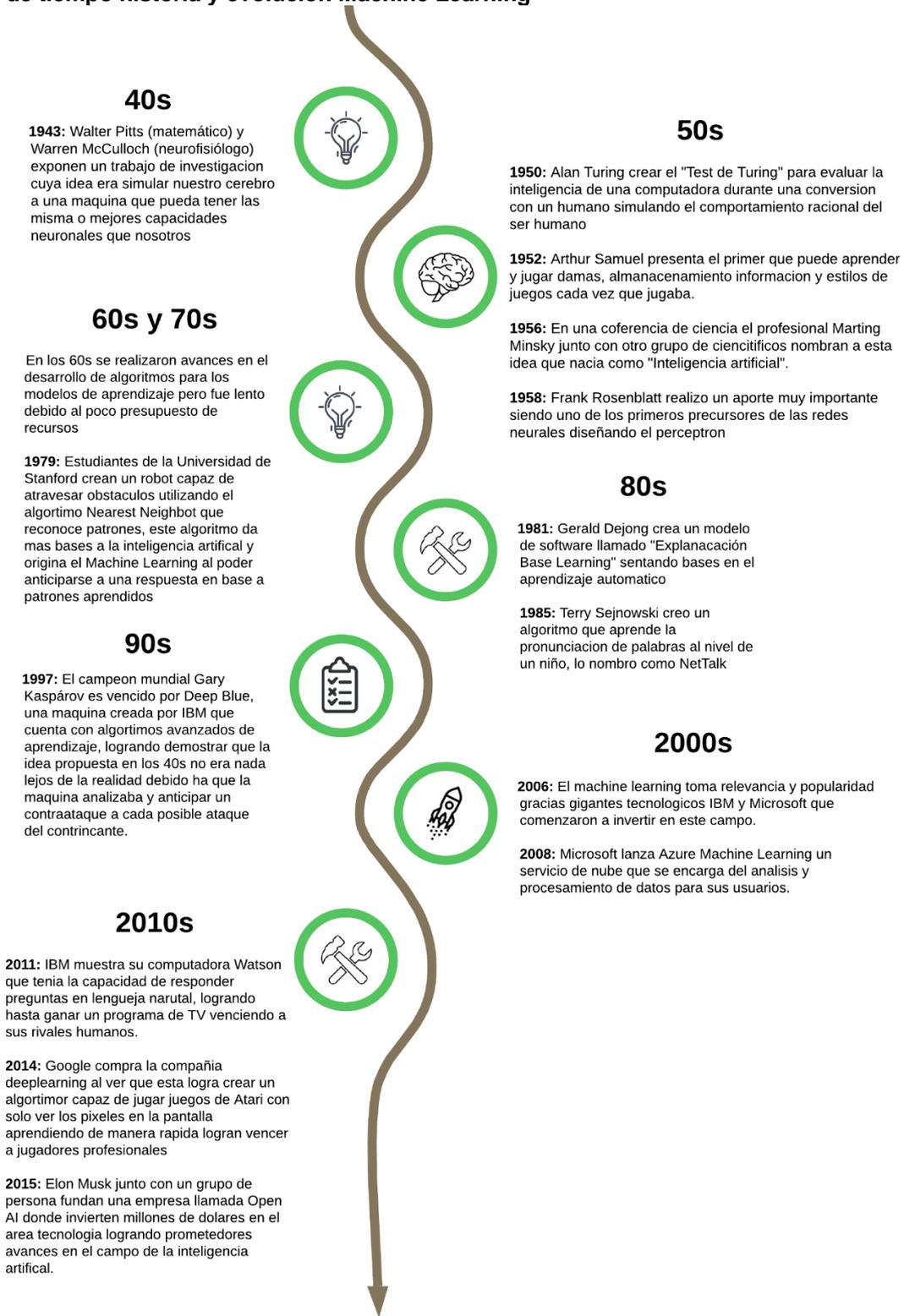
Lo llamativo de esta tecnología es que simula el comportamiento o aprendizaje humano mientras combina la rapidez y precisión de una maquina al momento de realizar tareas complejas o procesos que requieran un análisis masivo de datos, adaptabilidad en tareas cambiantes y elaboración de predicciones de manera constate; debido a esto el machine Learning es considerado una tecnología poderosa y valiosa por los beneficios que ofrece a diferentes tipos de organizaciones o empresas al ser estos aplicados a cualquier ámbito o para fines específicos. El conocer su origen y características nos ayuda a entender cuáles fueron las razones por las cuales se sigue desarrollando y solicitando personas calificadas para desarrollar este tipo de herramientas.

2.3.1.1 Historia y evolución

El desarrollo del Machine Learning comienza como una idea muy lejana de la realidad; no obstante, gracias al aporte realizado por: científicos, matemáticos, docentes, psicólogos y empresas tecnológicas ha logrado un desarrollo acelerado a partir del 2022, siendo unos de los campos más valorados dentro del contexto de la inteligencia artificial; ya que el Machine Learning mediante el uso de algoritmos provee a los sistema IA la capacidad para encontrar patrones en altos volúmenes de datos mediante la aplicación de análisis predictivo.

A continuación, se presenta una ilustración de la evolución del Machine Learning. Ramírez & Cárdenas (s/f):

Figura 3
Línea de tiempo historia y evolución Machine Learning



Nota: Línea del tiempo de la evolución del Machine Learning con el pasar de las décadas; adaptación de la historia de Machine Learning de Ramírez & Cárdenas (s/f)

2.3.1.2 El aprendizaje

De acuerdo con Amazon Web Services, Inc.(2024) en Machine Learning existen diferentes formas de aprendizaje, esto se debe a que siempre suele a ver un enfoque o tipo de respuesta diferente esperado por parte del desarrollador; los estilos de aprendizaje son 4 donde se incluye el supervisado, no supervisado, semi supervisado y por refuerzo, ninguno de los 4 son iguales cada uno puede aplicarse en diferentes situaciones debido a que cada uno puede llegar a ser clave cuando se requiera resolver un problema en específico. A continuación, se presentan los 4 estilos de aprendizaje:

1. El aprendizaje supervisado se utiliza cuando se desea obtener una respuesta a partir de datos que han sido estructurados, la respuesta esperada por parte del modelo ya se la conoce sin embargo este se implementa para que el modelo aprenda a partir de un conjunto de datos y pueda realizar predicciones hasta dar con la respuesta correcta la cual nosotros ya la conocemos; cuando se realiza el entrenamiento de un modelo bajo la técnica de aprendizaje supervisado se debe tener la data de entrenamiento etiquetada y saber al resultado que se desea llegar ya que así confirmamos que está realizando adecuadamente el desarrollo, cuando se le proporcione la data al modelo debemos hacerlo sin darle la respuesta ya que así podremos validar después si la misma es coherente.
2. El aprendizaje no supervisado no requiere que los datos estén estructurados o etiquetados, esta técnica de machine Learning se utiliza cuando desconocemos la respuesta que podemos obtener y tenemos datos que están desorganizados es decir no tiene etiquetas, sin embargo, el algoritmo puede terminar una respuesta encontraron patrones, realizando agrupaciones y con la detección de rarezas en los datos.
3. El aprendizaje semi supervisado integran los dos aprendizajes mencionados anteriormente, en este aprendizaje se utiliza datos, pero se dividen en dos grupos, el primero grupo contiene un conjunto pequeño de datos etiquetados mientras en el segundo grupo tiene una gran cantidad de datos no etiquetados; lo que se pretende aquí es minimizar el esfuerzo debido a que se comienza entrenando al modelo con el primero grupo

pequeño de datos y una vez entrenado se utiliza el segundo grupo el cual se llena por el mismo modelo etiquetando a los datos.

4. El aprendizaje por esfuerzo se caracteriza por su sistema de recompensas que se habilita cada vez que el modelo alcance o logre un objetivo este ganara puntos que le servirán para recibir pistas sobre que pasos debe seguir, este no es utilizado en ambientes empresariales debido a si nivel de sesgo aun así es óptimo cuando se necesita desarrollar video juegos.

2.3.1.3 Proceso de implementación de un proyecto ML

Cuadrado & Angulo (s/f) explican que para llevar a cabo un proyecto de Machine Learning se requiere seguir 8 fases con el fin de llevar un control sobre lo que se hará, esto puede significar un éxito en nuestra implementación ya que sabemos lo que se hará de manera secuencial, esto es considerado una buena práctica. A continuación, se presenta el proceso:

Figura 4
Fases de implementación ML



Nota: Representación de las etapas que se siguen en un proyecto de Machine Learning.

2.3.1.4 Datos Sintéticos

A acuerdo con Gamo (s/f) los datos sintéticos es un concepto que nace a partir de la IA generativa que se basa en producir datos a partir de datos reales logrando crear datos con las mismas distribuciones estadísticas que el conjunto original, para

realizar esta lógica se aplican algoritmos y construyen modelos de Machine Learning.

Los datos sintéticos pueden garantizar legitimidad y credibilidad a partir de un conjunto de datos verdaderos es decir que son generados de diversos procesos que pueden darse en una organización o empresa. En cuanto a la integridad y privacidad de la información los datos sintéticos son viables ya que cumplen las normativas vigentes sobre la ley de protección de datos debido a que se derivan de datos originales y verdaderos.

La importancia de estos radica en su aplicación dentro del Machine Learning y Deep Learning, al funcionar como un recurso para mejorar o implementar distintos tipos de modelos cuando no exista la data suficiente o simplemente se tenga un conjunto pequeño que no sea suficiente para entrenar el modelo; este concepto no cobra significado hasta decir que los datos por su naturaleza es lo que más se trata dentro del campo de la inteligencia artificial, los científicos de datos ocupan gran parte de su tiempo dedicándose a la validación y tabulación de datos como parte del proceso de preparación para después realizar el entrenamiento, el conjunto de datos de entrenamiento que pueden ser reales o sintéticos se considera la base para todo científico de ciencia de datos ya de esto depende en gran parte el rendimiento y los resultados del modelo.(Báez et al., s/f)

De acuerdo con cómo se define el proyecto, es posible incluir datos sintéticos en caso de ser necesario ya que sirve para probar el funcionamiento del modelo en caso de haber ausencia de datos o no tener la data suficiente como para entrenar y probar el modelo de Machine Learning.

2.3.2 Lenguajes de Programación

Los lenguajes de programación se consideran una herramienta básica para la producción de software (Challenger-Pérez et al., 2014), en contexto general se dividen en dos tipos :

- **Bajo nivel:** Los lenguajes de programación de bajo nivel no se pueden transferir de un computador a otro, están estrechamente ligados al computador para el cual fueron diseñados, esto crea una dependencia fundamentalmente hacia el equipo, no puede ser leído ni entendido con

facilidad por los seres humanos.

- **Alto nivel:** Los lenguajes de programación de alto nivel fueron creados con la finalidad de un mayor entendimiento ya que utiliza palabras conocidas por los seres humanos y hace que dar instrucciones al computador sea más sencillo; son generalmente portables lo que significa una independencia hacia el equipo donde se ejecutan y así permite a los programadores enfocarse en el código que están realizando.

En sí, los lenguajes de programación, tanto de bajo nivel como de alto nivel, fueron diseñados para dar una respuesta a ciertos problemas, para ejecutar acciones que el usuario requiera dependiendo de su necesidad y así, solventarla de una manera eficaz. Así mismo como el ser humano progresa, va aptándose a los lenguajes de programación que van surgiendo, ya que cada vez más y más tecnologías van desarrollándose, lo que lleva al ser humano a incrementar su conocimiento una era digital que está en constante avance (Alay & Sevillano, 2022).

2.3.3 Lenguajes de programación: Python

Python es considerado como un lenguaje de alto nivel debido a que contiene implícitas estructuras de datos como conjuntos, listas, tuplas y diccionarios; estas estructuras permiten realizar tareas complejas en una reducida cantidad de líneas de código. Así mismo, cuenta con facilidades para la programación orientada a objetos, funcional e imperativa, por lo que se lo reconoce como un lenguaje multiparadigmas. Python fue creado por Guido van Rossum, un programador holandés a finales de los 80 y principio de los 90 y fue basado en el lenguaje ABC (Challenger-Pérez et al., 2014).

Es considerado como el lenguaje de programación más popular para la ciencia de datos debido a que permite olvidarse de la parte tediosa de la programación y facilita en centrarse en las ideas y de cómo poner estos conceptos en acción a través de un entorno amigable (Mirjalili & Raschka, 2020).

Para el entrenamiento de nuestro modelo de Machine Learning el utilizar Python en su versión 3.10.5. ya que esto es un requerimiento de la empresa.

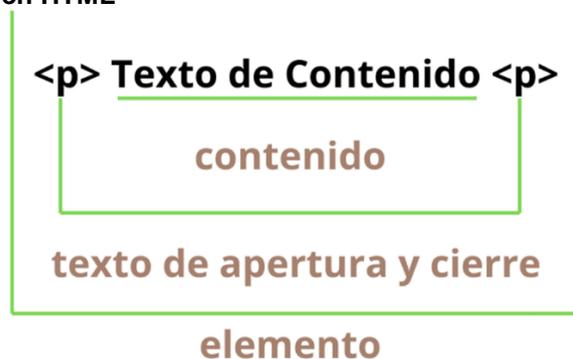
2.3.4 Framework: Laravel

Cíceri (2019) describe al framework como un subsistema general que contiene estructuras y componentes de software interconectados y predefinidos que sirven como fundamento para el desarrollo y la organización del sistema; consta de una estructura de archivos y carpetas para mantener el código organizado, una arquitectura que permite el desarrollo del proyecto, seguridad, robustez y soporte.

Según Arias (2017), define a PHP como un lenguaje interpretado libre que tiene como utilidad el desarrollo de aplicaciones dinámicas que interactúan de la mano con un servidor en la web. MediaWiki, Facebook, WordPress, Drupal, Magento utilizan este lenguaje en sus sistemas web. Surgió a mediados de 1994 por Rasmus Lerdorf y entre sus características principales están la velocidad, robustez y portabilidad. Es un lenguaje altamente modularizado, siendo ideal para su uso e instalación en servidores. Se considera de dominio específico es decir que es destinado netamente para el desarrollo web.

Laravel es el framework de PHP⁴ más utilizado en la actualidad; posee una buena curva de aprendizaje, tiene una documentación completa, simple y gratuita. En 2011, fue creado por Taylor Otwell debido a la necesidad de que exista un framework que incluyera funcionalidades fundamentales para el desarrollo web como una autenticación incorporada; esto lo llevo a crear el framework que hoy conocemos como Laravel. Laravel trabaja de la mano con HTML⁵ que genera un contenido dinámico desde el servidor hacia el navegador.

Figura 5
Etiqueta en HTML



Nota: Ejemplo de una etiqueta en HTML para definir un párrafo.

⁴ Hypertext Preprocessor (PHP)

⁵ HiperText Markup Language (HTML)

HTML hace referencia a un lenguaje de programación más comúnmente utilizado para construir páginas web; implementa una codificación de etiquetas las cuales se interpretan por el navegador de internet, así como Google Chrome o Microsoft Edge y este pasa a mostrar la página web al usuario (Luna, 2024).

Como parte de la creación del sistema web en el módulo de reportería cuyo fin es la generación automática de los informes utilizamos Laravel en su versión 11.9 que utiliza el lenguaje de programación HTML. De esta manera, nos alineamos el estándar que maneja la empresa para facilitar una futura integración de manera más efectiva y eficiente.

2.3.5 Blade

Blade se establece como un conjunto de plantillas nativo del framework de Laravel; permite la implementación de código de PHP con HTML de forma legible y sencilla debido a su sintaxis amigable para el usuario. Blade realiza un procesamiento y organización de vistas conocida como layout, lo cual es considerada como su mayor ventaja. Otra ventaja del uso de Blade, es de que es flexible en el uso de sus vistas ya que permite el manejo de PHP (Ortega et al., 2019).

Como parte de requerimiento por parte de la empresa, para el desarrollo de la interfaz gráfica del usuario en el prototipo de sistema web, implementamos Blade, una de las plantillas de Laravel en la cual se utiliza código de PHP con HTML.

2.3.6 Bootstrap

Bootstrap es un framework para el diseño y creación de sistemas web responsivos y modernos como requerimiento por parte de la empresa de seguridad con la finalidad de alinearnos al estándar previamente establecido que manejan y poder optimizar una futura integración del prototipo a sus sistemas.

Bootstrap fue creado por Mark Otto y Jacob Thornton, ambos siendo desarrolladores de Twitter, con la finalidad de optimizar el diseño de los sistemas web con un framework que sirve como guía para los desarrolladores de como emplear un buen manejo de estándares y prácticas. Según Ortega et al. (2019) entre sus características principales tenemos las siguientes:

- Facilidad en creación de sistemas web altamente adaptables a cualquier dispositivo.
- Sistemas web visualmente organizados.
- Comunidad amplia y activa.
- Agradable experiencia de desarrollo.

Para el desarrollo de nuestro prototipo, utilizamos Bootstrap para un desarrollo de interfaz atractiva en su diseño ya que fue requerimiento por parte de la empresa de seguridad y así cumplir con sus estándares requeridos en cuanto al diseño.

2.3.7 Librerías: FPDF

En contexto de lenguajes de programación, se determina a una librería como un conjunto de funciones, clases o métodos que se encuentran en un paquete con el fin de ser reutilizados para ahorrar tiempo y esfuerzo en el desarrollo por parte del usuario.

Conforme a Santillán (2015) , FPDF es una clase escrita en PHP que directamente desde el mismo, genera documentos en formato PDF⁶. La F de FPDF significa Free de gratis y libre, esta librería se usa para cualquier propósito y se modifica a su gusto para satisfacer las necesidades que el usuario presente.

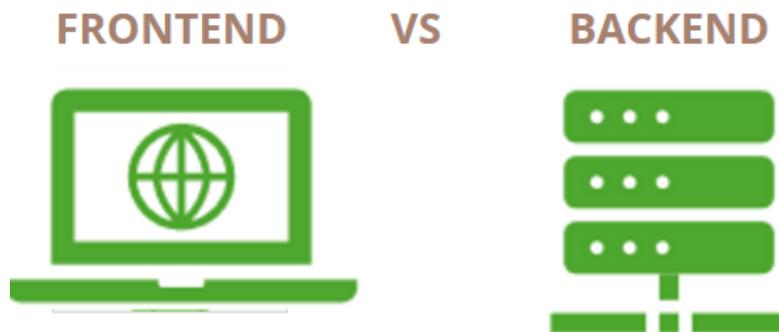
Para la generación del diseño del informe dinámico con las incidencias y recomendaciones para la empresa de seguridad, se utilizó la librería FPDF⁷ que fue impuesta por parte de la empresa como un requerimiento para el diseño de los informes, dado a que esta librería permite realizar un diseño flexible en cuanto a informes en PDF y así, brindar una presentación acorde al estándar que maneja la empresa.

⁶ Portable Document Format

⁷ Free Portable Document Format

2.3.8 Frontend y Backend

Figura 6
Frontend vs Backend



Nota: Representación gráfica entre frontend y backend.

En el desarrollo web, el Frontend se lo denomina como la interfaz gráfica que interactúa directamente con el usuario mediante el sistema web utilizado; engloba toda representación gráfica y funcionalidades básicas. Por otro lado, tenemos al Backend que se encarga de la manipulación de los datos que provienen desde el Frontend; por sí solo, el Backend no tiene mucho sentido, es necesario que exista un Frontend y una base de datos, es la conexión entre ambos (Pérez Ibarra et al., 2021).

Tabla 3
Comparativa entre Frontend y Backend

Frontend	Backend
Parte del sistema visible e interactiva para el usuario.	Parte del sistema que maneja la lógica y procesamiento detrás de la interfaz.
Tiene la funcionalidad de crear interfaces gráficas que interactúan directamente con el usuario.	Tiene la funcionalidad de gestionar y procesar los datos además de servir información al Frontend.
Entre los lenguajes comunes usados están: HTML, CSS, JavaScript (y frameworks como React, Angular, Vue).	Entre los lenguajes comunes usados están: PHP, Python, Ruby, Java, Node.js, C#.
Su principal objetivo es de hacer que la aplicación sea visualmente atractiva y fácil de usar.	Su principal objetivo es de asegurar que la aplicación funcione correctamente, sea segura y eficiente.

Nota: Cuadro comparativo que explica las diferencias entre Frontend y Backend.

2.3.9 Base de Datos

Conforme a Beynon-Davies (2018), una base de datos se define como una colección que tiene como característica principal el ser organizada y compuesta por datos; su finalidad es de representar un contexto definido en base a la organización para la cual fue creada. Es importante que posea integridad, garantizando que los datos sean precisos y no llegue a existir inconsistencias en el manejo de su contenido.

2.3.10 Base de Datos: MYSQL

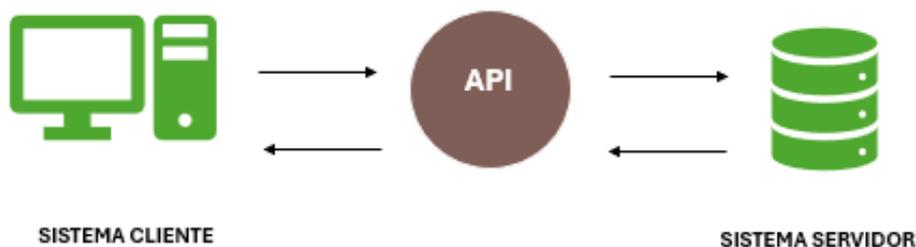
MySQL es un reconocido sistema gestor de bases de datos ampliamente destacado por su simplicidad y notable rendimiento. Es una opción llamativa para aplicaciones comerciales, como de entretenimiento justamente por tener una mínima curva de aprendizaje , por la rapidez en su implementación y un alto grado de estabilidad (L. A. C. Santillán et al., 2014).

Empleamos la base de datos de MySQL con la finalidad de gestionar los datos, por solicitud específica de parte de la empresa de seguridad; alineándonos con el estándar utilizado por la organización para facilitar una futura integración.

En la siguiente gráfica se muestra la interacción entre los sistemas y la API

2.3.11 API

Figura 7
Funcionamiento de API



Nota: Refleja la interacción de los sistemas con la API.

Una API⁸ es un conjunto de reglas y estándares definidos por un software que permiten establecer una comunicación de forma estructurada y eficiente entre los diferentes sistemas; estas reglas actúan como un intermediario que optimiza el intercambio de información y la interacción entre un sistema cliente que realiza la formulación de la petición y un sistema servidor que emite una respuesta hacia esa petición (Torres-Salinas & Arroyo-Machado, 2022).

2.3.12 Entorno de Desarrollo Integrado

De acuerdo con Red Hat (2023) un entorno de desarrollo integrado (IDE) es un

⁸ Interfaz de Programación de Aplicaciones (API)

implemento de software que agiliza la elaboración de aplicaciones combinando diferentes componentes que utiliza comúnmente un programador que desarrolla aplicaciones, estos componentes se juntan en una sola interfaz de usuario (GUI). Aquí podremos encontrar el editor de código que cuenta con funcionalidades como resaltar sintaxis, autocompletado dependiendo del lenguaje y la detección de errores de manera constante. Provee automatización en tareas repetitivas como la compilación de código fuente a binario, para después empaquetarlo y ver cuáles son los resultados de las pruebas de dicho proceso. Cuenta con un depurador de código esto ayuda a identificar y encontrar errores que estén en conflicto mostrándolos de forma gráfica.

2.3.12.1 Características distintivas

Entre los entornos de desarrollo integrados (IDE) llegan a apreciarse varios aspectos que muestran sus características más importantes, la compatibilidad con los lenguajes de programación logra que algunos IDE logren ser multilenguaje y otros admitir solo ciertos lenguajes en específico. Por ejemplo, IntelliJ se creó con el objetivo de ser utilizada con Java únicamente mientras que otros entornos como Eclipse llegan a tener la posibilidad de ser usados diferentes lenguajes como lo es Python, Java y XML, así haciéndolos multilenguaje. Se recomienda que antes de instalar o ejecutar un IDE veamos si es compatible con el sistema operativo que maneja el equipo, por ejemplo podrán existir IDE creados únicamente para sistemas Linux y otros para Windows, esto concluye que siempre debemos anticiparnos antes de seleccionar un IDE no solo porque únicamente trabaja con el lenguaje que queremos sino que también es 100% compatible con nuestra máquina; por otro lado cuando se trabaja en la Nube no se requiere ningún tipo de especificaciones técnicas o software requerido ya que solo funciona a base de internet. (Red Hat, 2023)

Las funciones de automatización de los IDE se distinguen de las funcionalidades básicas como un editor de texto, compilaciones en periodos de tiempo y depurador debido a la capacidad que tiene para optimizar y agilizar los procesos difíciles del desarrollo; Aquí se incluyen funciones como la refactorización de código que reorganiza y mejora el código sin cambiar la funcionalidad del mismo, búsqueda avanzada de código para encontrar y solucionar problemas de forma rápida y el componente de soporte para integraciones y entregas constante. Las características mencionadas ayudan a darle a los usuarios una personalización dentro

del flujo de trabajo y en caso de requerirlo poder adaptarse a las necesidades de un proyecto. (Red Hat, 2023)

2.3.12.2 Ventajas

Red Hat (2023) menciona las ventajas cuando se utiliza un IDE dentro de un proyecto de desarrollo de software con un equipo de trabajo:

- El IDE permite trabajar con diversas herramientas o extensión gracias a la gran capacidad de compatibilidad que tiene además de poder integrar herramientas de una manera más fácil, eso beneficia al trabajo en equipo ya que los desarrolladores no necesitan esperar o tomarse el tiempo de preconfigurar nuevas herramientas, eso hace que se trabaje de manera eficiente además tener varias herramientas en un solo entorno.
- Existe una integridad cuando se desarrolla debido a que si bien todo lo desarrollan en diferentes instancias cada uno utiliza las mismas herramientas adicionales esto mejora el flujo de trabajo y fortalece que cada programador tenga la posibilidad de desarrollar de la forma que más le guste.
- Algunos IDEs incluyen herramientas que puede generar código de manera automática con solo pedirlo, esto reduce el tiempo de implementación y la culminación de un componente más rápido.

No es obligatorio desarrollar aplicaciones utilizando un IDE. Por ejemplo, se puede utilizar editores de texto como Vim o Emacs que son programas básicos de escritura de código; estos editores no incluyen herramientas específicas, debido a esto si el desarrollador las necesitara tendrá que instalarlas por separado, algunas de estas son compiladores, depuradores o sistemas de control de versión siendo esta instalación y configuración de forma manual; En un entorno empresarial esto cambia teniendo ventajas como ahorrar de tiempo, estandarizar el ambiente de trabajo y automatizar tareas esenciales como resultado se incrementa la productividad y la organización. (Red Hat, 2023)

2.3.12.3 Tipos de IDE

De acuerdo Amazon Web Services Inc. (2024) los entornos de desarrollo integrado pueden ser de varios tipos por eso es importante saber cuál usar, esto depende de lo que se pretenda desarrollar debido a que en un proyecto de software se puede llegar a dividir en diferentes equipos que se encarguen de un área en específico; a pesar de eso se pueden clasificar en dos: IDE locales, IDE en la nube.

2.3.12.4 IDE locales

Los IDE locales son aquellos entornos que se manejan en el equipo local, se descargan y se instalan directamente desde internet, dependiendo en que área (Frontend, Backend, base de datos) se pretenda desarrollar se deberá escoger el adecuado. Su característica principal es la personalización permitiendo la descarga e instalación de bibliotecas o librerías para crear diferentes proyectos con diferentes estilos de codificación, también se caracterizan por ser plataformas multilenguaje logrando manejar diferentes lenguajes de programación; Sin embargo debido a su personalización pueden llegar a crearse proyectos con requisitos que pueden consumir muchos recursos resultando en un bajo rendimiento del equipo si no se tienen buenas prestaciones.

2.3.12.5 IDE en la nube

Así como existen IDE locales también existen en la nube, ambos se pueden diferenciar, pero tiene la misma característica para editar, compilar y ejecutar código sin ninguna restricción, La ventaja de utilizar un IDE en la nube es que se omite el paso de la instalación que puede ser tediosa dentro del dispositivo local. Si bien ambos permiten realizar los mismo uno lo hace mejor que otro de la siguiente manera: Ambiente de desarrollo ya preinstalado, acceso de inmediato a través de internet y mejoras en tiempos de respuesta. (Amazon Web Services, Inc., 2024b)

Para la elaboración de este proyecto se utiliza VisualStudio como entorno de desarrollo porque determinamos que cumple con todos los requisitos que debe tener el proyecto siendo compatible con cada tecnología que tengamos que implementar como por ejemplo librerías, extensiones y plugins este tendría el objetivo de ser usado en el transcurso del proyecto mientras se realice el desarrollo del sistema web basado en Machine Learning. Se utiliza este tipo software ya que ayudarían al proyecto a

facilitar su desarrollo, evitar un mal uso de la sintaxis y realizar la compilación del código para previsualizar desde el navegador.

2.3.13 Plataforma de control de versiones

Microsoft (2023) define al sistema de control de versiones como un método que ayuda a los desarrolladores a tener un seguimiento de las diferentes instancias que ha tenido su código a lo largo del tiempo, cada vez que la persona realiza nuevos cambios dentro de su proyecto crea una nueva versión de la instancia en la que se encuentra; no obstante ellos cuentan con una instancia principal donde podrán ver los últimos cambios, no obstante pueden regresar a las diferentes versiones de su código sin afectar a ninguno de los cambios que había realizado.

Si no existiría el sistema de control de versión los desarrolladores tendrían diferentes instancias de su proyecto o código guardadas en distintos archivos en su equipo, esto provoca que se consuma espacio y sea poco útil; al no tener una versión principal tiende a causar confusión y quizás eliminar una copia del código incorrecto. Para esto se utilizan estos sistemas que mantienen administrado todas las versiones posibles que existan del código.(Microsoft, 2023)

Existen diferentes maneras de generar clasificaciones sobre los sistemas de control de versiones, generalmente los dos más comunes son los centralizado y distribuidos. Los centralizados basan se crean y generan dentro de un repositorio siendo mucho más eficiente; Por otro lado, si optamos por distribuidos el código o archivos se va a mantener en diferentes repositorios.

A continuación, se presenta una tabla de clasificaciones del tipo de sistemas de control de versiones (GitLab, 2023):

Tabla 4:
Clasificación de sistemas de control de versiones

Tipo	Descripción	Uso común
Distribuido	Los usuarios tienen acceso a un repositorio desde varias ubicaciones donde ellos cargaran sus cambios o archivos, al contar con varios repositorios esto puede mejorar la flexibilidad.	Puede ser conveniente para aquellas personas que laboran desde diferentes ubicaciones o cuentan con un trabajo remoto.
Centralizado	Todo el almacenamiento se centra en un repositorio único que sea que se encuentra en un servidor o maquina local.	Suele usarse más en proyecto de desarrollo de software de cualquier ámbito debito a que se necesitara

Tipo	Descripción	Uso común
		compartir el código y rastrear archivos,
Basado en bloques	Presenta el bloqueo de archivos importantes dándole la preferencia a una persona única siendo esta la que puede modificar un archivo a la vez.	Se aplica en proyecto donde se requiera evitar conflictos ya sea porque se edita, borrar o actualiza algo como diseño CAD o documentos.
Optimista	Aquí cada desarrollador trabaja de manera individual su código sin tenerlos conectado a nada, en caso de realizar algún cambio este es enviado al servidor para que lo analice y fusione.	Útil en entornos que requieran una revisión más detallada sobre cuáles fueron los cambios y ambientes con menos conflictos.

Nota: La siguiente tabla muestra las 4 diferentes maneras en las que se pueden clasificar los sistemas de control de versiones.

El sistema de generación de informes basado en Machine Learning utiliza un repositorio en GitHub donde se almacena todo el código fuente del proyecto y un software de control de versiones distribuido donde existe una instancia principal que contendrá los últimos avances del sistema, cada desarrollador tendrá una instancia secundaria para trabajar de manera independiente así evitando que trabajen sobre la misma instancia lo que puede ocasionar confusiones y una desorganización ya que no se podría establecer que avances ha realizado cada uno. Se utiliza GitKraken como software de control de versión para este proyecto al ser compatible con GitHub lugar donde se encuentra el repositorio con el código fuente del sistema.

Capítulo III: Metodología De La Investigación

En el siguiente capítulo se explica la metodología aplicada para desarrollar el proyecto de estudio y la metodología utilizada para la construcción del prototipo.

Tipo Investigación y enfoque metodológico

Para el desarrollo de este proyecto se emplea un enfoque de investigación mixta descriptiva, se usa investigación mixta porque nos permite trabajar con datos tanto cualitativos como cuantitativos donde se incluyen datos numéricos, visuales, registros, verbales, textuales con el fin de tener un mejor entendimiento sobre el objeto de estudio; esto permite contrastar datos cuantificables y no cuantificables así logrando comprobar o no los resultados esperados durante la investigación; este tipo de investigación nos brinda un mayor panorama acerca del fenómeno o problemática planteada y por consiguiente podemos alinearlos a los objetivos, alcances y a la pregunta de investigación planteada. (Hernández, 2018)

Dentro del estudio, la investigación mixta por su lado cualitativo corresponde a la interpretación de los requerimientos durante el levantamiento de información con la parte interesada de la empresa el cual se dio de manera verbal (reunión de trabajo con el jefe de PROTEMAXI), la participación de la empresa fue clave debido a que se pudo conocer cuáles eran las características y funcionalidades del proceso actual de generación de informes y cómo este influye en las operaciones de la empresa; esto permitió definir cuáles iban a ser las características y aspectos del prototipo ajustándolo a la automatización y al análisis con Machine Learning teniendo en cuenta que dicha solución debe de estar alienada a las necesidades y requerimientos de la empresa de seguridad. Lo correspondiente al lado cuantitativo de la investigación se realizó una revisión general de los reportes de las incidencias registradas por el personal de seguridad de la empresa que corresponden a cada cliente, la data que proporciona estos reportes ayuda a la alimentación y permite el diseño del modelo de machine Learning que es un complemento esencial para la generación de reportes del prototipo.

El enfoque descriptivo se aplica porque consta de la búsqueda y recolección de información que gira en torno al objeto de estudio y que es útil para contestar la

pregunta de investigación; permite definir y comprender el elemento investigado utilizando mediciones o criterios, es importante definir de donde proviene la información.(Nieto, 2018). Dentro del proyecto, este tipo de enfoque ayuda a entender de donde tienen que venir la información relacionada a los registros de incidencias, estos se analizan e interpretan para entender el comportamiento y funcionamiento de las variables para después poder relacionarlo con el prototipo del sistema; una vez interpretada la información recolectada es más fácil determinar que variables son las que más las más óptimas y relevantes para obtener mejores resultados en los informes.

Población y muestra

Para la selección de la población se utilizan 409 registros de novedades que fueron proporcionadas por PROTEMAXI, dichos datos representan registros de incidencias que han sido reportadas desde el año 2023 hasta el año 2025 y corresponden a los diferentes clientes que tienen contratado el servicio de seguridad; estos datos son esenciales para el entrenamiento del Modelo de Machine Learning y también para su evaluación, el análisis de estos datos con el modelo ayuda a encontrar patrones y tendencias para generar recomendaciones dentro del informe.

Puesto que la población no es tan grande y puede ser manejada, se opta por usar la misma población como muestra dentro del estudio, reducir la cantidad de datos podría afectar a la efectividad del modelo de Machine Learning.

Los 409 registros de novedades contienen datos numéricos, de tipo fechas, texto que detallan el título, detalle, estado y fecha de la novedad; a nivel general existen valores algunas columnas que contiene una inconsistencia de datos o un campo vacío. A continuación, se mencionan y describen cada uno de los campos del dataset:

- **Id:** columna tipo número con cero valores faltantes
- **título:** columna tipo texto con cero valores faltantes
- **id_agente:** columna tipo número con cero valores faltantes
- **id_cliente:** completo columna texto número con cero valores faltantes

- **id_localidad:** columna tipo número con cero valores faltantes
- **id_garita:** columna tipo número con cero valores faltantes
- **fecha_novedad:** columna tipo fecha con cero valores faltantes, pero presenta campos donde tiene una mezcla de campo tipo fecha y texto
- **TIPO_NOVEDAD:** columna tipo texto con 179 campos vacíos
- **tipo_incidente:** columna de tipo texto con 401 campos vacíos
- **tipo_hallazgo:** columna de tipo texto con 385 campos vacíos
- **tipo_actividad:** columna de tipo texto con 295 campos vacíos
- **tipo_protemaxi:** columna de tipo texto con 407 campos vacíos
- **detalle:** columna de tipo texto con cero valores faltantes, pero presenta 1 valor de cero en un registro
- **personas_involucradas:** columnas de tipo texto con 1 campo vacío, presenta valores numéricos en diferentes registros
- **lugar_involucrado:** columna de tipo texto con 184 campos vacíos
- **centralista:** columna de tipo texto con 284 campos vacíos
- **estado_novedad:** columna de tipo texto con 179 campos vacíos
- **estado:** columna tipo texto con cero valores faltantes

Se revisa el formato de los datos proporcionados y se encuentran columnas que tienen tipos de valores mezclados que contienen, donde se interpretan inconsistencias erróneas en los datos; a continuación, se detallan cuáles son esas columnas:

- **id:** presenta el numero de la novedad registrada sin embargo se encuentran valores como "CLIENTE 30", "CLIENTE 33"
- **id_cliente:** se debería alojar el número de id del cliente sin embargo se

muestra un texto que describe el nombre del cliente, Ejemplo “Cliente 1”, “Cliente 2”.

- **personas_involucradas:** hay registros donde la columna tiene “0” y otro registro con “Xxx” lo que genera una inconsistencia la interpretar los datos.
- **fecha_novedad:** presenta valores tipo fecha con hora y campos con una combinación de fecha más el nombre del cliente, ejemplo: “7/11/2023 15:CLIENTE 3:58”, “CLIENTE 3/2/2024 8:30:00”.
- **detalle:** presenta un valor de “0” en un registro cuando es columna de tipo texto

La presencia de datos combinados o con valores inconsistentes puede alterar la operación del modelo de Machine Learning y su carga a la base de datos, por lo que se realiza un tratamiento a los campos de las columnas que no se adecuen al formato del que proceden (numérico, fecha, texto), este proceso forma parte del desarrollo del modelo de Machine Learning.

Instrumentos de recolección de datos

Para la obtención de la información del objeto de estudio se aplica la entrevista como técnica de recolección de datos, como se mencionó anteriormente esta entrevista fue realizada a 2 personas de la empresa que tenían conocimiento acerca del proceso de generación de informes y el planteamiento que tenían para mejorarlo; durante la reunión se realizaron una serie de preguntas con el fin de recabar información que sea relevante para el proyecto y para el desarrollo del prototipo, así logrando establecer requerimientos y funcionalidades.

En la entrevista se trató sobre el porqué se requiere de una solución de Tecnología de la Información para automatizar el proceso de generación de reportes mediante la aplicación de la inteligencia artificial, mencionando funcionalidades y aspectos generales que debe tener el prototipo; se indagó también acerca del funcionamiento del proceso actual y que esperan de la utilización del prototipo como ayuda en su operación.

Entre los puntos importantes se mencionaron las herramientas de desarrollo que se requiere utilizar, la forma y funcionalidad del prototipo, la data que se analiza, cumplimiento acorde al formato del informe, variables bajo las que se requiere generar el informe; los puntos mencionados son ampliados dentro la construcción del prototipo dentro de la fase de levantamiento de los requerimientos.

Metodología de desarrollo de prototipo

La metodología de desarrollo de software en cascada pertenece a la familia de las metodologías tradicionales en la implementación de sistemas, se divide en 5 etapas donde se establecen objetivos y requerimientos, se diseña y desarrolla software y se realizan pruebas para comprobar los objetivos y resultados esperados inicialmente; debido a la forma en la que se implementa que es clara y estructurada se logra determinar los objetivos de manera rápida.

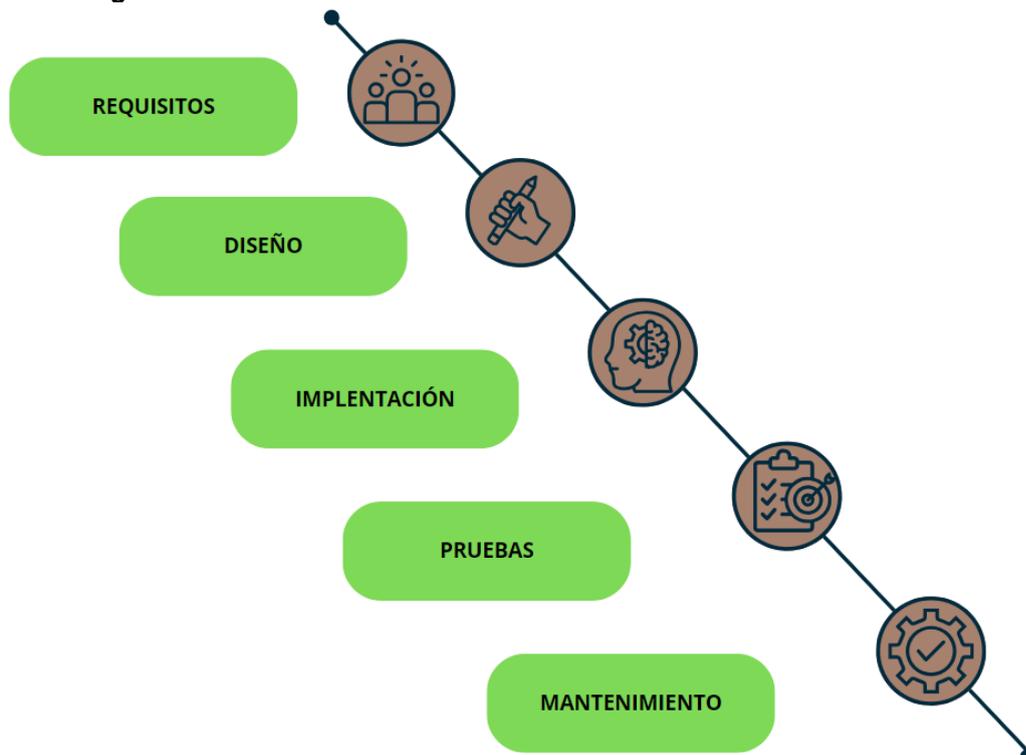
J. A. Barrera & Barrera (2020) define a la metodología de cascada como simple pero efectiva cuando se requiere ejecutar un proyecto siempre y cuando este predefinido y se haya dejado claro todo lo que se va a realizar, es unas de las más fáciles de implementar debido a que su proceso de revisión y desarrollo que sigue pasos en concreto sin tendencia al cambio, esto lo vuelve ideal para proyectos pequeños donde se haya establecido los alcances y objetivos del proyecto teniendo en cuenta lo que se desarrollará y las limitaciones que existen.

Algo relevante de esta metodología es que su primera etapa es clave ya que dependiendo de qué tan bien se haya realizado la definición de objetivos y requerimiento dependerá el éxito del proyecto debido a que en estas etapas se realiza el levantamiento de requerimientos la planificación; en la metodología se deben seguir los pasos de manera secuencial y no se quiere realizar un retroceso a un paso anterior evitando la repetición de pasos lo cual beneficia la entrega del producto final.

Aplicando esta metodología se llega a tener una mejor trazabilidad del proyecto facilitando la culminación de hitos, actividades y el no causar retrasos realizando cambios recurrentes durante el desarrollo del sistema ya que existen requisitos establecidos.

A continuación, se ilustran las fases de la metodología seleccionada:

Figura 8
Metodología cascada



Nota: Se muestra un modelo de las 5 fases que conforman la metodología en cascada, aquí se mantiene una ruta secuencial donde la fase más importante es la primera.

Las fases del modelo de metodología en cascada conforman diferentes partes del proyecto que se ha sido desarrollado, por esta razón es importante describir como se integran estas fases dentro del proyecto:

(Olivera & Alonso, 2021) definen a la primera etapa del modelo como la primordial donde se establecen objetivos, limitaciones y cuáles serán las características y reglas de negocio que contendrá el prototipo, se resalta que de esta fase dependerá gran parte del éxito del proyecto. En la fase de diseño se estructuran las tecnologías tanto a nivel de hardware como software a utilizar en el sistema, aquí se realiza también una planificación de actividades acorde a las especificaciones técnicas que hayan sido solicitadas por el usuario durante el levantamiento de información.

En la fase de implementación se realiza el desarrollo del prototipo del sistema de generación de informes junto con la construcción del modelo de Machine Learning, en esta fase deberá existir una relación y consistencia con las fases de requerimientos y diseño ya que las dos primeras fases determinarán cual será el aspecto y reglas

que definirán el software; esto nos asegura presentar un software en base a lo solicitado así evitando realizar cambios inoportunos.

Una vez finalizada la implementación del prototipo se realizarán pruebas donde se busque evaluar como el prototipo del sistema responde a cada una de las características y funcionalidades que fueron plasmadas en la fase de implementación, si durante las pruebas se debe realizar correcciones se las hará siempre y cuando este dentro los objetivos; la evaluación es clave porque permite darle validez al software midiendo la capacidad y calidad con la que genera informes y analiza datos con el fin de obtener resultados acorde a los esperado.

Capítulo IV: Propuesta Tecnológica

El siguiente capítulo se explica cómo se lleva a cabo el desarrollo del proyecto para realizar la construcción del prototipo del sistema, el contenido del capítulo se estructura de acuerdo con las metodologías indicadas en el capítulo 3.

Levantamiento de información

El levantamiento de información se realizó en dos reuniones programadas que se tuvo con el representante legal de la empresa auspiciante donde se explicó cuáles eran las necesidades que tenía la empresa y cuáles eran los aspectos que debía de tener el prototipo, durante la reunión se tomó nota de dichas necesidades y funcionalidades. A continuación, se detalla la información recopilada:

1. El sistema proporciona a los usuarios informes generados de manera automatizada utilizando inteligencia artificial y utilizando data proporcionada la empresa, la data que se utilizará para generar el informe proviene de las incidencias reportadas que se haya realizado al mes por lo guardias, es decir este informe generado será del mes y se deberá generar por cliente.
2. Actualmente el proceso de elaboración de informes es realizado por el analista de reportes de la matriz de PROTEMAXI el cual revisa reporte por reporte de las incidencias que han sido registradas durante todo el mes, esta información deberá ser analizada por la persona responsable y elaborar el informe correspondiente de acuerdo al cliente y al mes; la empresa tiene la necesidad de automatizar este proceso donde toda esta información generada día a día se pueda consolidar en un informe final sin la necesidad de una persona.
3. El sistema debe ejecutarse a través de la web y su desarrollado tiene que ser hecho con Laravel tanto a nivel de back-end como front-end. El sistema se realiza utilizando Laravel porque en caso de realizar una integración con el sistema de registro de incidencias de PROTEMAXI deberá existir la facilidad y compatibilidad entre ambos sistemas.

4. El informe generado por el prototipo deberá seguir el formato estandarizado que se tiene en la entrega de informes a los clientes, esto quiere decir que cuando se vaya a realizar la generación del informe este deberá contar con las mismas características y campos que el informe de entrega por parte de PROTEMAXI.
5. El prototipo del sistema deberá ser desarrollado en un entorno fuera del sistema actual que tiene PROTEMAXI el cual permite registrar incidencias reportadas por el personal de seguridad, esto con el fin de mantener la integridad del sistema y no causar posibles afecciones.
6. La información proporcionada de las novedades serviría como data de procesamiento para la elaboración del informe del mes por cliente, entre los datos mencionados en la reunión se encuentran descripción, título, cliente, fecha, hora y nombres; la empresa menciona que estos datos se utilizarán para realizar el análisis y así desarrollar el informe.
7. El sistema deberá manejar una relación estricta donde el número de informes generados deberá estar ligado al número de clientes que tenga, así mismo las fechas de las incidencias reportadas deberán estar acorde al mes y al cliente.
8. Dentro del front-end del Sistema deberá mostrarse la lista de clientes que estén dentro de la data proporcionada ya que debido a temas de integridad de datos no se puede proporcionar la lista completa; en esta lista de cliente el usuario podrá generar los informes por clientes de acuerdo con el mes.
9. No se deberá mostrar el nombre del cliente que tiene contratado el servicio de PROTEMAXI con temas de confidencialidad e integridad de datos.
10. La data proporcionada con respecto a las novedades que se reportan día a día es la data que se utilizará para el análisis utilizando el modelo de Machine Learning.
11. Dentro del informe existe información que se genera del resultado del análisis de la data de las novedades por el modelo de Machine Learning y

data que proviene directamente de la base de datos que sirve para llenar campos y tablas que son precargadas.

Definición del prototipo

De acuerdo con la información recopilada en el levantamiento de los requerimientos se traducen a las características de la implementación, el prototipo se define como un sistema web que permite generar informes mensuales por cliente de las incidencias que son reportadas día a día, la automatización de los informes trae consigo la utilización de Machine Learning por lo que no se realiza una automatización convencional y tampoco 100% de extracción e inserción debido a que existe información dentro del informe que es analizada y puesta como información de valor.

Características y funcionalidades

A continuación, se mencionan las características y funcionalidades:

- Generación automatizada de informes que sigan el formato del informe estándar de la empresa
- Los resultados que genere el análisis del modelo de machine Learning se incluyen dentro del informe
- En la pantalla se debe mostrar el listado de cliente y el mes o periodo del informe que se va a generar junto con un botón para generar el informe
- El modelo de machine Learning debe analizar y generar resultados respecto a la data de novedades mencionada en la población y muestra del capítulo 3
- El informe generado deberá descargarse en un archivo en formato PDF

Requisitos funcionales y técnicos

A continuación, se indican los requisitos funcionales y técnicos:

- El prototipo debe cumplir con los siguientes requisitos:

- Se utiliza Laravel como Framework para el desarrollo web del sistema.
- Se utiliza MySQL como gestor de base de datos.
- Los informes generados deben seguir con el formato estándar de informes de PROTEMAXI.
- Los informes se generan de acuerdo con el cliente y mes.
- Los informes deben ser generados en formato PDF.
- Debe ser un sistema independiente al sistema actual de PROTEMAXI.

Herramientas para el desarrollo

A continuación, se mencionan las tecnologías utilizadas para el desarrollo del software:

Tabla 5
Clasificación de sistemas de control de versiones

Capa	Tecnologías
Front-end	Laravel (HTML, CSS, PHP)
Back-end	Laravel (PHP)
Base de Datos	MySQL, SQL, XAMPP
Desarrollo	GitKraken, Composer, VisualStudioCode, GitHub
Procesamiento inteligente	Python, Excel

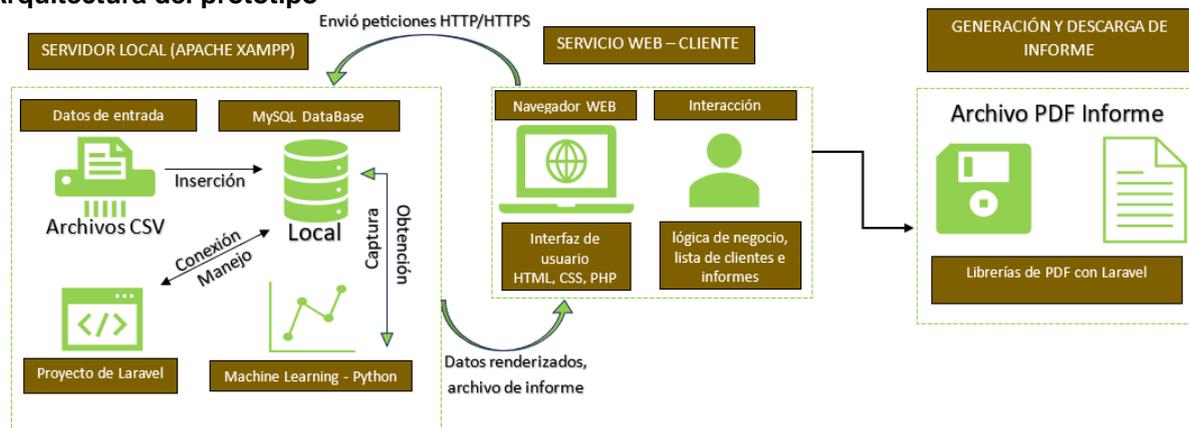
Nota: La siguiente tabla muestra las herramientas que se utilizan para la construcción del prototipo.

Arquitectura del sistema

Para la representación del sistema se elabora una ilustración que permite tener una visión de cuáles son los componentes que lo conforman y como estos se relacionan entre sí para hacer que el sistema funcione de manera correcta.

A continuación, se presenta el grafico de la arquitectura del sistema:

Figura 9
Arquitectura del prototipo



Nota: Se muestran el componente se estructura el prototipo de sistema de generación de reportes.

La arquitectura del sistema al tener diferentes componentes puede ser dividida en capas de acuerdo con la naturaleza o funcionamiento que tenga cada uno de ellos para el prototipo, aquí:

1. El almacenamiento y procesamiento es el componente clave debido a que aquí se obtienen los datos que servirán para la generación del reporte, así como los datos que se utilizarán para ser analizados por el modelo de Machine Learning; aquí interviene herramientas de alojamiento de datos como el tipo de base de datos que se utiliza y lenguajes para construir modelos para procesar datos como lo es Python.
2. El componente generador es la capa principal debido a que automatiza el proceso de generación de reportes y lo combina con los resultados arrojados por el modelo de Machine Learning; en este componente se realizan las configuraciones de conectividad con la base de datos de incidencia. Está construido bajo los requerimientos de la empresa, esto es utilizando Laravel.
3. El ultimo componente es el producto final es decir el informe el cual es descargable y puede ser visualizado en formato PDF, el resultado de este componente depende de cómo se haya realizado la implementación de la data y la construcción del sistema, esto quiere decir que si falta data el informe puede verse incompleto y si existe poco desarrollo en el software puede tener problemas al momento de visualizarse.

solicitó acceso a la información del registro de incidencias que la empresa tenía porque aquí se encontraba información como fecha en que se registró la incidencia, la localidad en donde ocurre, personal que la registra, los detalles que contiene esta, personas involucradas, etc.

A continuación, se presenta el archivo Excel con el registro de incidencias proporcionado por PROTEMAXI:

Figura 11
Registro de Incidencias

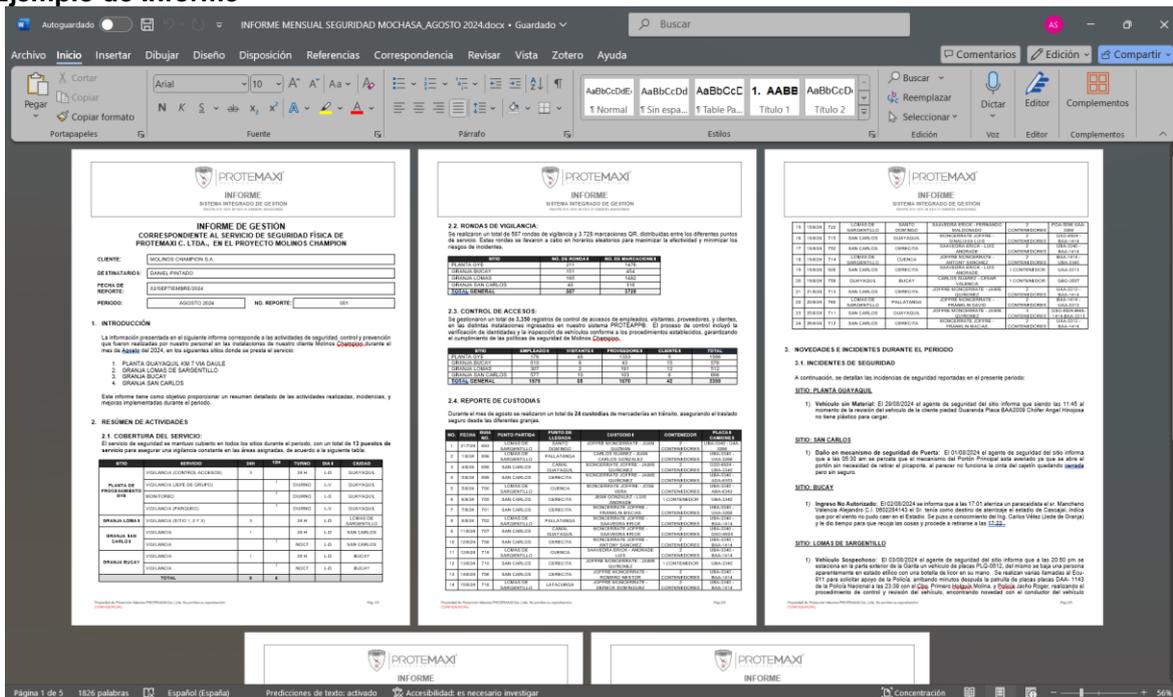
ID	Fecha	H. evento	Id. cliente	Id. localidad	Id. gerencia	Fecha de ocurrencia	TIPO NOVEDAD	tipo incidencia	tipo actividad	tipo personal	detalle	personas involucradas	lugar involucrado	controlada	estado	resultado
128	10/21/2023	0:00	95	CLIENTE 1	29	52					Se entrega llaves de ofi. Sr. Vera a Sr. Soledispe Kelvin	xxx				APROBADO
129	10/21/2023	0:00	95	CLIENTE 1	29	52					vehículo Volkswagen picco GTL 7985 Solo base con autorización previa		0			APROBADO
130	10/21/2023	13:50	95	CLIENTE 1	29	52					SALE MOVIL GOK 159 SR. KELVIN SOLEDISPA CON SR. GALARZA LLEVAN MERCADERIA A CLIENTE : CARAGUAY AUTORIZA VIA TELLAR ING JOSE MARIA VERA. EL LUNES REGULARIA LA SALIDA		0			APROBADO
131	10/21/2023	13:50	95	CLIENTE 1	29	52					ING FEUD ALEX SALE EN MOVIL GTL 7988 CON HOR. 2099. CON DESTINO A TALLERES AUTORIZADO POR ING DURAN JAVIER		0			APROBADO
132	10/22/2023	7:50	95	CLIENTE 1	29	52					SR JAMIE PITA DANIEL -AREA IMPRESORA INGRESA ATRASADO A LABORAR AUTORIZADO POR SR. CHOZE MITTE EDUARDO		0			APROBADO
133	10/22/2023	9:09	95	CLIENTE 1	29	52					SR ORIE GONZALEZ ADRIAN ABEA TELARES INGRESA ATRASADO A LABORAR AUTORIZA SR MACIAS BRAVO JIPSON LODE DE TURNO		0			APROBADO
134	10/22/2023	11:07	95	CLIENTE 1	29	52					SALE EN CAMIONETA SR VICTOR PACHECO I electricos A PLANTA 2 AUTORIZADO RETORNA 18:15		0			APROBADO
135	10/22/2023	8:09	95	CLIENTE 1	29	53					INGRESA SR MARCO RAMIREZ LEÓN 19:50		0			APROBADO
136	10/25/2023	1:39	95	CLIENTE 1	29	53					PACIENTE SR CHRISTOPHER MENDOZA ZAMBRANO EDGAR 22 AÑO:AREA LAMINADORA		0			APROBADO
137	25/10/2023	3:28:00	95	CLIENTE 1	29	53					SR QUIIMI LINDAO JOSUE ADRIAN 9852055127 INGRESA ATRASADO A AREA CORTADORA... AUTORIZA SR LUIS TUMBACAO		0			APROBADO
138	10/29/2023	11:50	95	CLIENTE 1	29	52					SR ERICK YAGUAL (TELARES) DESEABA INGRESAR A LAS 11:50. SR JIPSON MACIAS LE INDICA QUE A ESA HORA YA NO PUEDE INGRESAR		0			APROBADO
139	10/29/2023	11:59	95	CLIENTE 1	29	52					DR JHONNY GOYA INGRESA A DISPENSARIO RETIITA CARPETA QUE UTILIZARA MAÑANA EN PLANTA 2.		0			APROBADO
140	10/29/2023	8:45	96	CLIENTE 1	29	53					PERSONAL ATRASADO		0			APROBADO
141	10/29/2023	8:48	96	CLIENTE 1	29	53					19H41 SR OMAR MERO		0			APROBADO
142	10/29/2023	8:51	96	CLIENTE 1	29	53					20H20 INGRESA SR VICTOR MORALES		0			APROBADO
143	11/1/2023	20:12	95	CLIENTE 1	29	53					SR RAMIREZ TOMALA JOSE (IMPRESORA) AUTORIZA ING BRAZALES		0			APROBADO
144	11/1/2023	19:40	95	CLIENTE 1	29	53					ATRASADO ENTRA SR RAMIREZ TOMALA JOSE I		0			APROBADO

Nota: Archivo Excel que contiene 409 novedades reportadas en los períodos: 2023, 2024 y 2025

Debido a que se requería seguir estrictamente el formato del informe se solicitó que se nos proporcionará un ejemplo de un informe realizado con el fin de cumplir con los estándares y requerimientos de la empresa.

A continuación, se presenta un ejemplo del informe proporcionado:

Figura 12
Ejemplo de Informe



Nota: Archivo Word que contiene informe de incidencias del mes de agosto del 2024

Para obtener los datos que utilizamos en nuestro prototipo, tuvimos que realizar una extracción de datos a partir de la muestra que nos impartió la empresa de seguridad en un archivo Excel con 409 registros. Este archivo fue transformado en formato CSV para realizar el trabajo de carga a la base de datos con la cual se va a implementar en el prototipo.

Creación de proyecto de Laravel

Para la creación del proyecto de Laravel donde va a recaer la codificación para la página web del módulo de reportaría y donde se va a realizar la conexión a la base de datos para la generación de reportes, se utiliza el siguiente comando:

```
composer create-project laravel/laravel proyectoLaravel
```

Figura 13
Comando de creación

```
C:\Windows\System32\cmd.exe
Microsoft Windows [Versión 10.0.22631.4602]
(c) Microsoft Corporation. Todos los derechos reservados.

C:\Users\USER\Desktop\TesisProyecto>composer create-project laravel/laravel proyectoLaravel
Creating a "laravel/laravel" project at "./proyectoLaravel"
Installing laravel/laravel (v11.5.1)
- Downloading laravel/laravel (v11.5.1)
- Installing laravel/laravel (v11.5.1): Extracting archive
Created project in C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel
> @php -r "file_exists('.env') || copy('.env.example', '.env');"
Loading composer repositories with package information
Updating dependencies
Lock file operations: 110 installs, 0 updates, 0 removals
- Locking brick/math (0.12.1)
- Locking carbonphp/carbon-doctrine-types (3.2.0)
- Locking dflydev/dot-access-data (v3.0.3)
- Locking doctrine/inflector (2.0.10)
- Locking doctrine/lexer (3.0.1)
- Locking dragonmantank/cron-expression (v3.4.0)
- Locking egulias/email-validator (4.0.3)
- Locking fakerphp/faker (v1.24.1)
- Locking filp/whoops (2.16.0)
- Locking fruitcake/php-cors (v1.3.0)
- Locking graham-campbell/result-type (v1.1.3)
- Locking guzzlehttp/guzzle (7.9.2)
- Locking guzzlehttp/promises (2.0.4)
- Locking guzzlehttp/psr7 (2.7.0)
- Locking guzzlehttp/uri-template (v1.0.3)
- Locking hamcrest/hamcrest-php (v2.0.1)
- Locking laravel/framework (v11.38.2)
- Locking laravel/pail (v1.2.1)
- Locking laravel/pint (v1.20.0)
- Locking laravel/prompts (v0.3.3)
```

Figura 14
Ejecución

```
C:\Windows\System32\cmd.exe
- Locking laravel/framework (v11.38.2)
- Locking laravel/pail (v1.2.1)
- Locking laravel/pint (v1.20.0)
- Locking laravel/prompts (v0.3.3)
- Locking laravel/sail (v1.40.0)
- Locking laravel/serializable-closure (v2.0.1)
- Locking laravel/tinker (v2.10.0)
- Locking league/commonmark (2.6.1)
- Locking league/config (v1.2.0)
- Locking league/flysystem (3.29.1)
- Locking league/flysystem-local (3.29.0)
- Locking league/mime-type-detection (1.16.0)
- Locking league/uri (7.5.1)
- Locking league/uri-interfaces (7.5.0)
- Locking mockery/mockery (1.6.12)
- Locking monolog/monolog (3.8.1)
- Locking myclabs/deep-copy (1.12.1)
- Locking nesbot/carbon (3.8.4)
- Locking nette/schema (v1.3.2)
- Locking nette/utils (v4.0.5)
- Locking nikitic/php-parser (v5.4.0)
- Locking nunomaduro/collision (v8.5.0)
- Locking nunomaduro/termwind (v2.3.0)
- Locking phar-io/manifest (2.0.4)
- Locking phar-io/version (3.2.1)
- Locking phpoption/phpoption (1.9.3)
- Locking phpunit/php-code-coverage (11.0.8)
- Locking phpunit/php-file-iterator (5.1.0)
- Locking phpunit/php-invoker (5.0.1)
- Locking phpunit/php-text-template (4.0.1)
- Locking phpunit/php-timer (7.0.1)
- Locking phpunit/phpunit (11.5.3)
- Locking psr/clock (1.0.0)
```

Figura 15
Instalación de paquetes y recursos

```
C:\Windows\System32\cmd.exe
- Locking phpunit/phpunit (11.5.3)
- Locking psr/clock (1.0.0)
- Locking psr/container (2.0.2)
- Locking psr/event-dispatcher (1.0.0)
- Locking psr/http-client (1.0.3)
- Locking psr/http-factory (1.1.0)
- Locking psr/http-message (2.0)
- Locking psr/log (3.0.2)
- Locking psr/simple-cache (3.0.0)
- Locking psy/psysh (v0.12.7)
- Locking ralouphie/getallheaders (3.0.3)
- Locking ramsey/collection (2.0.0)
- Locking ramsey/uuid (4.7.6)
- Locking sebastian/cli-parser (3.0.2)
- Locking sebastian/code-unit (3.0.2)
- Locking sebastian/code-unit-reverse-lookup (4.0.1)
- Locking sebastian/comparator (6.3.0)
- Locking sebastian/complaxity (4.0.1)
- Locking sebastian/diff (6.0.2)
- Locking sebastian/environment (7.2.0)
- Locking sebastian/exporter (6.3.0)
- Locking sebastian/global-state (7.0.2)
- Locking sebastian/lines-of-code (3.0.1)
- Locking sebastian/object-enumerator (6.0.1)
- Locking sebastian/object-reflector (4.0.1)
- Locking sebastian/recursion-context (6.0.2)
- Locking sebastian/type (5.1.0)
- Locking sebastian/version (5.0.2)
- Locking staabm/side-effects-detector (1.0.5)
- Locking symfony/clock (v7.2.0)
- Locking symfony/console (v7.2.1)
- Locking symfony/css-selector (v7.2.0)
- Locking symfony/deprecation-contracts (v3.5.1)
```

Figura 16
Creación de archivos y directorios

```
C:\Windows\System32\cmd.exe
- Installing phar-io/version (3.2.1): Extracting archive
- Installing phar-io/manifest (2.0.4): Extracting archive
- Installing myclabs/deep-copy (1.12.1): Extracting archive
- Installing phpunit/phpunit (11.5.3): Extracting archive
66 package suggestions were added by new dependencies, use `composer suggest` to see details.
Generating optimized autoload files
> Illuminate\Foundation\ComposerScripts::postAutoloadDump
> @php artisan package:discover --ansi

 INFO  Discovering packages.

laravel/pail ..... DONE
laravel/sail ..... DONE
laravel/tinker ..... DONE
nesbot/carbon ..... DONE
nunomaduro/collision ..... DONE
nunomaduro/termwind ..... DONE

81 packages you are using are looking for funding.
Use the `composer fund` command to find out more!
> @php artisan vendor:publish --tag=laravel-assets --ansi --force

 INFO  No publishable resources for tag [laravel-assets].

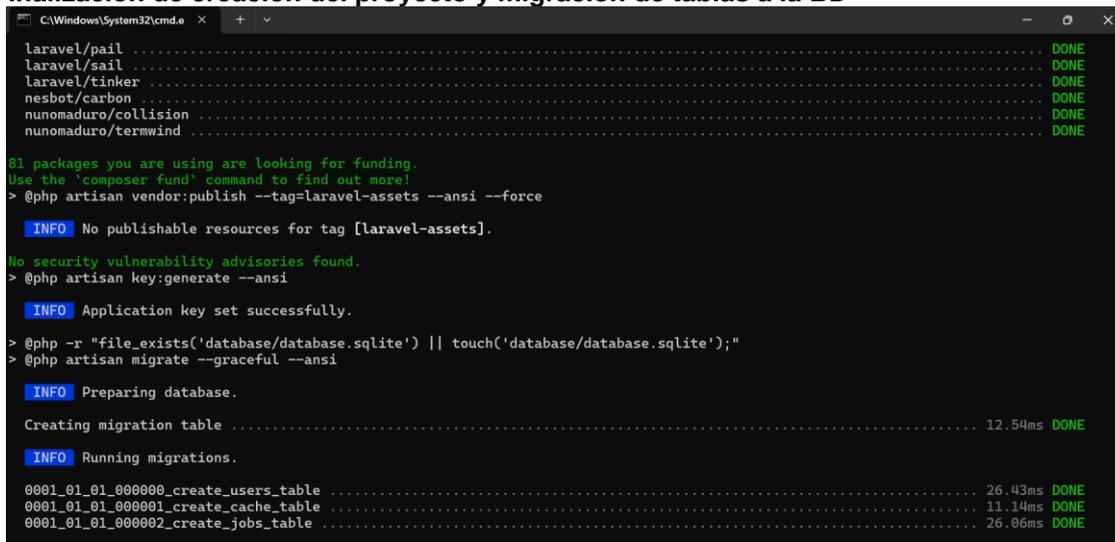
No security vulnerability advisories found.
> @php artisan key:generate --ansi

 INFO  Application key set successfully.

> @php -r "file_exists('database/database.sqlite') || touch('database/database.sqlite');"
> @php artisan migrate --graceful --ansi

 INFO  Preparing database.
```

Figura 17
Finalización de creación del proyecto y migración de tablas a la BD



```
C:\Windows\System32\cmd.exe
laravel/pail ..... DONE
laravel/sail ..... DONE
laravel/tinker ..... DONE
nesbot/carbon ..... DONE
nunomaduro/collision ..... DONE
nunomaduro/termwind ..... DONE

81 packages you are using are looking for funding.
Use the 'composer fund' command to find out more!
> @php artisan vendor:publish --tag=laravel-assets --ansi --force

  INFO No publishable resources for tag [laravel-assets].

No security vulnerability advisories found.
> @php artisan key:generate --ansi

  INFO Application key set successfully.

> @php -r "file_exists('database/database.sqlite') || touch('database/database.sqlite');"
> @php artisan migrate --graceful --ansi

  INFO Preparing database.

Creating migration table ..... 12.54ms DONE

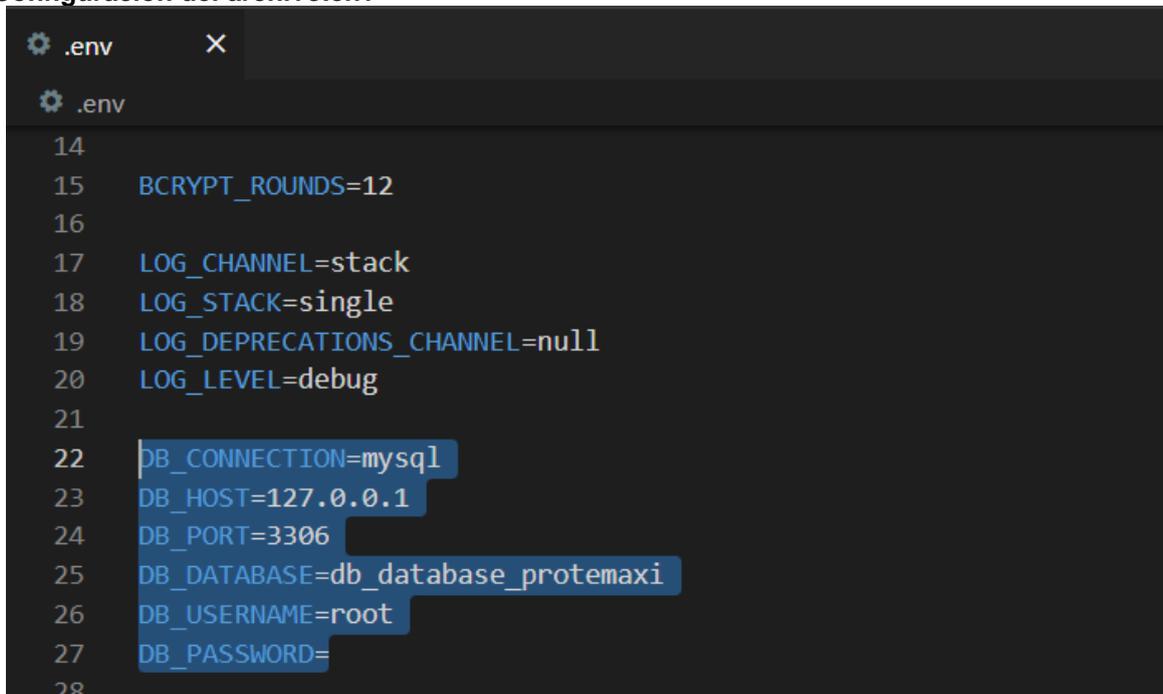
  INFO Running migrations.

0001_01_01_000000_create_users_table ..... 26.43ms DONE
0001_01_01_000001_create_cache_table ..... 11.14ms DONE
0001_01_01_000002_create_jobs_table ..... 26.06ms DONE
```

Configuración del archivo.env

Posterior a la creación del proyecto de Laravel, se requiere de editar la configuración detallada en el **archivo. env** del proyecto. Misma configuración en la que recae las credenciales para el acceso y conexión a la base de datos. Como tal se necesita que se modifique el nombre de la base de datos, el usuario y contraseña.

Figura 18
Configuración del archivo.env



```
.env
14
15 BCRYPT_ROUNDS=12
16
17 LOG_CHANNEL=stack
18 LOG_STACK=single
19 LOG_DEPRECATIONS_CHANNEL=null
20 LOG_LEVEL=debug
21
22 DB_CONNECTION=mysql
23 DB_HOST=127.0.0.1
24 DB_PORT=3306
25 DB_DATABASE=db_database_protemaxi
26 DB_USERNAME=root
27 DB_PASSWORD=
28
```

Nota: Archivo. env que muestra parámetros del tipo de conexión de base de datos ip del host y nombre la base de datos para poner conectar el proyecto de Laravel con MySQL

Creación de Modelos MVC

En Laravel se utiliza la arquitectura de software de MVC (Modelo-Vista-Controlador). Los modelos representan las tablas que se encuentran en la base de datos de una forma que permite una interacción de manera organizada.

Para la creación de Modelos, se utilizó los siguientes comandos:

```
php artisan make: model Detalle -m
```

Figura 19
Creación de modelo Detalle

```
C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel>php artisan make:model Detalle -m
INFO Model [C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel\app\Models\Detalle.php] created successfully.
INFO Migration [C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel\database\Migrations\2025_01_15_055352_create_detalles_table.php] created successfully.
```

```
php artisan make:model Recomendacion -m
```

Figura 20
Creación de modelo Recomendación

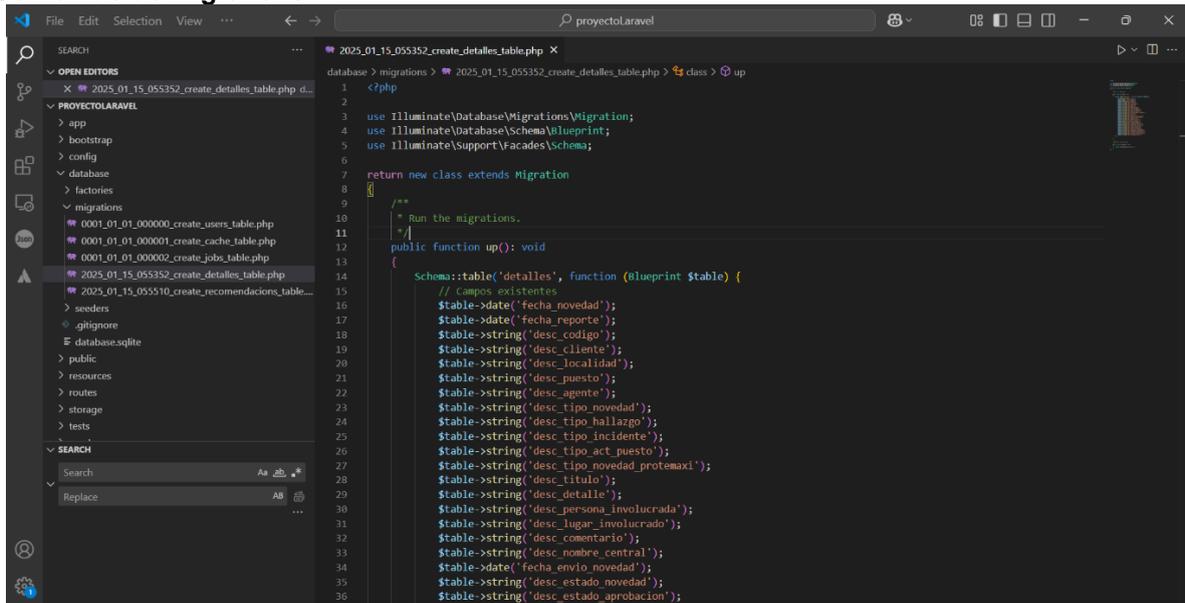
```
C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel>php artisan make:model Recomendacion -m
INFO Model [C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel\app\Models\Recomendacion.php] created successfully.
INFO Migration [C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel\database\Migrations\2025_01_15_055510_create_recomendacions_table.php] created successfully.
```

Migrations

El migrations en Laravel constituye de una forma de administrar el esquema de la base de datos conectada al proyecto desde Laravel mismo.

Permite crear, realizar modificaciones e incluso eliminaciones cuando es necesario todo desde el proyecto mismo con funciones en la carpeta de migrations.

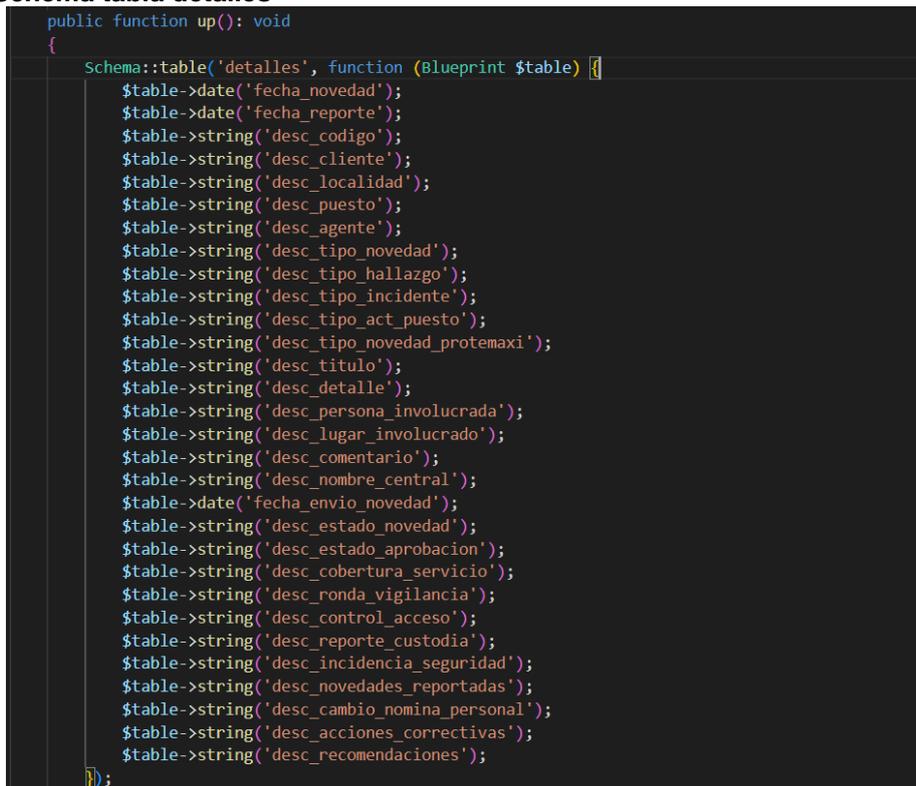
Figura 21
Contenido de Migrations



The screenshot shows a code editor with a file explorer on the left and a code editor on the right. The file explorer shows a project structure with folders like 'app', 'bootstrap', 'config', 'database', 'factories', 'migrations', 'seeds', 'storage', and 'tests'. The 'migrations' folder is expanded, showing several migration files. The selected file is '2025_01_15_055352_create_detalles_table.php'. The code editor shows the following PHP code:

```
1 <?php
2
3 use Illuminate\Database\Migrations\Migration;
4 use Illuminate\Database\Schema\Blueprint;
5 use Illuminate\Support\Facades\Schema;
6
7 return new class extends Migration
8 {
9     /**
10      * Run the migrations.
11      */
12     public function up(): void
13     {
14         Schema::table('detalles', function (Blueprint $table) {
15             // Campos existentes
16             $table->date('fecha_novedad');
17             $table->date('fecha_reporte');
18             $table->string('desc_codigo');
19             $table->string('desc_cliente');
20             $table->string('desc_localidad');
21             $table->string('desc_puesto');
22             $table->string('desc_agente');
23             $table->string('desc_tipo_novedad');
24             $table->string('desc_tipo_hallazgo');
25             $table->string('desc_tipo_incidente');
26             $table->string('desc_tipo_act_puesto');
27             $table->string('desc_tipo_novedad_protemaxi');
28             $table->string('desc_titulo');
29             $table->string('desc_detalle');
30             $table->string('desc_persona_involucrada');
31             $table->string('desc_lugar_involucrado');
32             $table->string('desc_comentario');
33             $table->string('desc_nombre_central');
34             $table->date('fecha_envio_novedad');
35             $table->string('desc_estado_novedad');
36             $table->string('desc_estado_aprobacion');
```

Figura 22
Schema tabla detalles



The screenshot shows a close-up of the migration code for creating the 'detalles' table. The code is as follows:

```
public function up(): void
{
    Schema::table('detalles', function (Blueprint $table) {
        $table->date('fecha_novedad');
        $table->date('fecha_reporte');
        $table->string('desc_codigo');
        $table->string('desc_cliente');
        $table->string('desc_localidad');
        $table->string('desc_puesto');
        $table->string('desc_agente');
        $table->string('desc_tipo_novedad');
        $table->string('desc_tipo_hallazgo');
        $table->string('desc_tipo_incidente');
        $table->string('desc_tipo_act_puesto');
        $table->string('desc_tipo_novedad_protemaxi');
        $table->string('desc_titulo');
        $table->string('desc_detalle');
        $table->string('desc_persona_involucrada');
        $table->string('desc_lugar_involucrado');
        $table->string('desc_comentario');
        $table->string('desc_nombre_central');
        $table->date('fecha_envio_novedad');
        $table->string('desc_estado_novedad');
        $table->string('desc_estado_aprobacion');
        $table->string('desc_cobertura_servicio');
        $table->string('desc_ronda_vigilancia');
        $table->string('desc_control_acceso');
        $table->string('desc_reporte_custodia');
        $table->string('desc_incidencia_seguridad');
        $table->string('desc_novedades_reportadas');
        $table->string('desc_cambio_nomina_personal');
        $table->string('desc_acciones_correctivas');
        $table->string('desc_recomendaciones');
    });
}
```

Figura 23
Schema tabla recomendaciones

```
1  <?php
2
3  use Illuminate\Database\Migrations\Migration;
4  use Illuminate\Database\Schema\Blueprint;
5  use Illuminate\Support\Facades\Schema;
6
7  return new class extends Migration
8  {
9      /**
10     * Run the migrations.
11     */
12     public function up(): void
13     {
14         Schema::create('recomendacions', function (Blueprint $table) {
15             $table->date('id');
16             $table->date('cliente');
17             $table->string('titulo');
18             $table->string('frecuencia');
19             $table->string('recomendacion');
20         });
21     }
22
23     /**
24     * Reverse the migrations.
25     */
26     public function down(): void
27     {
28         Schema::dropIfExists('recomendacions');
29     }
30 };
```

Estos esquemas de mis tablas definidas pasan a escribirse en la base de datos conectada, que previamente fue establecida en el archivo .env, por medio del siguiente comando:

php artisan migrate

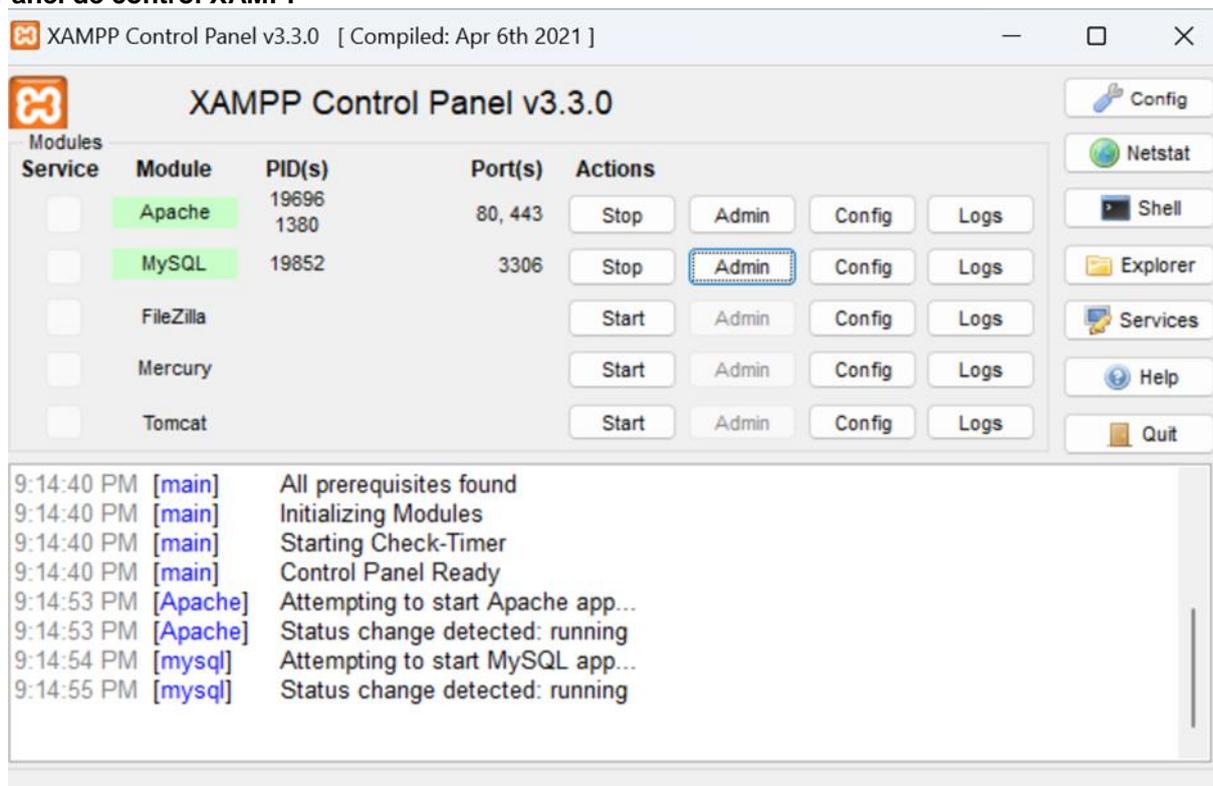
Figura 24
Ejecución migrate

```
PS C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel> php artisan migrate
INFO Running migrations.
2025_01_15_055352_create_detalle_table ..... 28.92ms DONE
2025_01_15_055510_create_recomendacions_table ..... 22.38ms DONE
```

Uso del XAMP

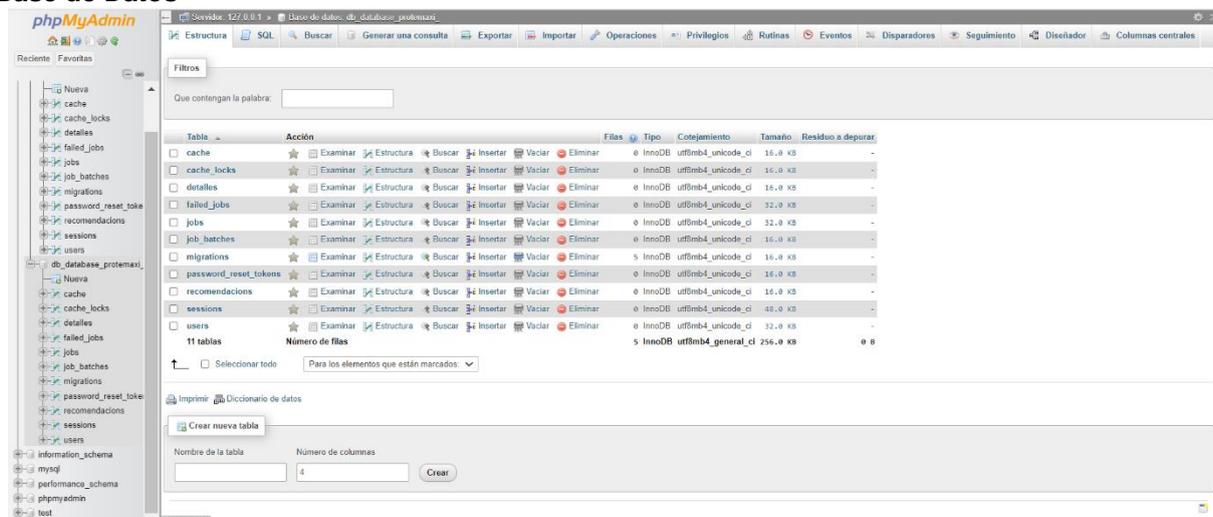
Para poder acceder a nuestra base MySQL de nuestro local, debemos tener activado el módulo de Apache y de MySQL en nuestro panel de control de XAMP.

Figura 25
Panel de control XAMPP



De esta manera si ingresamos a administrar nuestra base de datos Mysql podemos visualizar la estructura que hemos creado para el proyecto. Sin tener activado el módulo del panel de control de XAMP, no podríamos acceder a la base de datos.

Figura 26
Base de Datos

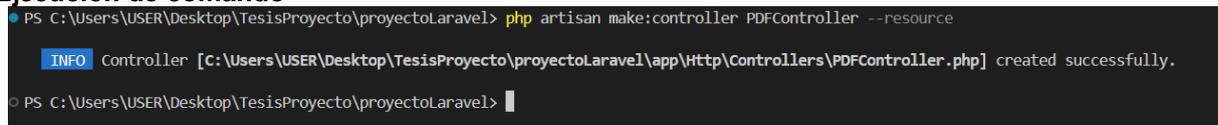


Creación de Controlador

En el contexto de Laravel, los controladores son los que contienen toda la lógica de negocio correspondiente al proyecto y se lo conoce como la capa intermedia entre la interfaz gráfica y los modelos. Generamos un controlador por medio del siguiente comando:

```
php artisan make:controller PDFController --resource
```

Figura 27
Ejecución de comando

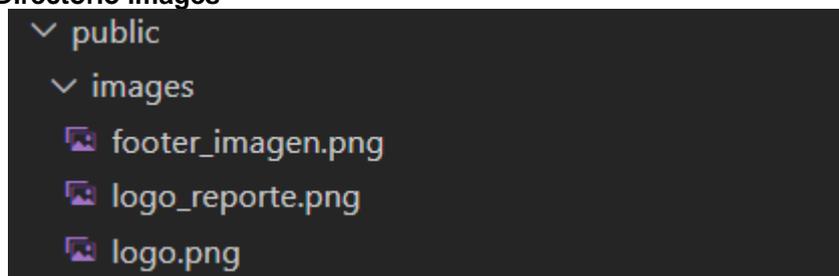


```
PS C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel> php artisan make:controller PDFController --resource
INFO Controller [C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel\app\Http\Controllers\PDFController.php] created successfully.
PS C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel>
```

Carpeta de Imágenes

De manera manual se creó la carpeta “imagenes” en el directorio cuya finalidad es de almacenar las imágenes que se van a utilizar a lo largo del proyecto para que recaigan en un repositorio centralizado.

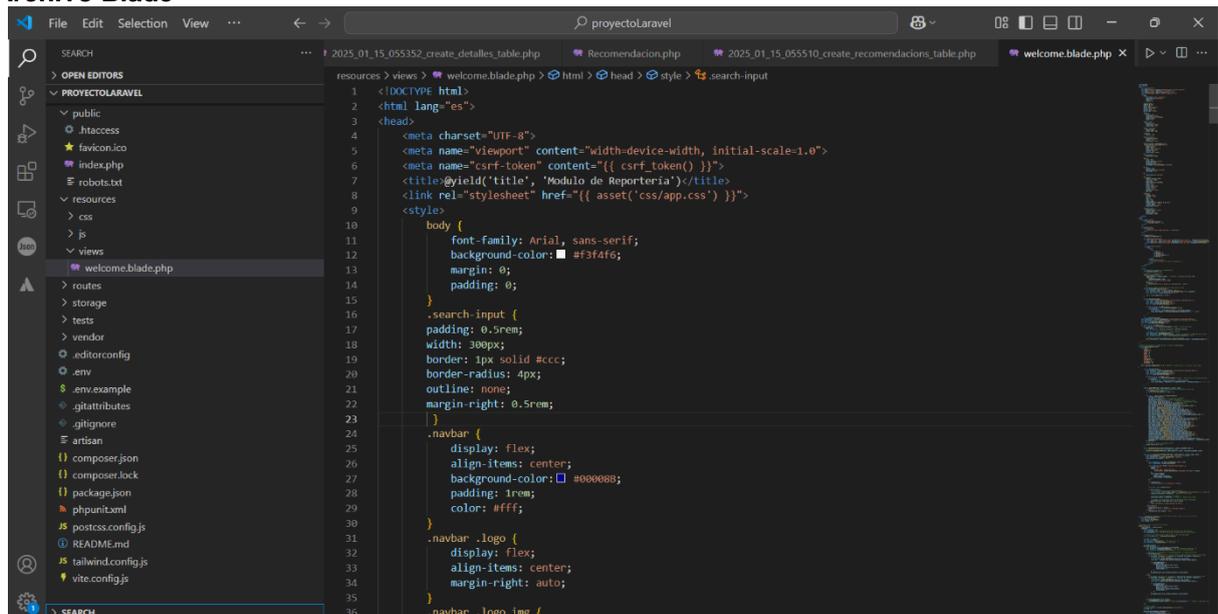
Figura 28
Directorio images



Vista con Blade

Utilizamos Blade con la finalidad de generar una vista, la interfaz gráfica con la cual el usuario interactúa directamente.

Figura 29
Archivo Blade

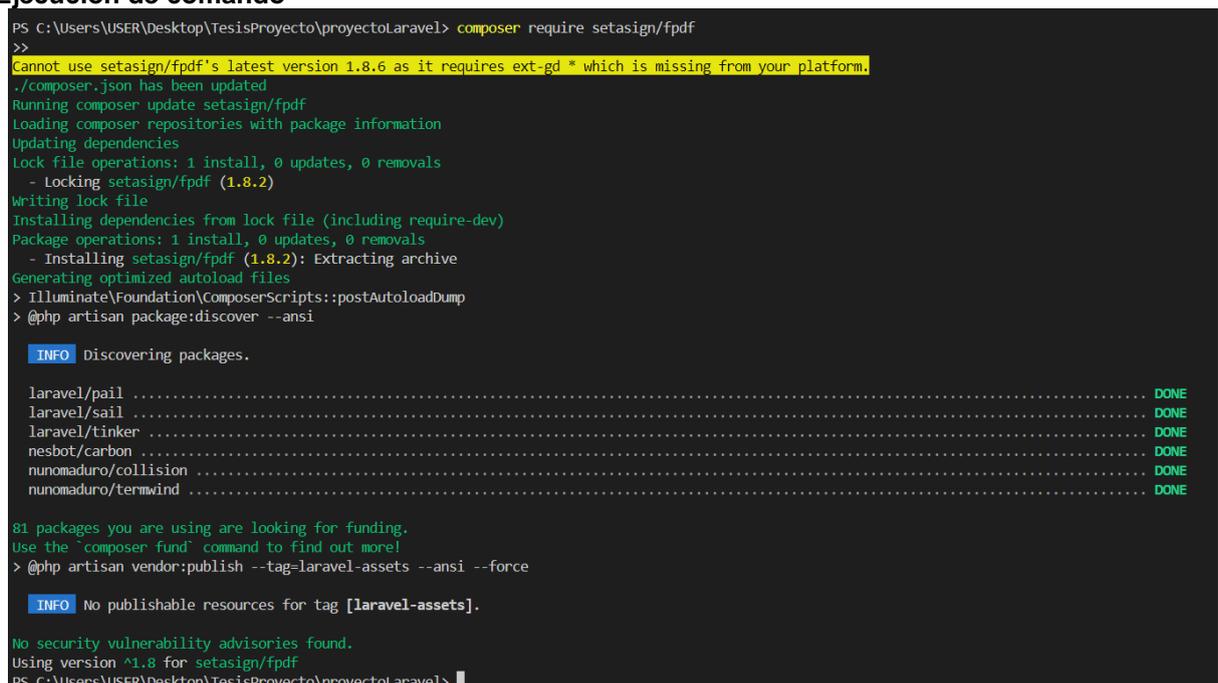


Instalación de la librería FPDF

Para la construcción de nuestro reporte dinámico implementamos la librería FPDF y para ello, primero debemos realizar la instalación de esta mediante el siguiente comando:

```
composer require setasign/fpdf
```

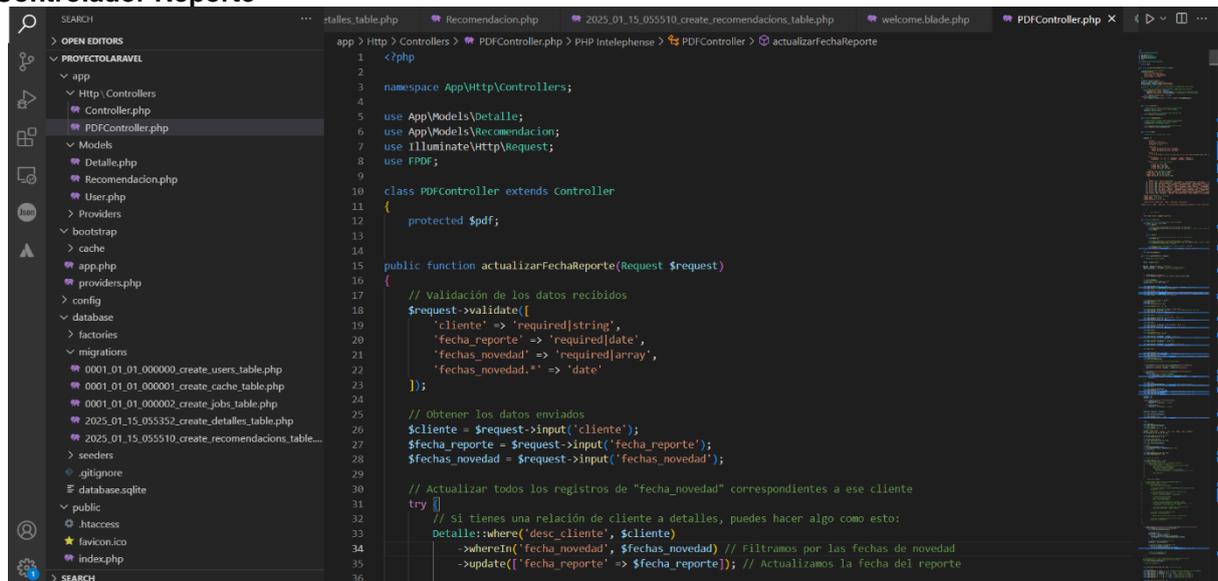
Figura 30
Ejecución de comando



Construcción del Reporte en el Controlador

Para construir el reporte mediante código lo realizamos desde el controlador que además es donde se realizan las peticiones http, recaen funciones definidas y la codificación del reporte.

Figura 31
Controlador Reporte



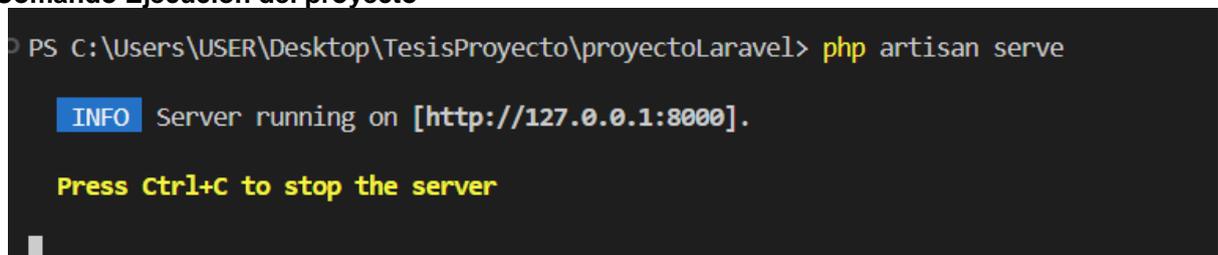
```
1 <?php
2
3 namespace App\Http\Controllers;
4
5 use App\Models\Detalle;
6 use App\Models\Recomendacion;
7 use Illuminate\Http\Request;
8 use FPDF;
9
10 class PDFController extends Controller
11 {
12     protected $pdf;
13
14
15     public function actualizarFechaReporte(Request $request)
16     {
17         // Validación de los datos recibidos
18         $request->validate([
19             'cliente' => 'required|string',
20             'fecha_reporte' => 'required|date',
21             'fechas_novedad' => 'required|array',
22             'fechas_novedad.*' => 'date'
23         ]);
24
25         // Obtener los datos enviados
26         $cliente = $request->input('cliente');
27         $fecha_reporte = $request->input('fecha_reporte');
28         $fechas_novedad = $request->input('fechas_novedad');
29
30         // Actualizar todos los registros de "fecha_novedad" correspondientes a ese cliente
31         try {
32             // Si tienes una relación de cliente a detalles, puedes hacer algo como esto:
33             Detalle::where('desc_cliente', $cliente)
34                 ->whereIn('fecha_novedad', $fechas_novedad) // Filtramos por las fechas de novedad
35                 ->update(['fecha_reporte' => $fecha_reporte]); // Actualizamos la fecha del reporte
36         } catch (Exception $e) {
37             // Manejar el error
38         }
39     }
40 }
```

Ejecución del proyecto

Para ejecutar nuestro proyecto de Laravel y así visualizar la vsita que hemos creado con Blade corremos el siguiente comando en la consola:

```
php artisan serve
```

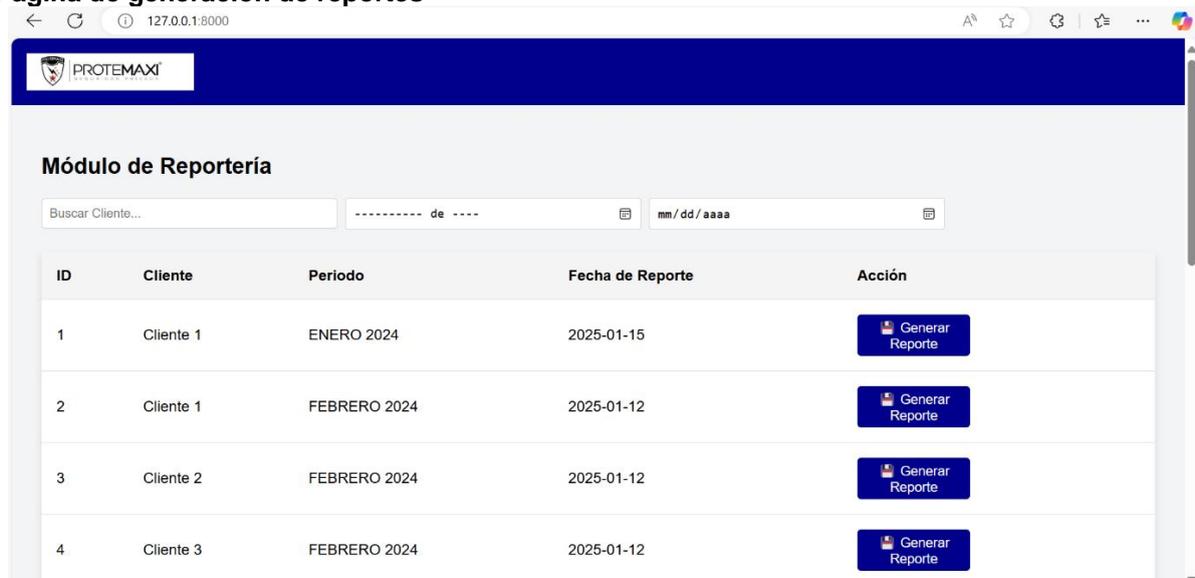
Figura 32
Comando Ejecución del proyecto



```
PS C:\Users\USER\Desktop\TesisProyecto\proyectoLaravel> php artisan serve
INFO Server running on [http://127.0.0.1:8000].
Press Ctrl+C to stop the server
```

Nos dirigimos a la URL que muestra la consola donde se está ejecutando nuestro proyecto para poder visualizarlo.

Figura 33
Página de generación de reportes



En el módulo de reportería, se estableció que la tabla este agrupada a nivel de granularidad de cliente y periodo; también se consideró que cada vez que se genera un reporte nuevo refleje esa modificación en la columna fecha de reporte de manera automática.

Al hacer clic en el botón “*Generar Reporte*” se descarga el reporte correspondiente al registro seleccionado. Se agregaron filtros de cliente, periodo y fecha reporte respectivamente a la página web con finalidad de que sea amigable para el usuario.

Desarrollo del modelo de Machine Learning

Comparación y elección del tipo de modelo

Previo al desarrollo del modelo de Machine Learning es importante decidir qué modelo se ajusta a los requerimientos y necesidades de la empresa, esto dependerá en gran medida de lo que se deseó obtener. Se necesita poder encontrar patrones que evalúen una frecuencia que después pueda ser clasificada en diferente nivel de incidencia, esto tomando en consideración la precisión con la que debe contar dicho modelo. Lo que se pretende aquí es analizar los resultados de cada modelo probado, exponer sus ventajas y desventajas.

Aquí se busca encontrar patrones que identifiquen frecuencias y utilizar estos

patrones para poder predecir futuras incidencias en base a la frecuencia con que suceden los diferentes tipos de incidentes de seguridad y poder generar recomendaciones por cada incidente.

Se han seleccionado 3 modelos para su comparativa y posterior elección:

- Random Forest
- Support vector machine
- Gradient Boosting Regressor

Cada uno de estos modelos ofrecen ventajas que pueden ser útiles al momento de tratar datos de incidencias de seguridad:

Se escogió Random Forest porque este modelo puede manejar datos con muchas características o diferente tipo de formato, es útil si se desea buscar predicciones en base al nivel de importancia de cada variable con la que se esté trabajando. Su enfoque de promediar múltiples árboles podría limitarlo al hacerlo menos preciso que otras opciones.

Se escogió Support Vector Machine debido a su capacidad para trabajar con diferentes variables que contengan múltiples características dentro de un mismo dataset para trabajar tanto clasificación como regresión esto por si en algún futuro se deseara cambiar la naturaleza del propósito del modelo, esto lo vuelve un algoritmo especial al poder identificar patrones en diferentes escenarios, sin embargo, en problema donde se manejen datos tabulares puede ver limitado. Se debe tener en cuenta que no es un modelo que trabaje de manera perfecta con grandes volúmenes de datos o que puedan incrementarse en un futuro.

El modelo de Gradient Boosting Regressor se escogió debido a su peculiar técnica de Ensemble Learning que permite poder trabajar con diferentes modelos en un mismo "Modelo padre" cuyos hijos son modelos básicos pero que juntos pueden lograr mejorar la precisión de manera progresiva logrando corregir error tras error en cada iteración. Este modelo es útil cuando se busca tener excelente precisión y buen nivel de clasificación en los distintos niveles de incidencias. Es un modelo que puede seguirse optimizando sin necesidad de realizar ajustes tan complejos en sus

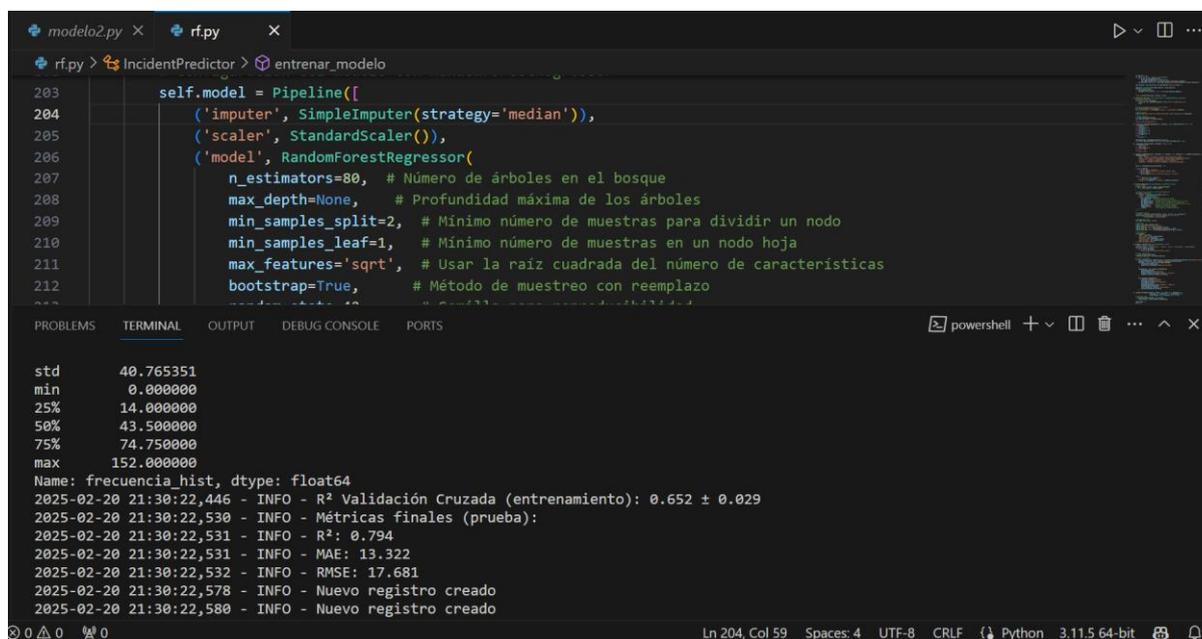
hiperparámetros, se puede añadir una optimización de funciones de pérdida usando `huber` para tener un mejor manejo de outliers.

Para cada modelo se mantienen las mismas variables de entrada con el fin de no afectar la salida esperada, por otro lado, las métricas de cada modelo se tratan de mantener similitudes para que no existan descompensaciones al momento de realizar las diferentes ejecuciones. La data que se utilizará en cada modelo será la misma con el fin de mantener una igualdad y ver los diferentes resultados bajo los mismos tipos de datos.

A continuación, se muestran los parámetros y las métricas resultantes de cada uno de los 3 modelos que han sido objeto de prueba con el fin de escoger uno, la selección final del modelo se hizo de acuerdo con los resultados arrojados y a su aplicabilidad dentro del prototipo del sistema web de generación de informes de incidencias:

Métricas y parámetros del modelo Random Forest

Figura 34
Random Forest Regressor



```
203 self.model = Pipeline([
204     ('imputer', SimpleImputer(strategy='median')),
205     ('scaler', StandardScaler()),
206     ('model', RandomForestRegressor(
207         n_estimators=80, # Número de árboles en el bosque
208         max_depth=None, # Profundidad máxima de los árboles
209         min_samples_split=2, # Mínimo número de muestras para dividir un nodo
210         min_samples_leaf=1, # Mínimo número de muestras en un nodo hoja
211         max_features='sqrt', # Usar la raíz cuadrada del número de características
212         bootstrap=True, # Método de muestreo con reemplazo
213     ))
214 ])
```

std 40.765351
min 0.000000
25% 14.000000
50% 43.500000
75% 74.750000
max 152.000000

Name: frecuencia_hist, dtype: float64
2025-02-20 21:30:22,446 - INFO - R² Validación Cruzada (entrenamiento): 0.652 ± 0.029
2025-02-20 21:30:22,530 - INFO - Métricas finales (prueba):
2025-02-20 21:30:22,531 - INFO - R²: 0.794
2025-02-20 21:30:22,531 - INFO - MAE: 13.322
2025-02-20 21:30:22,532 - INFO - RMSE: 17.681
2025-02-20 21:30:22,578 - INFO - Nuevo registro creado
2025-02-20 21:30:22,580 - INFO - Nuevo registro creado

Para el proceso de análisis y procesamiento de datos se utiliza validación cruzada por lo que se divide de manera parcial los datos de entrenamiento, esto con el fin de generar mejores métricas dentro del modelo al hacerlo más robusto. Con

respecto a al cálculo R2 dentro del grupo de entrenamiento se obtiene un 65.2% lo que significa que aproximadamente más del 50% de la variabilidad de los datos de entrenamiento puede ser explicados por el modelo; con una desviación estándar de 0.029 lo que representa el rendimiento del modelo en las diferentes partes de los datos de entrenamiento, el valor obtenido indica que no existe una variación grande por lo tanto es estable.

Con respecto al conjunto de datos de prueba tenemos un 79.4% en R2 que pertenece a la variabilidad de la variable objetivo, al ser este valor mayor que el valor de los datos de entrenamiento podemos decir que el modelo se comportará mejor ante datos de prueba que con los datos de entrenamiento en validación cruzada.

El error absoluto medio (MAE) es la diferencia entre los valores reales con los valores predichos, en este caso tiene un valor de 13.322 unidades, dentro del modelo este margen de error no considera signos de error y en vez de eso mide la magnitud promedio del error que puede haber sin penalizar valores que son externos. Se tiene un valor de 17.681 de la raíz del error cuadrático medio es decir el RMSE este tiene relación con el MAE debido a que es una métrica que influye en relación con MAE esto quiere decir que si el MAE es alto el RMSE será más alto debido a que penaliza más los errores grandes, esto puede indicar que los errores tienen cierta dispersión lo que significa que existen ciertas diferencias en los casos de predicciones junto con los valores reales

Métricas y parámetros del modelo SVM

Figura 35
Support Vector Machine

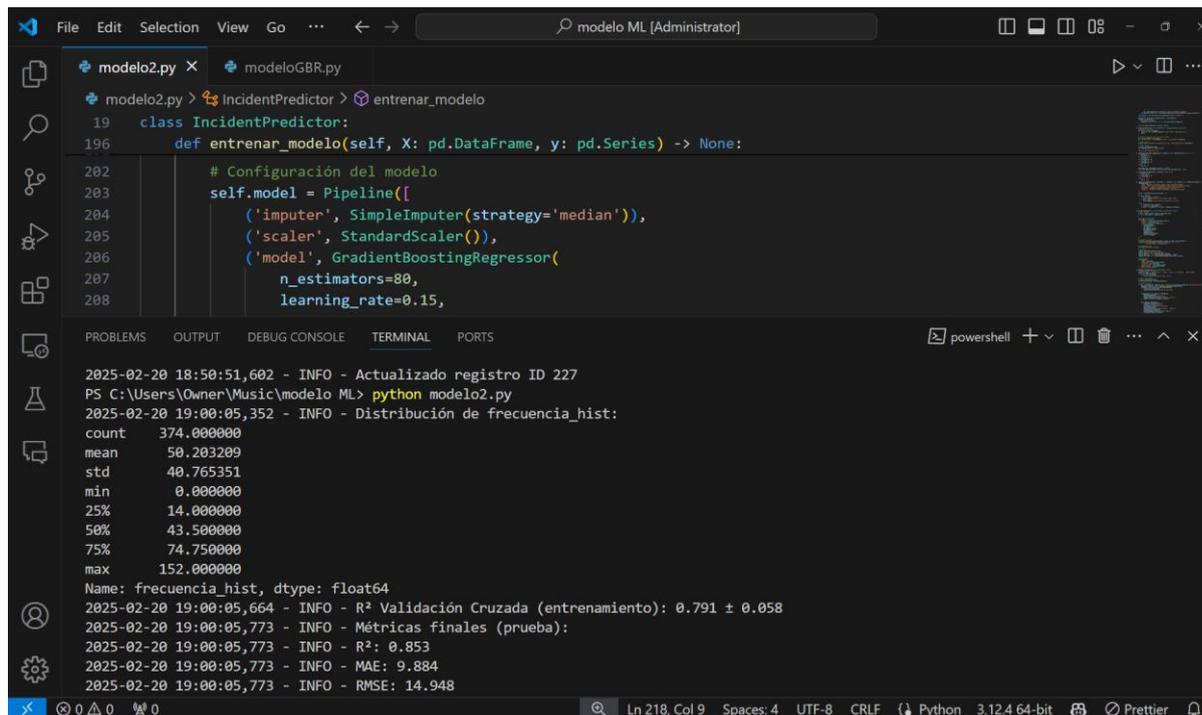
```
class IncidentPredictor:
    def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None:
        # Configuración inicial del modelo con SVR
        self.model = Pipeline([
            ('imputer', SimpleImputer(strategy='median')),
            ('scaler', StandardScaler()),
            ('model', SVR(kernel='rbf', C=1.0, epsilon=0.1, gamma='scale'))
        ])
        # Definir los hiperparámetros que se probarán en la búsqueda
        param_grid = {
```

```
50%      43.500000
75%      74.750000
max      152.000000
Name: frecuencia_hist, dtype: float64
2025-02-24 22:00:05,321 - INFO - Mejores parámetros: {'model__C': 10, 'model__epsilon': 0.01, 'model__gamma': 'auto'}
2025-02-24 22:00:05,438 - INFO - R² Validación Cruzada (entrenamiento): 0.237 ± 0.098
2025-02-24 22:00:05,465 - INFO - Métricas finales (prueba):
2025-02-24 22:00:05,466 - INFO - R²: 0.370
2025-02-24 22:00:05,469 - INFO - MAE: 22.996
2025-02-24 22:00:05,471 - INFO - RMSE: 30.908
```

El modelo Support Vector Machine optimizado con los parámetros $C = 10$, $\epsilon = 0.01$ y $\gamma = 'auto'$ obtuvo un R^2 de 0.237 ± 0.098 en la validación cruzada con los datos de entrenamiento, lo que indica que el modelo logra explicar solo un 23.7% de la variabilidad en los datos de entrenamiento, con una variabilidad moderada entre los pliegues. En los datos de prueba, el modelo alcanzó un R^2 de 0.370, lo que sugiere que la capacidad de generalización del modelo es limitada, ya que solo explica el 37% de la variabilidad de los datos de prueba. Además, el MAE de 22.996 y el RMSE de 30.908 indican un margen de error relativamente grande en las predicciones, siendo el error absoluto y la dispersión más altos en comparación con otros modelos, lo que sugiere que el modelo podría no ser tan preciso en sus predicciones. **Se realiza una explicación del porque las métricas del modelo SVM son muy inferiores a comparación de los otros modelos debajo del cuadro comparativo de los resultados de los 3 modelo.**

Métricas y parámetros del modelo Gradient Boosting Regressor

Figura 36
Gradient Boosting Regressor



```
class IncidentPredictor:
    def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None:
        # Configuración del modelo
        self.model = Pipeline([
            ('imputer', SimpleImputer(strategy='median')),
            ('scaler', StandardScaler()),
            ('model', GradientBoostingRegressor(
                n_estimators=80,
                learning_rate=0.15,
```

```
2025-02-20 18:50:51,602 - INFO - Actualizado registro ID 227
PS C:\Users\Owner\Music\modelo ML> python modelo2.py
2025-02-20 19:00:05,352 - INFO - Distribución de frecuencia_hist:
count    374.000000
mean     50.203209
std      40.765351
min       0.000000
25%      14.000000
50%      43.500000
75%      74.750000
max      152.000000
Name: frecuencia_hist, dtype: float64
2025-02-20 19:00:05,664 - INFO - R² Validación Cruzada (entrenamiento): 0.791 ± 0.058
2025-02-20 19:00:05,773 - INFO - Métricas finales (prueba):
2025-02-20 19:00:05,773 - INFO - R²: 0.853
2025-02-20 19:00:05,773 - INFO - MAE: 9.884
2025-02-20 19:00:05,773 - INFO - RMSE: 14.948
```

Los resultados del modelo de Gradient Boosting Regressor al igual que los modelos posteriores se dividen en datos de prueba y datos de entrenamiento donde a los datos de entrenamiento se les aplica validación cruzada, aquí observamos que el R2 para el grupo de entrenamiento es del 79.1% lo que significa que el modelo logra explicar más del 79% de la diferenciación de los datos de entrenamiento, mientras que el R2 de prueba es del 85.3% lo que quiere decir que el modelo tiene una utilidad superior al de los modelos mencionados anteriormente, al ser mayor se define que el modelo captura bien los patrones del conjunto de prueba además de no estar sobre ajustado.

Por otro lado, el MAE de 9.884 lo que indica que esa es la diferencia de error absoluta entre los valores reales y las predicciones realizadas, este valor es más pequeño que el del modelo de Random Forest lo que indica que puede sugerir predicciones más precisas de manera absoluta. El RMSE es de 14.9 lo que indica que existe una cierta dispersión en la muestra de errores, aunque este valor es menor al que se ha venido calculando en los modelos de Random Forest y Support Vector Machine.

Una vez obtenido los resultados de cada uno de los modelos se comparan las

métricas observadas, ventajas, desventajas y características con relación al prototipo desarrollado:

Tabla 6
Cuadro comparativo

Modelo	Gradient Boosting Regressor	Support Vector Machine	Random Forest Regressor
Métricas Observadas	MAE: 9.88 RMSE: 14.94 R2 Prueba: 0.82 R2 Entrenamiento: 0.791	MAE: 22.99 RMSE: 30.90 R2 Prueba: 0.37 R2 Entrenamiento: 0.23	MAE: 13.32 RMSE: 17.68 R2 Prueba: 0.79 R2 Entrenamiento: 0.65
Ventajas	<ul style="list-style-type: none"> - Mejor rendimiento en las métricas - Manejo natural de features no lineales - Robusto contra outliers - Bueno para datasets medianos y grandes - Menos propenso a sobreajuste - Alto rendimiento predictivo - Capacidad para manejar datos faltantes 	<ul style="list-style-type: none"> - Análisis de múltiples tipos de datos en un mismo dataset. - Eficiente en dataset pequeños o medianos - Uso en casos de clasificación o regresión 	<ul style="list-style-type: none"> - Fácil de interpretar - No requiere escalado de datos - Menos propenso al overfitting
Desventajas	<ul style="list-style-type: none"> - Puede ser computacionalmente intensivo 	<ul style="list-style-type: none"> - Costo computacional alto - Escalabilidad limitada a grandes volúmenes de datos 	<ul style="list-style-type: none"> - Costo computacional alto - Puede requerir más memoria

Nota: Esta es una tabla comparativa cuyo fin es contrastar los resultados obtenidos de las métricas de cada modelo más sus ventajas y desventajas dentro de la práctica.

¿Por qué la implementación de SVM en comparación con las métricas de los dos modelos ha tenido un bajo rendimiento o unas métricas bajas?

A pesar de que SVM es un modelo que tiene utilidad en múltiples tipos de datos y en casos de uso de clasificación y regresión este puede verse afectado debido a la forma en como se presentan los datos y el problema que se plantea al modelo. SVM es un modelo que está diseñado para trabajar en problemas con datos que tienen una estructura clara es decir datos que sean linealmente separables es decir datos que puedan ser separados en clases, en este caso si bien se usa también una clasificación y predicción, los datos son tabularmente complejos teniendo una mezcla de

características numéricas, categóricas y de texto. Cuando yo aplico el modelo SVM en este contexto el modelo dice que cualquier característica de la data debe ser numérica y escalable esto quiere decir que se debe realizar un preprocesamiento a parte del que se realiza con el fin de codificar variables categóricas y de texto; esto genera un problema ya que en el caso del texto se debe reducir su dimensionalidad. El problema radica en que cuando se cree la matriz de características esta será muy dispersa y grande esto causa que no sea escalable. Cabe mencionar que este modelo no analiza o haya interacción de manera automática entre las características de los datos.

¿Qué ventajas tiene el modelo Gradient Boosting Regressor con respecto a los demás modelos?

- **Mejores métricas:** El modelo de Gradient Boosting Regressor logra mejores métricas en R2, MAE y RMSE debido al enfoque que tiene para un manejo de corrección de errores iterativo.
- **Flexibilidad:** Aunque es cierto que la complejidad del modelo puede provocar sobreajuste, en este se pueden modificar los hiperparámetros como puede ser el `max_depth`, `n_estimators` o el `Learning_rate` con el fin de evitar este problema.
- **Manejo de outliers:** Al poder integrar funciones como `huber` que son de pérdida este logra pasar o hacer frente a valores atípicos es decir logra ser menos sensible.
- **Interoperabilidad:** Aunque no llega a ser mejor que Random Forest, este modelo puede analizar importancia de las características, esto ayuda a entender que variables ayudan a generar mejores predicciones dentro del modelo.
- **Manejo de datos:** Capaz de poder manejar el dataset proporcionado que trae valores ausentes sin necesidad de realizar imputaciones difíciles.

Learning Seleccionado: Gradient Boosting Regressor

Se realiza una comparación manual de las métricas de cada uno de los modelos

mencionando cual es el mas optimo dependiente del resultado obtenido, en este caso se evalúa cada métrica.

1. Análisis de las métricas:

- **MAE (Error Absoluto Medio):** Gradient Boosting Regressor tiene el MAE más bajo (2.50), lo que indica que sus predicciones son, en promedio, las más cercanas a los valores reales.
- **RMSE (Error Cuadrático Medio):** Similar al MAE, Gradient Boosting Regressor también presenta el RMSE más bajo (3.92), sugiriendo una menor dispersión en los errores de predicción.
- **R2 (Coeficiente de Determinación) en Prueba:** Gradient Boosting Regressor lidera con un R2 de 0.82 en el conjunto de prueba. Esto significa que el modelo explica el 82% de la varianza en los datos observados durante la prueba, lo que demuestra una alta capacidad predictiva en datos no vistos.

2. Balance de características:

- Ofrece un buen equilibrio entre precisión y complejidad
- Más preciso que Random Forest en este caso específico

3. Adecuación al problema:

- Parece manejar mejor la naturaleza del problema
- El alto R2 sugiere que captura bien las relaciones en los datos
- Bueno para problemas de regresión con complejidad moderada

En contraste con los demás modelos la esta justificación se basa netamente en características específicas que tiene el modelo en comparación con los otros, siendo mejor que el SVM al estar diseñado de manera específica para manejar datos tabulares donde se manejan características numéricas, texto y categóricas de manera nativa donde se encuentran relaciones no lineales mientras que el SVM no. Sobrepasando al modelo de Random Forest gracias a su habilidad para crear modelo mas precisos y robustos debido a su magnífica habilidad para iterar errores residuales optimizando sus resultados. Random Forest solo entrenaría arboles de manera dependiente y saca un promedio Mientras que Gradient Boosting logra entrenar múltiples arboles de manera seguida y cada árbol va corrigiendo los errores del árbol anterior.

La decisión de utilizar Gradient Boosting Regressor no solo se basa en su superior rendimiento métrico, sino también en consideraciones prácticas de

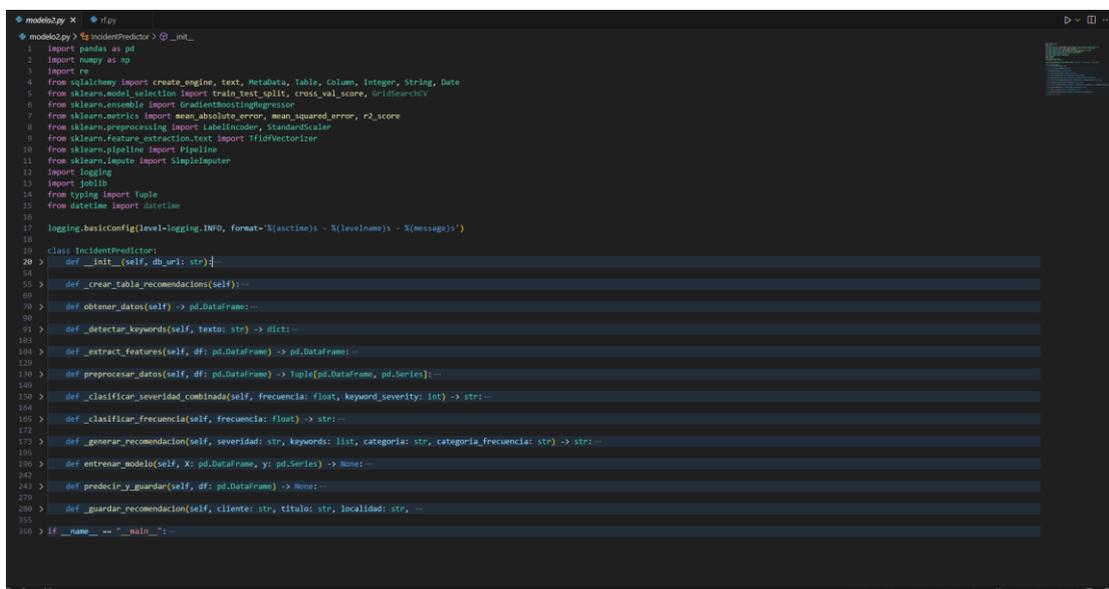
implementación, mantenimiento y valor comercial. El modelo demuestra una capacidad sobresaliente para capturar la complejidad inherente de los datos mientras mantiene un nivel de interpretabilidad y manejabilidad que lo hace ideal para aplicaciones del mundo real. Su rendimiento consistentemente superior en todas las métricas evaluadas, combinado con sus características prácticas favorables, lo posiciona como la mejor elección para este caso de uso específico.

Estructura y explicación del modelo

El modelo seleccionado para realizar la integración con el prototipo de Machine Learning ha sido desarrollado y mejorado debido a que en un principio las métricas no eran las óptimas, sin embargo, esto fue motivo que se siguiera un desarrollo continuo con el fin de mejorar las métricas del modelo tanto para los datos de prueba como para los de entrenamiento.

La estructura que se ha definido para el modelo donde se encuentra las funciones y las importaciones realizadas es la siguiente:

Figura 37
Estructura del modelo



```
1 import pandas as pd
2 import numpy as np
3 import re
4 from sqlalchemy import create_engine, text, MetaData, Table, Column, Integer, String, Date
5 from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
6 from sklearn.ensemble import GradientBoostingRegressor
7 from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
8 from sklearn.preprocessing import LabelEncoder, StandardScaler
9 from sklearn.feature_extraction.text import TfidfVectorizer
10 from sklearn.pipeline import Pipeline
11 from sklearn.impute import SimpleImputer
12 import logging
13 import sys
14 from typing import Tuple
15 from datetime import datetime
16
17 logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
18
19 class IncidentPredictor:
20     def __init__(self, db_url: str):
21
22     def __crear_tabla_recomendaciones(self):
23
24     def obtener_datos(self) -> pd.DataFrame:
25
26     def detectar_keywords(self, texto: str) -> dict:
27
28     def __extract_features(self, df: pd.DataFrame) -> pd.DataFrame:
29
30     def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]:
31
32     def clasificar_severidad_combinada(self, frecuencia: float, keyword_severity: int) -> str:
33
34     def clasificar_frecuencia(self, frecuencia: float) -> str:
35
36     def generar_recomendacion(self, severidad: str, keywords: list, categoria: str, categoria_frecuencia: str) -> str:
37
38     def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None:
39
40     def predecir_y_guardar(self, df: pd.DataFrame) -> None:
41
42     def guardar_recomendacion(self, cliente: str, titulo: str, localidad: str, ...
43
44 if __name__ == "__main__":
```

Librerías utilizadas

Las librerías que se utilizan dentro del modelo permiten dotar al programa de

diferentes cualidades con el fin de ofrecer un mejor rendimiento y optimización del modelo, durante el desarrollo y mejora del modelo de machine Learning se fueron agregando más librerías a las en un principio ya existentes.

Para el manejo de datos se utilizaron las librerías pandas, numpy y re ya que permiten:

- Manipular datos en una estructura de tipo DataFrame
- Permite limpieza, transformación y análisis de datos
- Numpy nos permite utilizar herramientas para cálculos numéricos y manipulación de arrays
- Son esencial para el preprocesamiento y cálculos matemáticos

Para la conexión a la base de datos se utiliza sqlalchemy para:

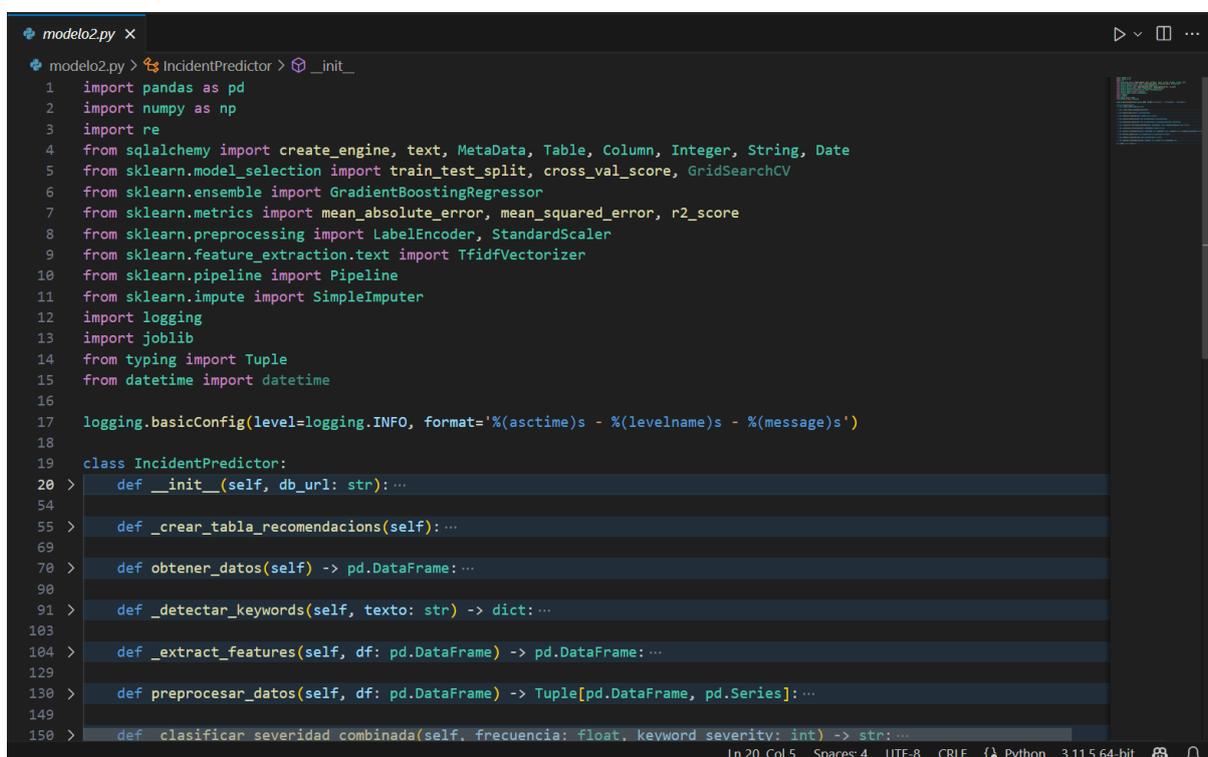
- Tener una conexión a la base de datos
- Permitir ejecutar consultas SQL en bruto
- Definir una estructura de las tablas a crear
- Poder definir tipo de datos en las columnas de las tablas creadas

Para enriquecer el modelo de Machine Learning se utilizan las siguientes librerías

- sklearn.model_selection
- sklearn.ensemble
- sklearn.metrics
- sklearn.preprocessing
- sklearn.feature_extraction.text
- sklearn.pipeline
- sklearn.impute

Las librerías especialmente orientadas al modelo de machine Learning seleccionado sirven para poder dividir los datos en conjuntos de entrenamiento y prueba, poder sacar métricas de evaluación en este caso para modelos de regresión, poder pasar variables categóricas a numéricas, poder normalizar los datos para obtener un mejor rendimiento del modelo, manejar valores faltantes y para poder aplicar el modelo de aprendizaje automático seleccionado basado en boosting de regresión

Figura 38



```
modelo2.py x
IncidentPredictor > _init_
1 import pandas as pd
2 import numpy as np
3 import re
4 from sqlalchemy import create_engine, text, MetaData, Table, Column, Integer, String, Date
5 from sklearn.model_selection import train_test_split, cross_val_score, GridSearchCV
6 from sklearn.ensemble import GradientBoostingRegressor
7 from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
8 from sklearn.preprocessing import LabelEncoder, StandardScaler
9 from sklearn.feature_extraction.text import TfidfVectorizer
10 from sklearn.pipeline import Pipeline
11 from sklearn.impute import SimpleImputer
12 import logging
13 import joblib
14 from typing import Tuple
15 from datetime import datetime
16
17 logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
18
19 class IncidentPredictor:
20 > def __init__(self, db_url: str): ...
54
55 > def _crear_tabla_recomendaciones(self): ...
69
70 > def obtener_datos(self) -> pd.DataFrame: ...
90
91 > def _detectar_keywords(self, texto: str) -> dict: ...
103
104 > def _extract_features(self, df: pd.DataFrame) -> pd.DataFrame: ...
129
130 > def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]: ...
149
150 > def clasificar_severidad_combinada(self, frecuencia: float, keyword severity: int) -> str: ...
Ln 20, Col 5 Spaces: 4 UTF-8 CRLF Python 3.11.5 64-bit
```

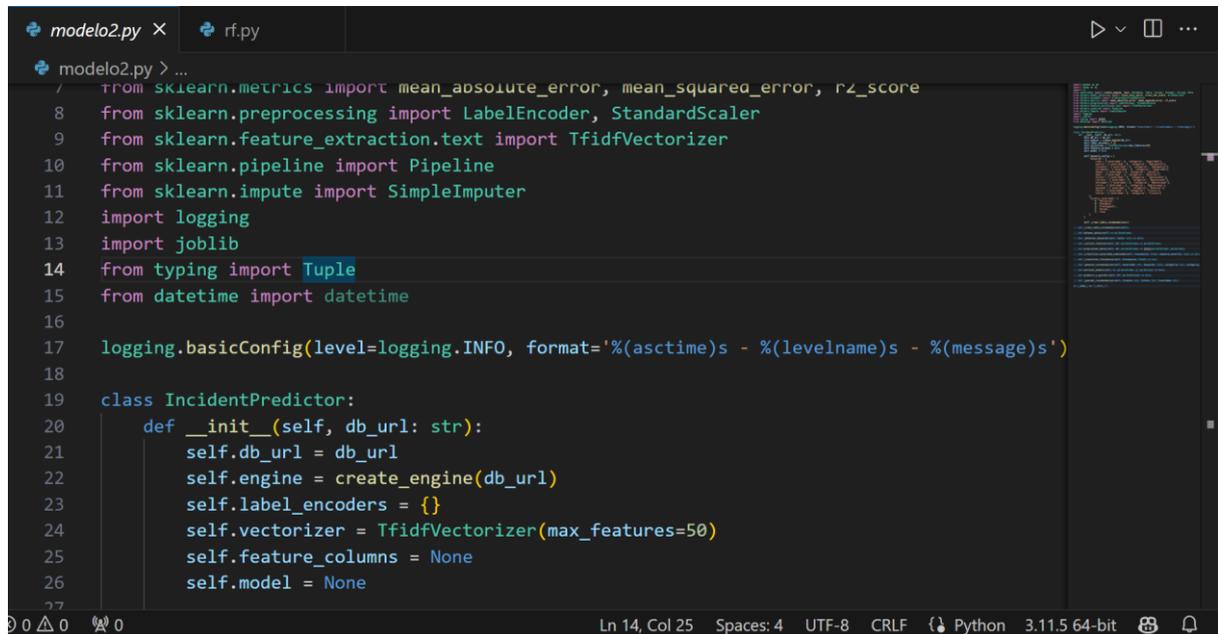
Configuración inicial y definición de clases

Se realiza una configuración del entorno y se define la clase principal de IncidentPredictor, se configura el sistema de logging con el fin de registrar cualquier mensaje que exista durante la ejecución del modelo ya que después se necesitara realizar una depuración y seguimiento del modelo.

La clase Incident Predictor es la clase principal con la inicia el modelo, esta clase engloba todo el conjunto de subclases o funciones que pueda llegar a tener el modelo. Esta clase se compone de variables, funciones, sentencias y sobre todo los componentes claves como la conexión a la base de datos MySQL y acciones de

vectorizar el texto.

Figura 39
Configuración inicial

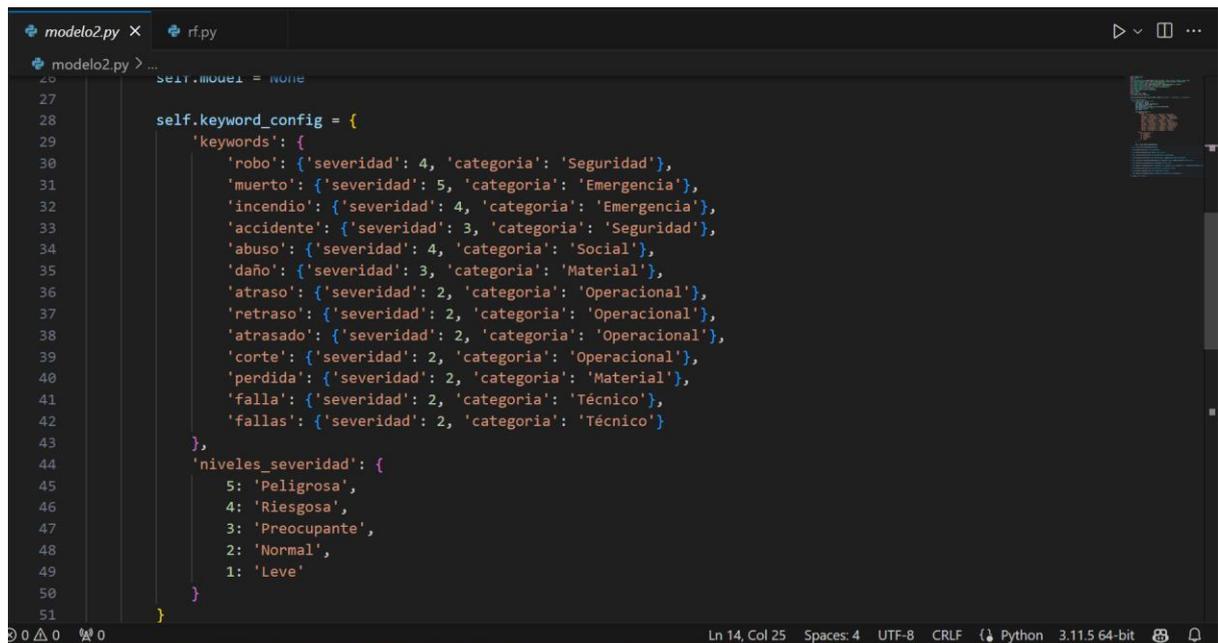


```
modelo2.py x rf.py
modelo2.py > ...
/ from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
8 from sklearn.preprocessing import LabelEncoder, StandardScaler
9 from sklearn.feature_extraction.text import TfidfVectorizer
10 from sklearn.pipeline import Pipeline
11 from sklearn.impute import SimpleImputer
12 import logging
13 import joblib
14 from typing import Tuple
15 from datetime import datetime
16
17 logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
18
19 class IncidentPredictor:
20     def __init__(self, db_url: str):
21         self.db_url = db_url
22         self.engine = create_engine(db_url)
23         self.label_encoders = {}
24         self.vectorizer = TfidfVectorizer(max_features=50)
25         self.feature_columns = None
26         self.model = None
27
```

Configuración de KeyWords y niveles de severidad

Esta parte tiene el propósito de establecer las palabras claves y niveles de severidad que puedan estar relacionados con los incidentes de seguridad, el diccionario de palabras se llama “keyword_config” dentro de este diccionario uno puede encontrar palabras claves como “robo”, “incendio”, palabras que este relacionado con los típicos incidentes de seguridad; en lo niveles de severidad que puede ser “Peligrosa”, “Riesgosa” van a tener su propio valor numérico, esta sección se encarga de poner analizar los incidentes y de acuerdo a eso poder asignarles una severidad y categoría basada en palabras encontrada, este no será el único analizar por el que pasara la data, a lo largo del modelo la data es predicha y se generan valores de frecuencia.

Figura 40
keywords

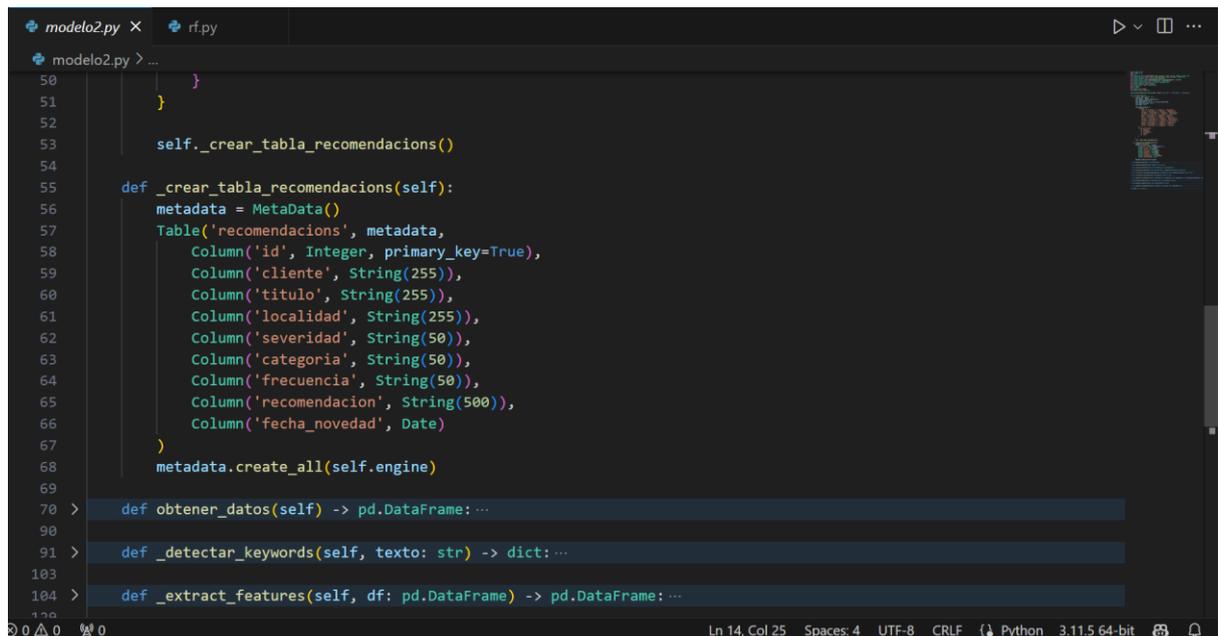


```
26 self.modelo = None
27
28 self.keyword_config = {
29     'keywords': {
30         'robo': {'severidad': 4, 'categoria': 'Seguridad'},
31         'muerto': {'severidad': 5, 'categoria': 'Emergencia'},
32         'incendio': {'severidad': 4, 'categoria': 'Emergencia'},
33         'accidente': {'severidad': 3, 'categoria': 'Seguridad'},
34         'abuso': {'severidad': 4, 'categoria': 'Social'},
35         'daño': {'severidad': 3, 'categoria': 'Material'},
36         'atraso': {'severidad': 2, 'categoria': 'Operacional'},
37         'retraso': {'severidad': 2, 'categoria': 'Operacional'},
38         'atrasado': {'severidad': 2, 'categoria': 'Operacional'},
39         'corte': {'severidad': 2, 'categoria': 'Operacional'},
40         'perdida': {'severidad': 2, 'categoria': 'Material'},
41         'falla': {'severidad': 2, 'categoria': 'Técnico'},
42         'fallas': {'severidad': 2, 'categoria': 'Técnico'}
43     },
44     'niveles_severidad': {
45         5: 'Peligrosa',
46         4: 'Riesgosa',
47         3: 'Preocupante',
48         2: 'Normal',
49         1: 'Leve'
50     }
51 }
```

Creación de la tabla recomendaciones

Es fase del script tiene la finalidad de poner crear una tabla en la base de datos con el fin de tener un lugar donde se puedan guardar las recomendaciones generada por el modelo, aquí se definen las siguientes tablas, así como se muestra a continuación:

Figura 41
Tabla recomendaciones



```
50     }
51 }
52
53     self._crear_tabla_recomendacions()
54
55 def _crear_tabla_recomendacions(self):
56     metadata = MetaData()
57     Table('recomendacions', metadata,
58         Column('id', Integer, primary_key=True),
59         Column('cliente', String(255)),
60         Column('titulo', String(255)),
61         Column('localidad', String(255)),
62         Column('severidad', String(50)),
63         Column('categoria', String(50)),
64         Column('frecuencia', String(50)),
65         Column('recomendacion', String(500)),
66         Column('fecha_novedad', Date)
67     )
68     metadata.create_all(self.engine)
69
70 > def obtener_datos(self) -> pd.DataFrame: ...
90
91 > def detectar_keywords(self, texto: str) -> dict: ...
103
104 > def extract_features(self, df: pd.DataFrame) -> pd.DataFrame: ...
```

Obtención de Datos

Durante esta etapa se realiza la obtención de data de incidentes que existen desde la base de datos, primero se realiza una consulta SQL con el fin de traer los datos de la base dentro de la tabla novedades, Aquí se filtran registros inválidos por ejemplo datos vacíos o con caracteres que no representen nada. Se aplicará un filtrado al título vacíos, las fechas inconsistentes o que no cuenten con una fecha adecuada. Estos datos serán obtenidos para entrenar el modelo y generar recomendaciones.

Figura 42
Obtención de datos

```
modelo2.py X it.py
incidentPredictor > _crear_tabla_recomendaciones
9 from sklearn.feature_extraction.text import TfidfVectorizer
10 from sklearn.pipeline import Pipeline
11 from sklearn.ignite import SimpleInquirer
12 import logging
13 import joblib
14 from typing import Tuple
15 from datetime import datetime
16
17 logging.basicConfig(level=logging.INFO, format='%(asctime)s - %(levelname)s - %(message)s')
18
19 class IncidentPredictor:
20     def __init__(self, db_url: str):
21
22     def _crear_tabla_recomendaciones(self):
23         metadata = Metadata()
24         Table('recomendaciones', metadata,
25              Column('id', Integer, primary_key=True),
26              Column('cliente', String(255)),
27              Column('titulo', String(255)),
28              Column('localidad', String(255)),
29              Column('severidad', String(50)),
30              Column('categoria', String(50)),
31              Column('frecuencia', String(50)),
32              Column('recomendacion', String(500)),
33              Column('fecha_novedad', Date)
34         )
35         metadata.create_all(self.engine)
36
37     def obtener_datos(self) -> pd.DataFrame:
38
39     def _detectar_keywords(self, texto: str) -> dict:
40
41     def _extract_features(self, df: pd.DataFrame) -> pd.DataFrame:
42
43     def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]:
44
45     def _clasificar_severidad_combinada(self, frecuencia: float, keyword_severity: int) -> str:
46
47     def _clasificar_frecuencia(self, frecuencia: float) -> str:
48
49     def _generar_recomendacion(self, severidad: str, keywords: list, categoria: str, categoria_frecuencia: str) -> str:
50
51     def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None:
52
53     def predecir_y_guardar(self, df: pd.DataFrame) -> None:
54
55     def _guardar_recomendacion(self, cliente: str, titulo: str, localidad: str, ...
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040
1041
1042
1043
1044
1045
1046
1047
1048
1049
1050
1051
1052
1053
1054
1055
1056
1057
1058
1059
1060
1061
1062
1063
1064
1065
1066
1067
1068
1069
1070
1071
1072
1073
1074
1075
1076
1077
1078
1079
1080
1081
1082
1083
1084
1085
1086
1087
1088
1089
1090
1091
1092
1093
1094
1095
1096
1097
1098
1099
1100
1101
1102
1103
1104
1105
1106
1107
1108
1109
1110
1111
1112
1113
1114
1115
1116
1117
1118
1119
1120
1121
1122
1123
1124
1125
1126
1127
1128
1129
1130
1131
1132
1133
1134
1135
1136
1137
1138
1139
1140
1141
1142
1143
1144
1145
1146
1147
1148
1149
1150
1151
1152
1153
1154
1155
1156
1157
1158
1159
1160
1161
1162
1163
1164
1165
1166
1167
1168
1169
1170
1171
1172
1173
1174
1175
1176
1177
1178
1179
1180
1181
1182
1183
1184
1185
1186
1187
1188
1189
1190
1191
1192
1193
1194
1195
1196
1197
1198
1199
1200
1201
1202
1203
1204
1205
1206
1207
1208
1209
1210
1211
1212
1213
1214
1215
1216
1217
1218
1219
1220
1221
1222
1223
1224
1225
1226
1227
1228
1229
1230
1231
1232
1233
1234
1235
1236
1237
1238
1239
1240
1241
1242
1243
1244
1245
1246
1247
1248
1249
1250
1251
1252
1253
1254
1255
1256
1257
1258
1259
1260
1261
1262
1263
1264
1265
1266
1267
1268
1269
1270
1271
1272
1273
1274
1275
1276
1277
1278
1279
1280
1281
1282
1283
1284
1285
1286
1287
1288
1289
1290
1291
1292
1293
1294
1295
1296
1297
1298
1299
1300
1301
1302
1303
1304
1305
1306
1307
1308
1309
1310
1311
1312
1313
1314
1315
1316
1317
1318
1319
1320
1321
1322
1323
1324
1325
1326
1327
1328
1329
1330
1331
1332
1333
1334
1335
1336
1337
1338
1339
1340
1341
1342
1343
1344
1345
1346
1347
1348
1349
1350
1351
1352
1353
1354
1355
1356
1357
1358
1359
1360
1361
1362
1363
1364
1365
1366
1367
1368
1369
1370
1371
1372
1373
1374
1375
1376
1377
1378
1379
1380
1381
1382
1383
1384
1385
1386
1387
1388
1389
1390
1391
1392
1393
1394
1395
1396
1397
1398
1399
1400
1401
1402
1403
1404
1405
1406
1407
1408
1409
1410
1411
1412
1413
1414
1415
1416
1417
1418
1419
1420
1421
1422
1423
1424
1425
1426
1427
1428
1429
1430
1431
1432
1433
1434
1435
1436
1437
1438
1439
1440
1441
1442
1443
1444
1445
1446
1447
1448
1449
1450
1451
1452
1453
1454
1455
1456
1457
1458
1459
1460
1461
1462
1463
1464
1465
1466
1467
1468
1469
1470
1471
1472
1473
1474
1475
1476
1477
1478
1479
1480
1481
1482
1483
1484
1485
1486
1487
1488
1489
1490
1491
1492
1493
1494
1495
1496
1497
1498
1499
1500
1501
1502
1503
1504
1505
1506
1507
1508
1509
1510
1511
1512
1513
1514
1515
1516
1517
1518
1519
1520
1521
1522
1523
1524
1525
1526
1527
1528
1529
1530
1531
1532
1533
1534
1535
1536
1537
1538
1539
1540
1541
1542
1543
1544
1545
1546
1547
1548
1549
1550
1551
1552
1553
1554
1555
1556
1557
1558
1559
1560
1561
1562
1563
1564
1565
1566
1567
1568
1569
1570
1571
1572
1573
1574
1575
1576
1577
1578
1579
1580
1581
1582
1583
1584
1585
1586
1587
1588
1589
1590
1591
1592
1593
1594
1595
1596
1597
1598
1599
1600
1601
1602
1603
1604
1605
1606
1607
1608
1609
1610
1611
1612
1613
1614
1615
1616
1617
1618
1619
1620
1621
1622
1623
1624
1625
1626
1627
1628
1629
1630
1631
1632
1633
1634
1635
1636
1637
1638
1639
1640
1641
1642
1643
1644
1645
1646
1647
1648
1649
1650
1651
1652
1653
1654
1655
1656
1657
1658
1659
1660
1661
1662
1663
1664
1665
1666
1667
1668
1669
1670
1671
1672
1673
1674
1675
1676
1677
1678
1679
1680
1681
1682
1683
1684
1685
1686
1687
1688
1689
1690
1691
1692
1693
1694
1695
1696
1697
1698
1699
1700
1701
1702
1703
1704
1705
1706
1707
1708
1709
1710
1711
1712
1713
1714
1715
1716
1717
1718
1719
1720
1721
1722
1723
1724
1725
1726
1727
1728
1729
1730
1731
1732
1733
1734
1735
1736
1737
1738
1739
1740
1741
1742
1743
1744
1745
1746
1747
1748
1749
1750
1751
1752
1753
1754
1755
1756
1757
1758
1759
1760
1761
1762
1763
1764
1765
1766
1767
1768
1769
1770
1771
1772
1773
1774
1775
1776
1777
1778
1779
1780
1781
1782
1783
1784
1785
1786
1787
1788
1789
1790
1791
1792
1793
1794
1795
1796
1797
1798
1799
1800
1801
1802
1803
1804
1805
1806
1807
1808
1809
1810
1811
1812
1813
1814
1815
1816
1817
1818
1819
1820
1821
1822
1823
1824
1825
1826
1827
1828
1829
1830
1831
1832
1833
1834
1835
1836
1837
1838
1839
1840
1841
1842
1843
1844
1845
1846
1847
1848
1849
1850
1851
1852
1853
1854
1855
1856
1857
1858
1859
1860
1861
1862
1863
1864
1865
1866
1867
1868
1869
1870
1871
1872
1873
1874
1875
1876
1877
1878
1879
1880
1881
1882
1883
1884
1885
1886
1887
1888
1889
1890
1891
1892
1893
1894
1895
1896
1897
1898
1899
1900
1901
1902
1903
1904
1905
1906
1907
1908
1909
1910
1911
1912
1913
1914
1915
1916
1917
1918
1919
1920
1921
1922
1923
1924
1925
1926
1927
1928
1929
1930
1931
1932
1933
1934
1935
1936
1937
1938
1939
1940
1941
1942
1943
1944
1945
1946
1947
1948
1949
1950
1951
1952
1953
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538
2539
2540
2541
2542
2543
2544
2545
2546
2547
2548
2549
2550
2551
2552
2553
2554
2555
2556
2557
2558
2559
2560
2561
2562
2563
2564
2565
2566
2567
2568
2569
2570
2571
2572
2573
2574
2575
2576
2577
2578
2579
2580
2581
2582
2583
2584
2585
2586
2587
2588
2589
2590
2591
2592
2
```

Figura 44
Detección de keywords

```

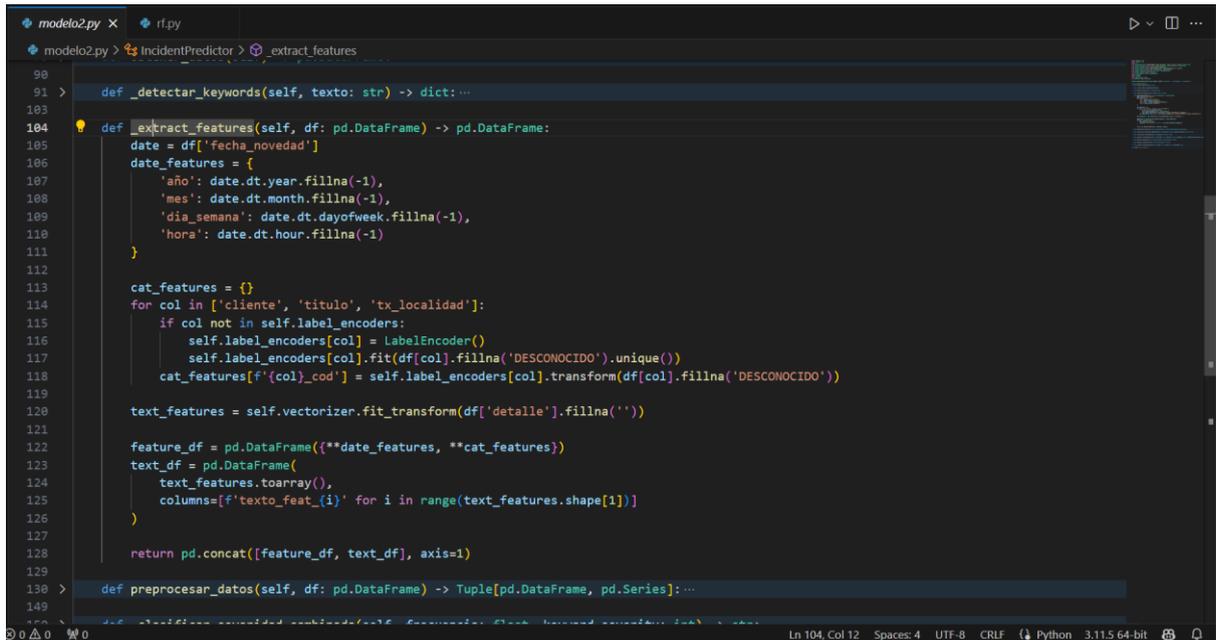
10 from sklearn.feature_extraction.text import TfidfVectorizer
11 from sklearn.pipeline import Pipeline
12 from sklearn.impute import SimpleImputer
13 import logging
14 import joblib
15 from typing import Tuple
16 from datetime import datetime
17
18 logging.basicConfig(level=logging.INFO, format="%asctimes - %(levelname)s - %(message)s")
19
20 class IncidentPredictor:
21     def __init__(self, db_url: str):
22
23     def _crear_tabla_recomendaciones(self):
24
25     def obtener_datos(self) -> pd.DataFrame:
26
27     def detectar_keywords(self, texto: str) -> dict:
28         texto = texto.lower()
29         encontradas = {'keywords': [], 'max_severidad': 1, 'categoria': 'General'}
30
31         for kw, config in self.keyword_config['keywords'].items():
32             if re.search(fr'\b{re.escape(kw)}\b', texto):
33                 encontradas['keywords'].append(kw)
34                 if config['severidad'] > encontradas['max_severidad']:
35                     encontradas['max_severidad'] = config['severidad']
36                 encontradas['categoria'] = config['categoria']
37
38         return encontradas
39
40     def _extract_features(self, df: pd.DataFrame) -> pd.DataFrame:
41
42     def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]:
43
44     def _clasificar_severidad_combinada(self, frecuencia: float, keyword_severity: int) -> str:
45
46     def _clasificar_frecuencia(self, frecuencia: float) -> str:
47
48     def _generar_recomendacion(self, severidad: str, keywords: list, categoria: str, categoria_frecuencia: str) -> str:
49
50     def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None:
51
52     def predecir_y_guardar(self, df: pd.DataFrame) -> None:
53
54     def _guardar_recomendacion(self, cliente: str, titulo: str, localidad: str,
55
56     if __name__ == "__main__":

```

Extracción de características

El propósito de esta fase es convertir los datos en características que puedan ser leídas o interpretadas por el modelo de machine Learning, aquí se encuentra la función `extract_features` que extrae características de los datos donde se podrían alojar características como la fecha y la categoría, además se encarga de vectorizar el texto de la columna detalle usando `TfidfVectorizer`

Figura 45
Extracción características



```
90
91 > def _detectar_keywords(self, texto: str) -> dict:...
103
104 def _extract_features(self, df: pd.DataFrame) -> pd.DataFrame:
105     date = df['fecha_novedad']
106     date_features = {
107         'año': date.dt.year.fillna(-1),
108         'mes': date.dt.month.fillna(-1),
109         'dia_semana': date.dt.dayofweek.fillna(-1),
110         'hora': date.dt.hour.fillna(-1)
111     }
112
113     cat_features = {}
114     for col in ['cliente', 'titulo', 'tx_localidad']:
115         if col not in self.label_encoders:
116             self.label_encoders[col] = LabelEncoder()
117             self.label_encoders[col].fit(df[col].fillna('DESCONOCIDO').unique())
118             cat_features[f'{col}_cod'] = self.label_encoders[col].transform(df[col].fillna('DESCONOCIDO'))
119
120     text_features = self.vectorizer.fit_transform(df['detalle'].fillna(''))
121
122     feature_df = pd.DataFrame(**date_features, **cat_features)
123     text_df = pd.DataFrame(
124         text_features.toarray(),
125         columns=[f'texto_feat_{i}' for i in range(text_features.shape[1])]
126     )
127
128     return pd.concat([feature_df, text_df], axis=1)
129
130 > def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]:...
```

Preprocesamiento de los Datos

Realizamos el procesamiento de datos mediante la función *preprocesar_datos* que lo primero que hace es una extracción de las características entre ellas están las categóricas, de fecha y textuales. Posteriormente, pasa a crear la variable objetivo (y) mediante el cálculo de frecuencia por clientes y este resultado por la agrupación de fechas.

Figura 46
Función procesar datos

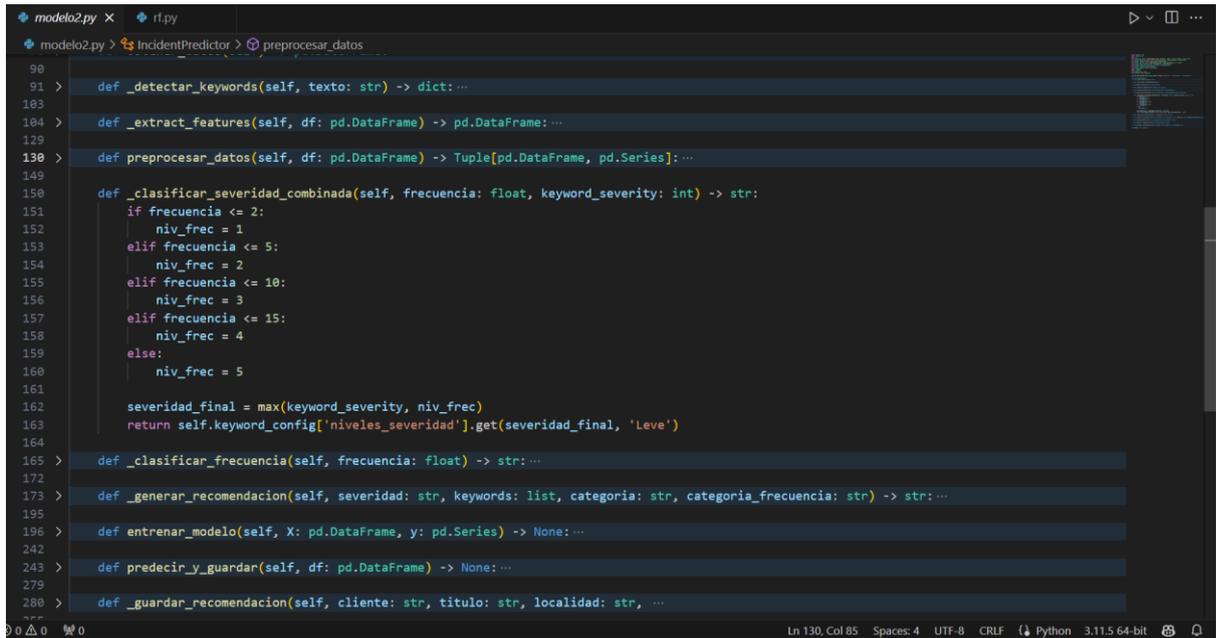
```
modelo2.py x r1.py
IncidentPredictor > _extract_features
90
91 > def _detectar_keywords(self, texto: str) -> dict: ...
103
104 > def _extract_features(self, df: pd.DataFrame) -> pd.DataFrame: ...
129
130 def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]:
131     # Detección de keywords
132     df['keywords_info'] = df.apply(
133         lambda row: self._detectar_keywords(f"{row['titulo']} {row['detalle']}"),
134         axis=1
135     )
136
137     # Calcular frecuencia histórica sin data leakage
138     df = df.sort_values('fecha_novedad')
139     df['frecuencia_hist'] = df.groupby(['cliente', 'tx_localidad']).cumcount()
140
141     # Validar target
142     logging.info("Distribución de frecuencia_hist:\n%s", df['frecuencia_hist'].describe())
143
144     # Generar features
145     X = self._extract_features(df)
146     self.feature_columns = X.columns.tolist()
147
148     return X, df['frecuencia_hist']
149
150 > def _clasificar_severidad_combinada(self, frecuencia: float, keyword_severity: int) -> str: ...
164
165 > def _clasificar_frecuencia(self, frecuencia: float) -> str: ...
172
173 > def _generar_recomendacion(self, severidad: str, keywords: list, categoria: str, categoria_frecuencia: str) -> str: ...
195
Ln 104, Col 67  Spaces: 4  UTF-8  CRLF  Python 3.11.5 64-bit
```

Clasificar severidad y frecuencia

Esta etapa del modelo sirve para por clasificar la severidad y la frecuencia de los incidentes, `_clasificar_severidad_combinada` se usa de manera general basada en la clave con la frecuencia histórica con el fin de determinar o mostrar que severidad va a ser. Por otro lado, tenemos `_clasificar_frecuencia` que busca clasificar en categorías ya sea baja, media y alta con el fin de basarse en el número que se obtuvo durante la ejecución.

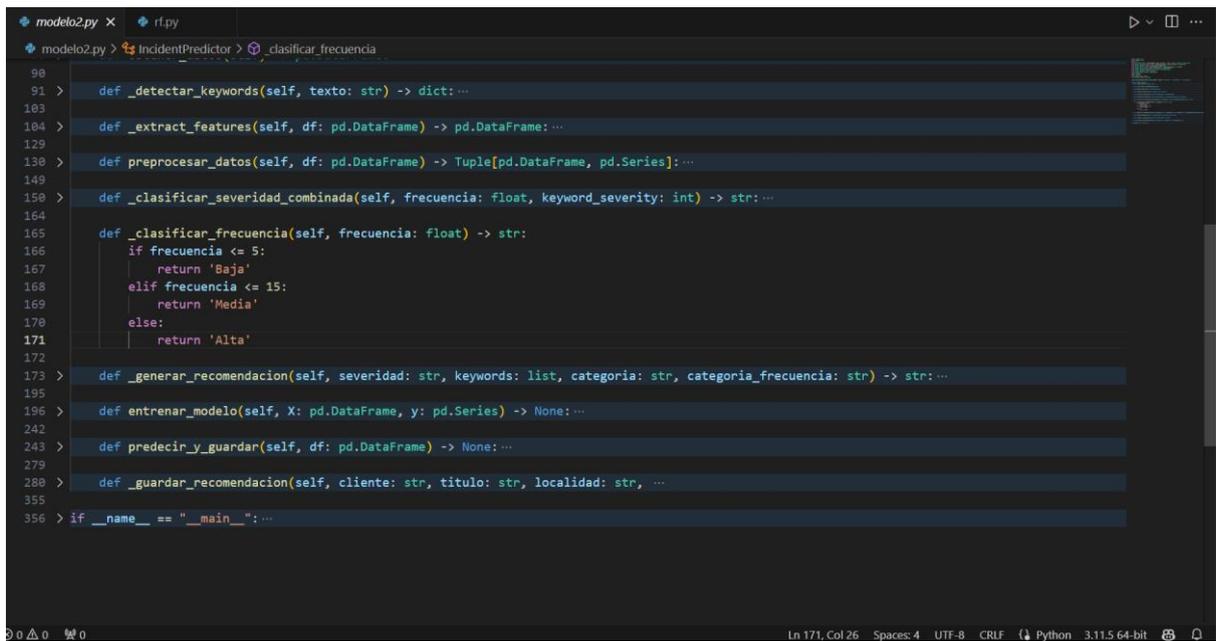
El objetivo de crear estas funciones es poder ayudar a categorizar los incidentes con el fin de poder generar recomendaciones correctas.

Figura 47
Severidad y frecuencia



```
90
91 > def _detectar_keywords(self, texto: str) -> dict: ...
103
104 > def _extract_features(self, df: pd.DataFrame) -> pd.DataFrame: ...
129
130 > def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]: ...
149
150 def _clasificar_severidad_combinada(self, frecuencia: float, keyword_severity: int) -> str:
151     if frecuencia <= 2:
152         niv_freq = 1
153     elif frecuencia <= 5:
154         niv_freq = 2
155     elif frecuencia <= 10:
156         niv_freq = 3
157     elif frecuencia <= 15:
158         niv_freq = 4
159     else:
160         niv_freq = 5
161
162     severidad_final = max(keyword_severity, niv_freq)
163     return self.keyword_config['niveles_severidad'].get(severidad_final, 'Leve')
164
165 > def _clasificar_frecuencia(self, frecuencia: float) -> str: ...
172
173 > def _generar_recomendacion(self, severidad: str, keywords: list, categoria: str, categoria_frecuencia: str) -> str: ...
195
196 > def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None: ...
242
243 > def predecir_y_guardar(self, df: pd.DataFrame) -> None: ...
279
280 > def _guardar_recomendacion(self, cliente: str, titulo: str, localidad: str, ...
...
```

Figura 48
Severidad y frecuencia



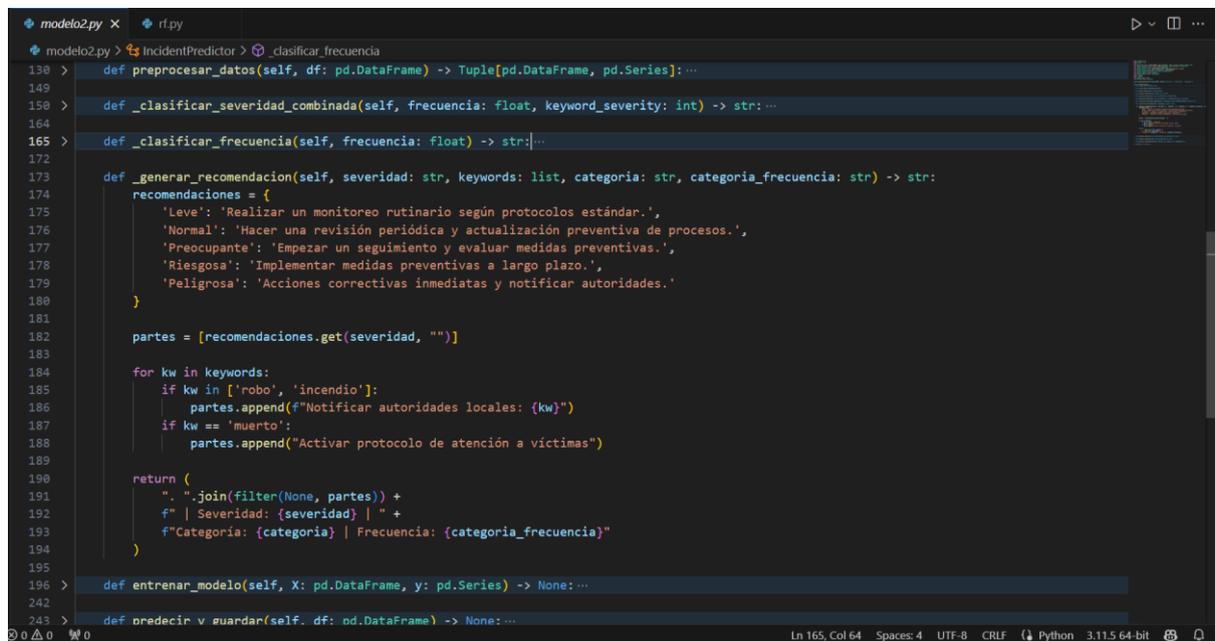
```
90
91 > def _detectar_keywords(self, texto: str) -> dict: ...
103
104 > def _extract_features(self, df: pd.DataFrame) -> pd.DataFrame: ...
129
130 > def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]: ...
149
150 > def _clasificar_severidad_combinada(self, frecuencia: float, keyword_severity: int) -> str: ...
164
165 def _clasificar_frecuencia(self, frecuencia: float) -> str:
166     if frecuencia <= 5:
167         return 'Baja'
168     elif frecuencia <= 15:
169         return 'Media'
170     else:
171         return 'Alta'
172
173 > def _generar_recomendacion(self, severidad: str, keywords: list, categoria: str, categoria_frecuencia: str) -> str: ...
195
196 > def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None: ...
242
243 > def predecir_y_guardar(self, df: pd.DataFrame) -> None: ...
279
280 > def _guardar_recomendacion(self, cliente: str, titulo: str, localidad: str, ...
355
356 > if __name__ == "__main__": ...
```

Generación de recomendaciones

Aquí se crean las recomendaciones basadas en la severidad, categoría y frecuencia de los incidentes. Aquí es donde se utiliza el diccionario de palabras donde cuando se identifique un nivel de severidad pueda darse una recomendación en base

a eso, además de también ponerle a la recomendación realizar algunas instrucciones.

Figura 49
Generación de recomendaciones



```
130 > def preprocesar_datos(self, df: pd.DataFrame) -> Tuple[pd.DataFrame, pd.Series]: ...
149
150 > def _clasificar_severidad_combinada(self, frecuencia: float, keyword_severity: int) -> str: ...
164
165 > def _clasificar_frecuencia(self, frecuencia: float) -> str: ...
172
173
174 def _generar_recomendacion(self, severidad: str, keywords: list, categoria: str, categoria_frecuencia: str) -> str:
175     recomendaciones = {
176         'Leve': 'Realizar un monitoreo rutinario según protocolos estándar.',
177         'Normal': 'Hacer una revisión periódica y actualización preventiva de procesos.',
178         'Preocupante': 'Empezar un seguimiento y evaluar medidas preventivas.',
179         'Riesgosa': 'Implementar medidas preventivas a largo plazo.',
180         'Peligrosa': 'Acciones correctivas inmediatas y notificar autoridades.'
181     }
182     partes = [recomendaciones.get(severidad, "")]
183
184     for kw in keywords:
185         if kw in ['robo', 'incendio']:
186             partes.append(f"Notificar autoridades locales: {kw}")
187         if kw == 'muerto':
188             partes.append("Activar protocolo de atención a víctimas")
189
190     return (
191         ". ".join(filter(None, partes)) +
192         f" | Severidad: {severidad} | " +
193         f"Categoria: {categoria} | Frecuencia: {categoria_frecuencia}"
194     )
195
196 > def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None: ...
242
243 > def predecir o guardar(self, df: pd.DataFrame) -> None: ...
```

Entrenamiento del modelo

Dentro de esta función el método recibe los dos parámetros que son X, que representa un DataFrame con las características del conjunto de datos, y Y que es la serie de valores que queremos predecir. Cuando el modelo pase estos datos lo primero que se realiza es la división del grupo de datos de entrenamiento y el grupo de datos de prueba que se realiza de manera de 80/20, así se garantiza que el modelo pueda aprender con una gran parte de los datos y que por otro lado ya cuando se vaya a realizar la prueba sea evaluado con datos que no ha visto, así podremos medir su capacidad de generalizar.

Figura 50
Entrenamiento del modelo

```
def entrenar_modelo(self, X: pd.DataFrame, y: pd.Series) -> None:
    # Split de datos
    X_train, X_test, y_train, y_test = train_test_split(
        X, y, test_size=0.2, random_state=42
    )

    # Configuración del modelo
    self.model = Pipeline([
        ('imputer', SimpleImputer(strategy='median')),
        ('scaler', StandardScaler()),
        ('model', GradientBoostingRegressor(
            n_estimators=80,
            learning_rate=0.15,
            max_depth=3,
            min_samples_split=15,
            min_samples_leaf=10,
            subsample=0.7,
            random_state=42,
            validation_fraction=0.2,
            n_iter_no_change=8
        ))
    ])

    # Validación cruzada
    cv_scores = cross_val_score(self.model, X_train, y_train, cv=5, scoring='r2')
    logging.info("R² Validación Cruzada (entrenamiento): %.3f ± %.3f",
                cv_scores.mean(), cv_scores.std())

    # Entrenamiento final
    self.model.fit(X_train, y_train)

    # Evaluación
    y_pred = self.model.predict(X_test)
    logging.info("Métricas finales (prueba):")
    logging.info("R²: %.3f", r2_score(y_test, y_pred))
    logging.info("MAE: %.3f", mean_absolute_error(y_test, y_pred))
    logging.info("RMSE: %.3f", np.sqrt(mean_squared_error(y_test, y_pred)))
```

Después se configura un Pipeline que se denomina una estructura que ayuda a facilitar el preprocesamiento y entrenamiento del modelo mediante una secuencia de diferentes procesos. Todo este proceso se divide en 3 partes primero la imputación de datos faltantes, después la estandarización de valores y por último el modelo de machine Learning; una vez se configure el modelo se hace una validación cruzada donde se realiza una partición en 5 es decir un CV de 5 con el fin de evaluar el rendimiento que llegue a tener el modelo en diferentes subconjuntos de datos de entrenamiento.

Figura 51
Pipeline

```
203 self.model = Pipeline(  
204     ('imputer', SimpleImputer(strategy='median')),  
205     ('scaler', StandardScaler()),  
206     ('model', GradientBoostingRegressor(  
207         n_estimators=80,  
208         learning_rate=0.15,  
209         max_depth=3,  
210         min_samples_split=15,  
211         min_samples_leaf=10,  
212         subsample=0.7,  
213         random_states=42,  
214         validation_fraction=0.2,  
215         n_iter_no_change=8  
216     ))  
217 )  
218  
219 # Validación cruzada  
220 cv_scores = cross_val_score(self.model, X_train, y_train, cv=5, scoring='r2')  
221 logging.info("R2 Validación Cruzada (entrenamiento): %.3f ± %.3f",  
222             cv_scores.mean(), cv_scores.std())  
223  
224 # Entrenamiento final  
225 self.model.fit(X_train, y_train)  
226  
227 # Evaluación  
228 y_pred = self.model.predict(X_test)  
229 logging.info("Métricas Finales (prueba):")  
230 logging.info("R2: %.3f", r2_score(y_test, y_pred))  
231 logging.info("MAE: %.3f", mean_absolute_error(y_test, y_pred))  
232 logging.info("RMSE: %.3f", np.sqrt(mean_squared_error(y_test, y_pred)))  
233  
234 # Guardar modelo  
235 joblib.dump(  
236     'model': self.model,  
237     'label_encoders': self.label_encoders,  
238     'vectorizer': self.vectorizer,  
239     'feature_columns': self.feature_columns,  
240     'keyword_config': self.keyword_config  
241 ), 'modelo_protetaxi.joblib')
```

Predicción y guardar resultados

El propósito principal de esta función es poder generar predicción y poder clasificar los resultados y almacenarlos en la base de datos mediante la tabla de recomendaciones que se había definido desde un inicio.

Aquí primero se verifica que el dataframe de entrada contenga las columnas que aparecen dentro de la captura de pantalla. En caso de haber datos faltantes saltará un error que impedirá continuar con la ejecución, una vez se validan los datos se programa para que se extraigan solamente las características en este caso las features más relevantes esto transforma el dataframe en un formato adecuado para que el modelo pueda realizar las predicciones necesarias.

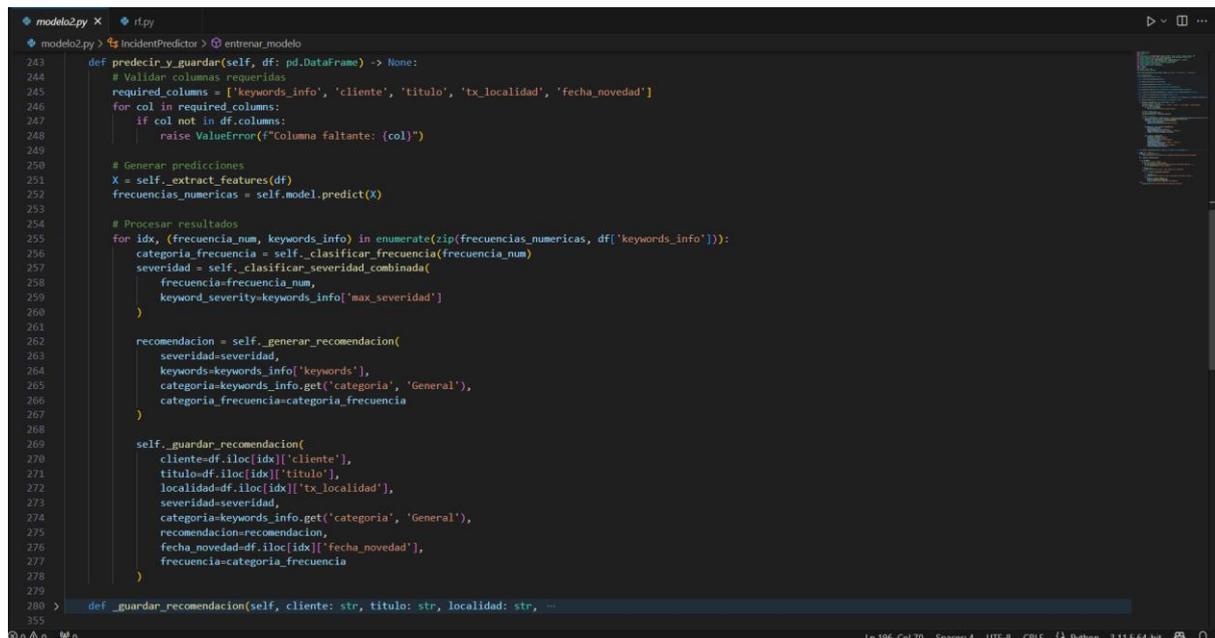
Para generar las predicciones se usa `self.model.predict(X)` de la cual se obtiene una lista de valores que son las frecuencias en formato numérico, esto valores son los que representan el valor estimado de la novedad o incidencia que se haya predicho en base a los datos proporcionados.

Una vez finaliza el proceso anterior ahora viene el procesamiento de los resultados obtenidos aquí es donde esta función realizara una clasificación de la

frecuencia en base a su valor número y la cambiara para que ahora la frecuencia coja un valor string que puede ser “alta”, “media” o “baja”. además, aparte de esto aquí se combinan los resultados del modelo de machine Learning con los de la clasificación inicial que se realizó se realiza una determinación de la severidad con el método `_clasificar_severidad_combinada` aquí es donde los resultados del modelo de machine Learning se combinan o se asocian con la severidad que se asocia de acuerdo con las palabras clave esto permite asignar un nivel de importancia a la incidencia.

Para finalizar se genera la recomendación final que es la que se verá en base a todos los resultados obtenidos, aquí se llamará al método `_generar_recomendacion` () que elaborará una recomendación en base a la severidad del evento, las palabras claves asociadas, la categoría de la incidencia y la clasificación de la frecuencia obtenida antes.

Figura 52
Predicción y guardar datos



```
def predecir_y_guardar(self, df: pd.DataFrame) -> None:
    # Validar columnas requeridas
    required_columns = ['keywords_info', 'cliente', 'titulo', 'tx_localidad', 'fecha_novedad']
    for col in required_columns:
        if col not in df.columns:
            raise ValueError(f"Columna faltante: {col}")

    # Generar predicciones
    X = self._extract_features(df)
    frecuencias_numericas = self.modelo.predict(X)

    # Procesar resultados
    for idx, (frecuencia_num, keywords_info) in enumerate(zip(frecuencias_numericas, df['keywords_info'])):
        categoria_frecuencia = self._clasificar_frecuencia(frecuencia_num)
        severidad = self._clasificar_severidad_combinada(
            frecuencia=frecuencia_num,
            keyword_severity=keywords_info['max_severidad']
        )

        recomendacion = self._generar_recomendacion(
            severidad=severidad,
            keywords=keywords_info['keywords'],
            categoria=keywords_info.get('categoria', 'General'),
            categoria_frecuencia=categoria_frecuencia
        )

        self._guardar_recomendacion(
            cliente=df.iloc[idx]['cliente'],
            titulo=df.iloc[idx]['titulo'],
            localidad=df.iloc[idx]['tx_localidad'],
            severidad=severidad,
            categoria=keywords_info.get('categoria', 'General'),
            recomendacion=recomendacion,
            fecha_novedad=df.iloc[idx]['fecha_novedad'],
            frecuencia=categoria_frecuencia
        )

def _guardar_recomendacion(self, cliente: str, titulo: str, localidad: str, ...
```

Disponibilidad de Resultados

Las predicciones y recomendaciones son almacenadas en una base de datos

MySQL según la configuración del conector MySQL, como tal se guarda en la tabla *recomendacions* y si no existe en la base de datos registrada en el conector pues tiene la codificación de crearla; caso contrario la actualiza.

Figura 53
Código guardar recomendación

```
modelo2.py x rf.py
modelo2.py > IncidentPredictor
275     recomendacion=recomendacion,
276     fecha_novedad=df.iloc[idx]['fecha_novedad'],
277     frecuencia=categoria_frecuencia
278 )
279
280 def _guardar_recomendacion(self, cliente: str, titulo: str, localidad: str,
281                          severidad: str, categoria: str, recomendacion: str,
282                          fecha_novedad: Date, frecuencia: str) -> None:
283     # Convertir fecha
284     if isinstance(fecha_novedad, pd.Timestamp):
285         fecha_novedad = fecha_novedad.to_pydatetime()
286
287     query_check = """
288     SELECT id
289     FROM recomendacions
290     WHERE
291         (titulo = :titulo AND fecha_novedad = :fecha_novedad)
292         OR
293         (titulo = :titulo AND MONTH(fecha_novedad) = :mes AND YEAR(fecha_novedad) = :año)
294     LIMIT 1
295     """
296
297     query_update = """
298     UPDATE recomendacions
299     SET
300         cliente = :cliente,
301         localidad = :localidad,
302         severidad = :severidad,
303         categoria = :categoria,
304         frecuencia = :frecuencia,
305         recomendacion = :recomendacion,
306         fecha_novedad = :fecha_novedad
307     WHERE id = :id
308     """
309
310     query_insert = """
311     INSERT INTO recomendacions
312     (cliente, titulo, localidad, severidad, categoria,
313     frecuencia, recomendacion, fecha_novedad)
```

Figura 54
Código actualizar recomendaciones

```
modelo2.py x rf.py
modelo2.py > IncidentPredictor
291         (titulo = :titulo AND fecha_novedad = :fecha_novedad)
292         OR
293         (titulo = :titulo AND MONTH(fecha_novedad) = :mes AND YEAR(fecha_novedad) = :año)
294     LIMIT 1
295     """
296
297     query_update = """
298     UPDATE recomendacions
299     SET
300         cliente = :cliente,
301         localidad = :localidad,
302         severidad = :severidad,
303         categoria = :categoria,
304         frecuencia = :frecuencia,
305         recomendacion = :recomendacion,
306         fecha_novedad = :fecha_novedad
307     WHERE id = :id
308     """
```

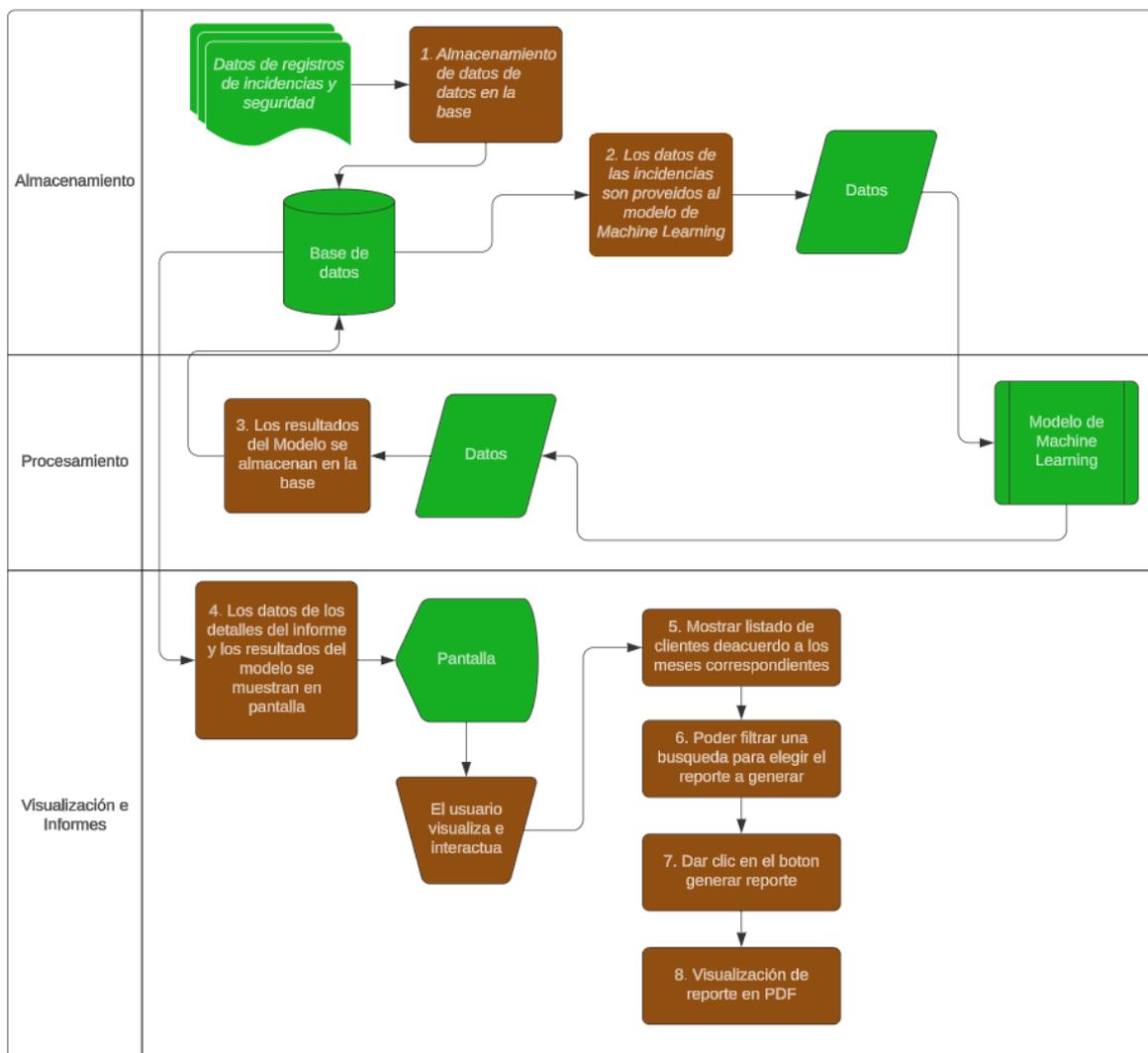
Como tal los datos quedan registrados en la base de datos de manera que cualquier usuario que los requiera consumir tenga la disponibilidad para realizarlo

Diagrama de flujo del prototipo

Para comprender el funcionamiento del prototipo, se presenta una gráfica que muestran los pasos o procesos que tiene que seguir el sistema para cumplir con la generación de los informes que son integrado de la mano con inteligencia artificial, este proceso es conocido como el flujo de trabajo del sistema y permite entender cuáles son las actividades para cumplir para que el resultado sea el desea o esperado.

A continuación, se presenta un diagrama del prototipo.

Figura 55
Diagrama de flujo del prototipo



Nota: La siguiente ilustración muestra cada uno de los pasos que se siguen para realizar la generación del informe desde la utilización de la data proporcionada hasta la visualización del informe generado.

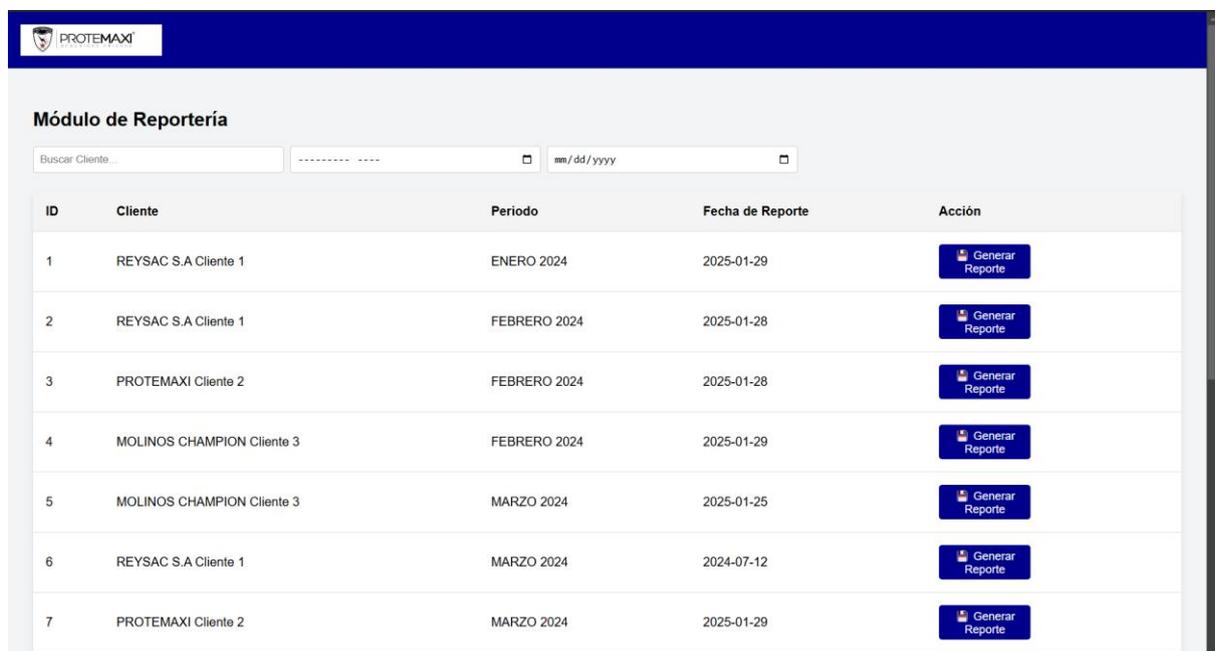
Resultados

Informe generado

Dentro página de generación de informes el usuario genera el archivo descargable en PDF al momento de darle clic al botón de “*generar informe*”, la estructura del informe se construye de 6 secciones, la última sección corresponde a la puesta de valor donde interviene la data proporcionada analizada por el modelo de Machine Learning cuyos resultados son puesto en forma de tabla donde se indican recomendaciones en base a la frecuencia de ocurrencia de las incidencia; el resto de secciones del informe corresponden a información que es precargada de la base de datos de PROTEMAXI, por motivos de seguridad dicha información se captura de la data proporcionada.

A continuación, se presenta la página de generación de informes donde se muestra la lista de clientes de acuerdo con el periodo correspondiente:

Figura 56
Página de generación de informes



ID	Cliente	Periodo	Fecha de Reporte	Acción
1	REYSAC S.A Cliente 1	ENERO 2024	2025-01-29	Generar Reporte
2	REYSAC S.A Cliente 1	FEBRERO 2024	2025-01-28	Generar Reporte
3	PROTEMAXI Cliente 2	FEBRERO 2024	2025-01-28	Generar Reporte
4	MOLINOS CHAMPION Cliente 3	FEBRERO 2024	2025-01-29	Generar Reporte
5	MOLINOS CHAMPION Cliente 3	MARZO 2024	2025-01-25	Generar Reporte
6	REYSAC S.A Cliente 1	MARZO 2024	2024-07-12	Generar Reporte
7	PROTEMAXI Cliente 2	MARZO 2024	2025-01-29	Generar Reporte

Al momento de dar clic en el botón “*generar reporte*”, se descargará un documento PDF que es el informe mensual de cada cliente que se realiza en base a la información de las novedades reportadas en el mes e información de seguridad

adicional que la empresa recopila.

A continuación, se presenta una captura del documento PDF que se genera:

Figura 57
Páginas del Informe



INFORME DE GESTIÓN - PROTEMAXI

CORRESPONDIENTE AL SERVICIO DE SEGURIDAD FÍSICA DE PROTEMAXI C. LTDA., EN EL PROYECTO CLIENTE 2

CLIENTE:	Cliente 2	
DESTINATARIOS:	DANIEL PINTADO	
FECHA DE REPORTE:	2025-01-29	
PERIODO:	FEBRERO 2024	NO. DE REPORTE: 3

1. INTRODUCCIÓN

La información presentada en el siguiente informe corresponde a las actividades de seguridad, control y prevención que fueron realizadas por nuestro personal en las instalaciones de nuestro cliente Cliente 2 durante el mes de FEBRERO, en los siguientes sitios donde se presta el servicio:

1. MATRIZ

Este informe tiene como objetivo proporcionar un resumen detallado de las actividades realizadas, incidencias, y mejoras implementadas durante el periodo.

2. RESÚMEN DE ACTIVIDADES

2.1. COBERTURA DEL SERVICIO

El servicio de seguridad se mantuvo cubierto en todos los sitios durante el periodo, con un total de 12 puestos de servicio para asegurar una vigilancia constante en las áreas asignadas, de acuerdo a la siguiente tabla:

SITIO	SERVICIO	24H	12H	TURNO	DIAS	CIUDAD
PLANTA DE PROCESAMIENTO	VIGILANCIA (CONTROL ACCESOS)	1	2	24 H	L-D	GUAYAQUIL
PLANTA DE PROCESAMIENTO	VIGILANCIA (JEFE DE GRUPO)	1	2	DIURNO	L-V	GUAYAQUIL
PLANTA DE PROCESAMIENTO	VIGILANCIA (PARQUEO)	1	2	DIURNO	L-V	GUAYAQUIL
GRANJA SAN CARLOS	VIGILANCIA	1	2	24 H	L-D	SAN CARLOS



PROTEMAXI
SEGURIDAD PRIVADA

INFORME
SISTEMA INTEGRADO DE GESTIÓN
FM-OPE-010 VER: 02 REV: 01 EMISSION: 20/AGO/2022

TOTAL GENERAL		4	8			
---------------	--	---	---	--	--	--

2.2. RONDAS DE VIGILANCIA:

Se realizaron un total de 15 rondas de vigilancia y 5 marcaciones QR, distribuidas entre los diferentes puntos de servicio. Estas rondas se llevaron a cabo en horarios aleatorios para maximizar la efectividad y minimizar los riesgos de incidentes.

SITIO	NO. DE RONDAS	NO. DE MARCACIONES
Granja Lomas	15	5
TOTAL GENERAL	15	5

2.3. CONTROL DE ACCESOS:

Se gestionaron un total de 14 registros de control de accesos de empleados, visitantes, proveedores, y clientes, en las distintas instalaciones ingresados en nuestro sistema PROTEAPP®. El proceso de control incluyó la verificación de identidades y la inspección de vehículos conforme a los procedimientos establecidos, garantizando el cumplimiento de las políticas de seguridad de Cliente 2.

SITIO	EMPLEADOS	VISITANTES	PROVEEDORES	CLIENTES	TOTAL
Granja San Carlos	1	2	1	1	5
Granja Bucay	1	1	1	1	4
Grafimpac S.A	2	1	1	1	5
TOTAL GENERAL	4	4	3	3	14

2.4. REPORTE DE CUSTODIAS

Durante el mes de FEBRERO se realizaron un total de 3 custodias de mercaderías en tránsito, asegurando el traslado seguro desde las diferentes granjas.

NO.	FECHA	GUIA NO.	PUNTO PARTIDA	PUNTO DE LLEGADA	CUSTODIOS	CONTENEDOR	PLACAS CAMIONES
1	0000-00-00	2024-07-31	LOMAS DE SARGENTILLO	SANTO DOMINGO	JOFFRE MONCERRATE - JUAN GUZMAN	2 CONTENEDORES	UBA-3340 - U4A-3266



PROTEMAXI
SEGURIDAD PRIVADA

INFORME
SISTEMA INTEGRADO DE GESTIÓN
FM-OPE-010 VER: 02 REV: 01 EMISION: 20/AGO/2022

2	0000-00-0 0	2024-08-0 1	LOMAS DE SARGENTILLO	PALLATANGA	CARLOS SUAREZ - JUAN CARLOS GONZALEZ	2 CONTENEDORE S	UBA-3340 - UAA-3266
3	0000-00-0 0	2024-08-0 4	SAN CARLOS	CAMAL GUAYAQUIL	MONCERRA TE JOFFRE - JAIME QUIÁONEZ	2 CONTENEDORE S	GSD-6924 - UBA-3340

3. NOVEDADES E INCIDENTES DURANTE EL PERIODO

3.1. INCIDENTES DE SEGURIDAD

A continuación, se detallan las incidencias de seguridad reportadas en el presente periodo:

- SITIO: LOCALIDAD29 - DETALLE: PARA SU INFORMACION; PERSONAL ATRASADO : 19:30 GEORGE CASTRO(PRENSA)-JONATHAN MEREJILDO(AREA BOLSO) AUT. SR. LUIS TUMBACO 19:30 DANNY PIGUAVE(MANTENIMIENTO) AUT. ING. DIONESIS PEREZ.
- SITIO: LOCALIDAD36 - DETALLE: VIERNES , 27 DE SEPTIEMBRE DEL 2024. SALUDOS CORDIALES . Informo siendo las 21:30 pm el generador de electricidad presenta novedades como que se ahoga y quiere apagarse le pongo a su conocimiento la situación que se está suscitando en el punto. Mediante llamadas y mensajes de Whatsapp se toma contacto con el Sr.ad. Andy Bajaña indicando que es una posible sobre carga que se está presentando y que proceda apagar 2 extractores de aire por galpón. Se procede según lo indicado pero al momento persiste la novedad. Se le informo a la central Protemaxi al Sr. de operaciones Segundo Constante y al jefe de seguridad interna de MOLINOS CHAMPIONS Ing. Daniel Pintado para q tengan conocimiento lo q está sucediendo en el sitio 3. Al momento siendo las 00:00 se apaga el generador quedando sin energia el sitio 3 y bajandose todas las cortinas de los galpones. Nuevamente se toma contacto con el jefe de seguridad interna de MOLINOS CHAMPIONS Ing. Daniel Pintado q indica le detalle toda la novedad sucita en el punto sobre la avería del generador de electricidad. Se toma contacto con el administrador Ing Andy Bajaña indicando q cuando llegue la energía gire la palanca para q haya energía en el sitio 3 indicando q el Sr Daniel Becerra viene a realizar levantamiento de cortinas y verificación de cajas de brekes y verificación de los motores de extractores de aire autorizando el ingreso por el sitio 1 y posterior al sitio 3. 00:24 am. Llega el Sr Daniel Becerra realiza la apertura en el sitio 3 y procede ir a las duchas para posterior realizar dicha labor indicada, realiza levantamientos de cortinas y verifica las cajas de brekes galpon por galpón y q todo este funcionando correctamente, realizada dicha labor procede nuevamente ir a las duchas para retirarse del punto. 01:33 am. Se procede abrir el candado para q salga el Sr Daniel Becerra y realiza el cierre sin novedad. ATT. HAROLD CASTRO AGENTE DE TURNO. GRANJA LOMAS - ALFA 3
- SITIO: LOCALIDAD29 - DETALLE: Siendo las 15:09 se procede a realizar la revisión al personal de comedor, dentro del bolso de la Colaboradora Fabiola Karina Lastra Arroyo se encuentra 1 funda con menestra, y a la sra, Rosa María Alejandro Palma 1 funda de arroz, se le procedió a comunicar la novedad al sr. Luis Ortiz de seguridad interna de Reysac mismo que autoriza la salida.
- SITIO: LOCALIDAD29 - DETALLE: Se entrega llaves de ofic Sr. Verá a Sr. Soledispa Kelvin
- SITIO: LOCALIDAD29 - DETALLE: Vehículo Volkswagen placa GTL 7985 .Solo Sale con Autorización Previa

3.2. NOVEDADES REPORTADAS EN PROTEAPP®

A lo largo de este periodo, se han identificado y reportado en nuestro sistema PROTEAPP® 15 novedades relevantes en todos los



PROTEMAXI
SEGURIDAD PRIVADA

INFORME
SISTEMA INTEGRADO DE GESTIÓN
FM-OPE-010 VER: 02 REV: 01 EMISIÓN: 20/AGO/2022

sitios, que destacan la importancia de nuestra gestión de vigilancia y seguridad, las cuales se resumen a continuación:

SITIO	INCIDENTES	HALLAZGOS	NOVEDADES DEL SITIO	TOTAL NOVEDADES REPORTADAS EN PROTEAPP
LOCALIDAD29	4	4	4	12
LOCALIDAD36	1	1	1	3
TOTAL GENERAL	5	5	5	15

3.3. CAMBIOS EN LA NOMINA DEL PERSONAL

En el siguiente cuadro se presenta el listado del personal que salió del proyecto en el periodo:

NO.	APELLIDOS NOMBRES	PUESTO	FECHA DE INGRESO	FECHA DE SALIDA
1	CASTRO SILVA JEFERSON RUSBEL	GRANJA LOMAS	2024-08-04	2024-08-19
2	JARAMILLO SOLORIZANO ALEJANDRO MICHAEL	GRANJA LOMAS	2024-05-22	2024-08-16
3	CEGARRA BRITO KEVIN MAURICIO	GRANJA LOMAS	2024-07-28	2024-08-27

4. ACCIONES CORRECTIVAS REPORTADAS POR EL CLIENTE

En este punto se detallan las solicitudes de acción correctiva (SAC's) reportadas por el cliente:

NO.	SITIO	FECHA DEL INCIDENTE	INCIDENTE	MEDIDA DE CONTROL
1	VIA A LA COSTA (BOTOTILLO)	2024-08-31	No se proporcionó alimentación a guardia en el sitio, en una hora adecuada. El supervisor arribó a las 11H40 al sitio cuando el personal ya se disponía a almorzar.	Concienciar al personal asignado la prioridad en la entrega de alimentación, cuando amerite.
2	VIA A LA COSTA (BOTOTILLO)	2024-08-21	El día 21/08/2024 se realizó 1 prueba poligráfica al señor VINICIO VLADIMIR TUFÍO TORO, chofer de la empresa RANSECORP.	N/A



5. VALORES AGREGADOS

Durante el mes de FEBRERO se proporcionaron los siguientes valores agregados solicitados por el departamento de seguridad física de Cliente 2 :

- 1) El día 21/08/2024 se realizó 1 prueba poligráfica al señor VINICIO VLADIMIR TUFIÑO TORO, chofer d

6. CONCLUSIONES Y RECOMENDACIONES

El servicio de seguridad proporcionado por PROTEMAXI durante FEBRERO cumplió con los requisitos esperados, garantizando la protección de las instalaciones de Cliente 2 en todos los puntos de servicio.

INCIDENCIA	FRECUENCIA	RECOMENDACION
Reducción de fotos tomar fotos	Baja	Revisar procedimientos trimestralmente.
Novedad en rea de pesaje	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
Daño en sistema eléctrico	Baja	Revisar procedimientos trimestralmente.
Transportistas que no respeta normas de seguridad	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
Fallo de energía en puesto de guardia	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
Daño reflectores galpón 4	Baja	Revisar procedimientos trimestralmente.
CONATO DE INCENDIO	Baja	Revisar procedimientos trimestralmente.

ELABORADO POR:	REVISADO POR:	REPORTADO A:
Jefe de Operaciones	Administrador del Contrato	Jefe de Seguridad Física
PROTEMAXI C. LTDA	PROTEMAXI C. LTDA	MOCHASA S.A.

Precisión del modelo de Machine Learning

Precisión del modelo está orientada a la identificación de incidencias de un cliente y periodo de acuerdo con la frecuencia en que ocurren, esto permite identificar patrones sobre incidencias que tengas el mismo comportamiento y sean repetitivas

para dar una recomendación de acuerdo con la frecuencia que puede ser baja, media o alta. Para el cálculo de la precisión se utiliza la siguiente formula:

$$Presición = \left(\frac{Numero\ de\ patrones\ correctos}{Número\ total\ de\ patrones\ reportados\ por\ el\ modelo} \right) * 100$$

El número de patrones correctos es la cantidad que representan el total de frecuencias reales que han sido validadas y encontradas de manera manual dentro de conjunto de datos proporcionado mientras que el número total de patrones reportados por el modelo proviene del total de frecuencias que el modelo ha logrado clasificar dentro del mismo conjunto de datos.

A continuación, se presentan las recomendaciones que da el modelo de machine Learning dentro de la sección 6 del informe generado que se presentó anteriormente, dentro de esta sección se encuentran las conclusiones y recomendaciones:

Figura 58
Recomendaciones Cliente 1

6. CONCLUSIONES Y RECOMENDACIONES

El servicio de seguridad proporcionado por PROTEMAXI durante JUNIO cumplió con los requisitos esperados, garantizando la protección de las instalaciones de Cliente 1 en todos los puntos de servicio.

INCIDENCIA	FRECUENCIA	RECOMENDACION
PERSONAL ATRASADO	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
Reporte desde Bit? cora	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
Salida de oficial	Media	Monitorear el incidente y establecer medidas preventivas mensuales.

Figura 59
Recomendaciones Cliente 4

6. CONCLUSIONES Y RECOMENDACIONES

El servicio de seguridad proporcionado por PROTEMAXI durante ABRIL cumplió con los requisitos esperados, garantizando la protección de las instalaciones de Cliente 4 en todos los puntos de servicio.

INCIDENCIA	FRECUENCIA	RECOMENDACION
ACCIDENTE LABORAL	Baja	Revisar procedimientos trimestralmente.

Figura 60
Recomendaciones Cliente 2

6. CONCLUSIONES Y RECOMENDACIONES

El servicio de seguridad proporcionado por PROTEMAXI durante FEBRERO cumplió con los requisitos esperados, garantizando la protección de las instalaciones de Cliente 2 en todos los puntos de servicio.

INCIDENCIA	FRECUENCIA	RECOMENDACION
Reducción de fotos tomar fotos	Baja	Revisar procedimientos trimestralmente.
Novedad en rea de pesaje	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
Daño en sistema eléctrico	Baja	Revisar procedimientos trimestralmente.
Transportistas que no respeta normas de seguridad	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
Fallo de energía en puesto de guardia	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
Daño reflectores galpón 4	Baja	Revisar procedimientos trimestralmente.
CONATO DE INCENDIO	Baja	Revisar procedimientos trimestralmente.

Figura 61
Recomendaciones Cliente 3

6. CONCLUSIONES Y RECOMENDACIONES

El servicio de seguridad proporcionado por PROTEMAXI durante MARZO cumplió con los requisitos esperados, garantizando la protección de las instalaciones de Cliente 3 en todos los puntos de servicio.

INCIDENCIA	FRECUENCIA	RECOMENDACION
PRUEBA	Baja	Revisar procedimientos trimestralmente.
FILTACION DE COMBUSTIBLE	Baja	Revisar procedimientos trimestralmente.
Garita con olor a rata.	Baja	Revisar procedimientos trimestralmente.
CON EL PANEL DE LA ALARMA, PARA DESACTIVACION DE LA MISMA	Baja	Revisar procedimientos trimestralmente.
NOVEDAD DE TRANSPORTE S	Baja	Revisar procedimientos trimestralmente.
Control de acceso	Baja	Revisar procedimientos trimestralmente.
NOVEDAD EN LA SALIDA DEL JEFE DE CONTROL INTERNO	Baja	Revisar procedimientos trimestralmente.
novedad en el ascensor de planta	Baja	Revisar procedimientos trimestralmente.
PARED CAIDA CAIDA	Baja	Revisar procedimientos trimestralmente.
NOVEDAD CON TRANSPORTISTA	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
NOVEDAD CON TRANSPORTE ROBISON	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
ORIFICIO EN PARED	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
bloqueo de elevador	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
FALTA DE ILUMINACION	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
A CON EL ACCESO BIOMETRICO DEL BA? O DE COLABORADOR	Media	Monitorear el incidente y establecer medidas preventivas mensuales.
NOVEDAD	Media	Monitorear el incidente y establecer medidas preventivas mensuales.

Para el cálculo de la precisión se toma en cuenta una incidencia de cada uno de los Clientes son su respectiva frecuencia que ha sido generado por el modelo y se la comparaba con la frecuencia real sacada de un análisis manual realizado; esta comparación sirve para verificar si la frecuencia está acorde a la realidad y ver si la recomendación proporcionada está acorde. A continuación, se presenta una tabla comparativa:

Tabla 7
Comparación frecuencia

Cliente	Mes	Título de Incidencia	Frecuencia Real	Frecuencia del Modelo
Cliente 1	JUNIO	Personal atrasado	Alta	Media
Cliente 2	FEBRERO	Novedad en rea de pesaje	Media	Media

Cliente	Mes	Título de Incidencia	Frecuencia Real	Frecuencia del Modelo
Cliente 3	MARZO	Personas sin identificar	Media	Media
Cliente 4	ABRIL	Accidente laboral	Baja	Baja
Cliente 5	JULIO	Novedad en rea de pesaje	Media	Media

Nota: Se presentan cada uno de los clientes de acuerdo con su periodo y la frecuencia que ha determinado el modelo de machine Learning con la información de la frecuencia real encontrada a mano

El modelo de Machine Learning se prueba en 4 clientes donde la mayoría tienen una relación entre la realidad y las respuestas que arroja el modelo, sin embargo, el cliente 1 ha fallado; las frecuencias lanzadas por el modelo del cliente 2, 3 y 4 fueron identificadas correctamente. A continuación, se indica el cálculo de la precisión en base a los datos descritos en la tabla 6:

$$Precisión = \left(\frac{4}{5}\right) * 100 = 80\%$$

El modelo de Machine Learning tiene una precisión del 80% en cuanto a las recomendaciones que puede dar cuando las contrastamos de manera frecuencial.

Tiempo de generación de informe

De acuerdo con lo mencionado por la empresa durante el levantamiento de información se destaca que manualmente la empresa demora 2 horas en realizar la creación del documento de informe por periodo mientras que a través del interfaz web del sistema de generación de informes, los informes se generan al instante en menos de 5 segundos.

Conclusiones

Tomando en consideración la construcción definida del prototipo y en base a los resultados obtenidos tanto por parte del sistema de generación de informes y el modelo de Machine Learning se da el cumplimiento de los objetivos generales, específicos y alcance definidos para el presente proyecto.

Se da el cumplimiento con la pregunta de hipótesis donde las variables de investigación muestran que el prototipo es capaz de analizar con un 80% de precisión en la clasificación de patrones de frecuencia en la data proporcionada de novedades, automatiza la generación de informes estandarizados reduciendo el tiempo en un 99,96% pasando de 2 horas a 3 segundos utilizando el mismo volumen de datos mejorando la eficiencia y calidad,

La implementación de un Modelo de Machine Learning como una componente innovador y moderno al momento de realizar una automatización en un proceso de generación de informes mensuales genera una ventaja en la toma de decisiones al obtener resultados que provienen de más de 100 registros de incidencias; el sistema o aplicativo web cumple con los estándares y funcionalidades requeridas a nivel técnico y funcional lo que permite ponerse a la altura de las necesidades planteadas en cuanto a una automatización inteligente.

El modelo de Machine Learning es capaz de hallar correlaciones entre variables como la fecha, título, descripción y tipo de incidentes de cada cliente, al hallar una correlación el modelo puede interpretar mejores patrones de frecuencias al saber que variables se relacionan entre sí; la implementación del algoritmo predictivo *Gradient Boosting Regressor* permite estimar las probabilidades en la aparición de nuevos incidentes en un futuro marcando su repetición; al tener identificados los patrones basados en un comportamiento particular junto con una frecuencia de ocurrencia el Modelo de Machine Learning genera exitosamente recomendaciones estandarizadas de acuerdo a la clasificación que tengas las incidencias analizada.

El sistema de generación de informes organiza la data de incidencias de acuerdo con el cliente y mes esto permite filtrar los reportes de cada cliente por periodo, esto permite genera informes consolidados de incidencias de acuerdo a las necesidades del usuario; Dentro del informe se integra los resultados del proceso de

análisis del modelo de Machine Learning que reemplaza tareas manuales que requieran un análisis sobre grandes volúmenes sobre la data de incidencias, así aumentando el nivel de precisión y eficiencia de dicho proceso optimizando la asignación de recursos.

La construcción de la base de datos hizo posible la conexión lógica entre el sistema de generación de reportes y el modelo de Machine Learning, esto permite que la data que proporcionada por la empresa de seguridad pueda ser manejada tanto para realizar el llenado directo de información dentro de ciertas secciones del informe y el análisis para después con esos resultados generar una puesta de valor dentro del informe; la base de datos sirve un medio de intercambio de información al interactuar con el modelo de Machine Learning lo que permite guardar los resultados del modelo en la base creando tablas donde la esa información analizada después es consultada por el sistema de generación de reportes para completar la secciones de recomendaciones.

La arquitectura del sistema se adecua a los requisitos técnicos mencionados por la empresa de seguridad, el uso del Framework de Laravel construido en PHP permite tener protocolos de seguridad ya integrados por defecto además de manejar peticiones utilizando APIRest que sirven para las interacciones funcionales del usuario y la carga de la interfaz de usuario; el uso de Python como una opción sencilla pero al mismo tiempo versátil para la creación de modelos de Machine Learning y poder realizar una conexión de manera dócil con la base de datos MariaDB que permite la obtención y envío de data por parte de ambos. Al utilizar el Framework mencionado se facilita la posible integración al sistema actual que maneja la empresa debido a que es el mismo esto permite que exista compatibilidad.

Recomendaciones

Se debe considerar que el servidor que se utiliza es de pocas prestaciones y es local por lo que es un sistema que se construyó en un entorno de desarrollo y practica por lo tanto tiene limitaciones; es importante entender que en caso de quererlo pasar a producción el prototipo del sistema ya se requeriría un servidor más robusto y con mayores prestaciones.

Para una mayor escalabilidad del prototipo se recomienda mejorar la infraestructura de la empresa porque si en un futuro se desea analizar otro tipo de informacion o variables que no se hayan contemplado en el proyecto se deberá mejorar componentes de hardware u optar por un software en la nube para un mejor manejo de datos.

Se recomienda a la empresa de seguridad mantener un registro de datos de manera completa al momento en que se realiza la obtención de los datos dentro del prototipo del sistema y en el modelo de Machine Learning con el fin mejorar la precisión o la capacidad de análisis en la data, esto implica que debe existir consistencia y coherencia en los datos.

Referencias Bibliográficas

- Alay, J. I. G., & Sevillano, R. P. C. (2022). Evolución de los sistemas de lenguaje de programación a lo largo de la historia. *E-IDEA Journal of Engineering Science*, 4(10), 14–26.
- Alvarez-Intriago, V., & Torres-Samaniego, F. (2019). Impacto de un Sistema Web para Optimizar Insumos en Negocio de Comida. *INVESTIGATIO*, 12, Article 12.
- Amazon Web Services, Inc. (2023). *¿En qué consiste la automatización inteligente?: Explicación sobre la automatización inteligente de procesos: AWS*. Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/intelligent-automation/>
- Amazon Web Services, Inc. (2024a). *¿Qué es el machine learning? - Explicación sobre el machine learning empresarial - AWS*. Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/machine-learning/>
- Amazon Web Services, Inc. (2024b). *¿Qué es un IDE? - Explicación de los entornos de desarrollo integrado - AWS*. Amazon Web Services, Inc. <https://aws.amazon.com/es/what-is/ide/>
- Báez, M., Barredo, S., Stinco, L., & Merlino, H. (s/f). *ENTORNO DE GENERACIÓN DE DATOS SINTÉTICOS*.
- Barrera, J. A., & Barrera, S. A. (2020). *Metodologías para el desarrollo de Proyectos*.
- Barrera, J. R. A., Torres, Y. E. S., & Alarcón, L. A. V. (2024). La Tecnología y su Impacto en la Gestión de Procesos y Estrategias de Automatización. *Ciencia Latina Revista Científica Multidisciplinar*, 8(4), Article 4. https://doi.org/10.37811/cl_rcm.v8i4.12822
- Barrezueta, H. D. P. (2021). *LEY ORGÁNICA DE PROTECCIÓN DE DATOS PERSONALES*. 38.

Beynon-Davies, P. (2018). *Sistemas de bases de datos*. Reverté.

<https://books.google.es/books?hl=es&lr=&id=XjbeDwAAQBAJ&oi=fnd&pg=PR5&dq=base+de+datos&ots=DJyXENFQEV&sig=sLiRSdi8YAUUsS8X61d67WJCgzBU>

Castañeda, B. (2023). *Vista de Impacto de un Sistema web de Control presupuestal para organizaciones*. Revista Ciencia y Tecnología.

<https://revistas.unitru.edu.pe/index.php/PGM/article/view/5767/5813>

Challenger-Pérez, I., Díaz-Ricardo, Y., & Becerra-García, R. A. (2014). El lenguaje de programación Python. *Ciencias Holguín*, 20(2), 1–13.

Cuadrado, D. A. M., & Angulo, D. P. S. (s/f). *UTILIZACIÓN DEL MACHINE LEARNING EN LA INDUSTRIA 4.0*.

Gamo, D. M. (s/f). *Autor(a): Sergio Arroni del Riego*.

GitLab. (2023). *¿Qué es el control de versiones?*

<https://about.gitlab.com/es/topics/version-control/>

Hernández, R. (2018). METODOLOGÍA DE LA INVESTIGACIÓN: LAS RUTAS CUANTITATIVA, CUALITATIVA Y MIXTA. En *METODOLOGÍA DE LA INVESTIGACIÓN: LAS RUTAS CUANTITATIVA, CUALITATIVA Y MIXTA* (1a ed., p. 753).

https://d1wqtxts1xzle7.cloudfront.net/64591365/Metodolog%C3%ADa_de_la_investigaci%C3%B3n._Rutas_cuantitativa__cualitativa_y_mixta-libre.pdf?1601784484=&response-content-disposition=inline%3B+filename%3DMETODOLOGIA_DE_LA_INVESTIGACION_LAS_RUTA.pdf&Expires=1734732463&Signature=A86zLrq26YuaDmCL4PFAm0KZ3DrkOXIRz9Sx1cH~HENRNxAWXufNOshEtuUnaxpajIDt70JNro-3Yg7ehAU38RA9fSyajLeg2krZ0-

q0gTLpT4TVKcbfItlxV~KdPTjEyzNTZJs~pSMJJuqG458vMtCK6X1ZPjtFmAx
HSEkhBzemLQqmFX08EHpACdFOCOTDjpcHPH6bnDXo~JriXrkD5Ce2Eqhp
V9sxkpSqeVOfSx8UhtwxM7zl~1SrAm5PTm-
lrS6eJ7~otoSqSqyPKsGuePbO3f8FWxnJx8nor6ipreBLxiyRCLdB-kGPBp-
81ut1~So4WM0wj1paChjgC-OBHQ__&Key-Pair-
Id=APKAJLOHF5GGSLRBV4ZA

International Business Machines Corporation. (2021, octubre 9). *¿Qué es la automatización inteligente?* <https://www.ibm.com/mx-es/topics/intelligent-automation>

IBM Business Process Manager 8.5.7. (2021, marzo 4).

<https://www.ibm.com/docs/es/bpm/8.5.7?topic=bpa-creating-process>

Maheswari, K. G., Siva, C., & Nalinipriya, G. (2023). Optimal cluster based feature selection for intrusion detection system in web and cloud computing environment using hybrid teacher learning optimization enables deep recurrent neural network. *Computer Communications*, 202, 145–153.
<https://doi.org/10.1016/j.comcom.2023.02.003>

Microsoft. (2023, octubre 5). *¿Qué es el control de versiones? - Azure DevOps.*

<https://learn.microsoft.com/es-es/devops/develop/git/what-is-version-control>

Mirjalili, V., & Raschka, S. (2020). *Python machine learning.* Marcombo.

<https://books.google.es/books?hl=es&lr=&id=5EtOEAAAQBAJ&oi=fnd&pg=PT5&dq=python&ots=erHZRwXDO5&sig=d0F8KRsnxkP2jMyWBqJWLbFPoM>

Nieto, N. T. E. (2018). *TIPOS DE INVESTIGACIÓN.*

Olivera, L. de la C. D., & Alonso, L. M. D. (2021). Modelos de Desarrollo de Software. *Revista Cubana de Ciencias Informáticas*, 15(1), 37–51.

Oracle. (2024). *¿Qué es el aprendizaje automático?*

<https://www.oracle.com/co/artificial-intelligence/machine-learning/what-is-machine-learning/>

Ortega, D. J. O., Asesor, O. R., & Pérez, M. A. B. (2019). Bootstrap y Laravel, herramientas para el desarrollo de aplicaciones web. *UNIVERSIDAD POLITÉCNICA DE SINALOA, Mexico.*

<http://repositorio.upsin.edu.mx/Fragmentos/tesinas/142016030030OvandoOrtegaDenzelJavier10843.pdf>

Pérez Ibarra, S. G., Quispe, J. R., Mullicundo, F. F., & Lamas, D. A. (2021).

Herramientas y tecnologías para el desarrollo web desde el FrontEnd al BackEnd. *XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja).*

<https://sedici.unlp.edu.ar/handle/10915/120476>

Ramírez, D. H., & Cárdenas, J. M. (s/f). *EL MACHINE LEARNING A TRAVÉS DE LOS TIEMPOS, Y LOS APORTES A LA HUMANIDAD.*

Red Hat. (2023, julio 31). *¿Qué es y para qué sirve un IDE?*

<https://www.redhat.com/es/topics/middleware/what-is-ide>

Santillán, Á. G. (2015). Lucha de clases. Comparativa de rendimiento de clases FPDF vs CLASSEZPDF mediante ejemplos ilustrativos. *Revista*

Iberoamericana para la Investigación y el Desarrollo Educativo ISSN: 2007-2619, 12. <http://www.1->

[11.ride.org.mx/index.php/RIDSESECUNDARIO/article/view/733](http://www.1-11.ride.org.mx/index.php/RIDSESECUNDARIO/article/view/733)

Santillán, L. A. C., Ginestà, M. G., & Mora, Ó. P. (2014). Bases de datos en MySQL. *Universitat oberta de Catalunya.*

https://www.academia.edu/download/54167148/MYSQL_MANUAL.pdf

Serrahima, Á. (2022). *AVANCES Y DESAFÍOS DE LA INTELIGENCIA ARTIFICIAL*

[Escuela técnica superior de ingeniería (ICAI)]. retrieve.pdf.

<https://repositorio.comillas.edu/rest/bitstreams/507551/retrieve>

Torres-Salinas, D., & Arroyo-Machado, W. (2022). APIs en contextos bibliométricos:

Introducción básica y corpus exhaustivo. *Anuario ThinkEPI*, 16.

<https://thinkepi.scimagoepi.com/index.php/ThinkEPI/article/view/91540>

What Is Machine Learning (ML)? | IBM. (2021, septiembre 22).

<https://www.ibm.com/topics/machine-learning>

Zhang, C. (Abigail). (2019). Intelligent Process Automation in Audit. *Journal of*

Emerging Technologies in Accounting, 16(2), 69–88.

<https://doi.org/10.2308/jeta-52653>

Anexos

Anexo 1. Preguntas de la entrevista de levantamiento de información

1. ¿Cómo se realiza el proceso de generación de informes?
2. ¿Quién es el responsable de realizar el proceso de generación de informes?
3. ¿Cuánto suele durar el proceso de generación de informes?
4. ¿Qué tan importante es el proceso de generación de informes para la toma de decisiones de la empresa?
5. ¿Se utiliza alguna técnica para analizar los datos e información que va dentro del informe?
6. ¿Qué es lo que se pretende conseguir con la automatización de este proceso y la integración con inteligencia artificial?
7. ¿Qué se pretende analizar dentro de la información o datos que se usan para la generación del informe?

Anexo 2. Levantamiento de información



Levantamiento de
información PROTEM,

Anexo 3. Archivo Informe generado



mi_reporte (3).pdf



DECLARACIÓN Y AUTORIZACIÓN

Nosotros, **Pazos Celi, Andrés Sebastián**, con C.C: # **0924444805** y **Vega Doylet, Priscilla Rebeca** con C.C: #**0931482079** autores del trabajo de titulación: “**Desarrollo de un prototipo de sistema web basado en machine learning para la identificación de tendencias y análisis de incidentes de la empresa de seguridad privada PROTEMAXI.**” previo a la obtención del título de **Ingeniero en Ciencias de la Computación** en la Universidad Católica de Santiago de Guayaquil.

1.- Declaro tener pleno conocimiento de la obligación que tienen las instituciones de educación superior, de conformidad con el Artículo 144 de la Ley Orgánica de Educación Superior, de entregar a la SENESCYT en formato digital una copia del referido trabajo de titulación para que sea integrado al Sistema Nacional de Información de la Educación Superior del Ecuador para su difusión pública respetando los derechos de autor.

2.- Autorizo a la SENESCYT a tener una copia del referido trabajo de titulación, con el propósito de generar un repositorio que democratice la información, respetando las políticas de propiedad intelectual vigentes.

Guayaquil, 17 de febrero de 2025

Nombre: **Pazos Celi, Andrés Sebastián**

C.C: **0924444805**

Nombre: **Vega Doylet, Priscilla Rebeca**

C.C: **0931482079**

REPOSITORIO NACIONAL EN CIENCIA Y TECNOLOGÍA

FICHA DE REGISTRO DE TESIS/TRABAJO DE TITULACIÓN

TEMA Y SUBTEMA:	Desarrollo de un prototipo de sistema web basado en machine learning para la identificación de tendencias y análisis de incidentes de la empresa de seguridad privada PROTEMAXI.		
AUTOR(ES)	Pazos Celi, Andrés Sebastián Vega Doylet, Priscilla Rebeca		
REVISOR(ES)/TUTOR(ES)	Yong Yong, Byron Severo		
INSTITUCIÓN:	Universidad Católica de Santiago de Guayaquil		
FACULTAD:	Ingeniería		
CARRERA:	Ingeniería en Ciencias de la Computación		
TÍTULO OBTENIDO:	Ingeniero en Ciencias de la Computación		
FECHA DE PUBLICACIÓN:	17 de febrero de 2025	No. DE PÁGINAS:	104 p.
ÁREAS TEMÁTICAS:	Inteligencia artificial, Automatización, Protección de datos, Análisis de datos,		
PALABRAS CLAVES/ KEYWORDS:	Aprendizaje Automático, Automatización inteligente, inteligencia artificial, prototipo, Seguridad, Python, Laravel, Patrones.		
RESUMEN/ABSTRACT:	<p>Este trabajo es la construcción e investigación de un prototipo de sistema web basado en Machine Learning, la automatización inteligente es una nueva forma de potenciar la automatización de procesos tradicionales con la Inteligencia artificial, PROTEMAXI una empresa que brinda servicio de seguridad a sus clientes dentro de la ciudad de Guayaquil tiene la necesidad de agilizar el proceso de generación de informes de las novedades reportadas mensualmente a la vez que la analiza para obtener información valiosa, comúnmente este proceso se hace de forma manual pero es poco preciso y puede contener errores; por este motivo el presente proyecto se basa en la construcción del prototipo del sistema que genera reportes automatizados de acuerdo al cliente y la fecha es decir el periodo integrando la tecnología del aprendizaje automático; La construcción del sistema fue hecha con los requerimientos tecnológicos que la empresa solicito como el Framework de Laravel para la interfaz web y para el modelo de Machine Learning Python. Desde las diferentes reuniones de levantamiento de información hasta el desarrollo del modelo de Machine Learning se utiliza una metodología de software en cascada que permite trabajar de manera más directa y sin petición a cambios no establecidos. Los objetivos del proyecto se basan en la implementación del empleo de técnicas de Machine Learning para generar resultados en base a los patrones encontrados, la elaboración del sistema web, la creación de una base de datos y el uso de las tecnologías requeridas por la empresa por termino de compatibilidad. Para finalizar, el proyecto concluye con la generación de informes de acuerdo con el cliente y al mes seleccionado donde se muestran datos precargados y los resultados del análisis del modelo de Machine Learning, gracias a esta implementación se logra reducir el tiempo de generación del informe y se obtienen patrones de frecuencia sobre las novedades ocurridas y que recomendaciones se pueden tomar al respecto.</p>		
ADJUNTO PDF:	<input checked="" type="checkbox"/> SI	<input type="checkbox"/> NO	
CONTACTO CON AUTOR/ES:	Andrés Sebastián Pazos Celi Teléfono: +593-987-147345 Priscilla Rebeca Vega Doylet Teléfono: +593-982-567167	E-mail: andrespazos1@hotmail.com rebecavegad@gmail.com	
CONTACTO CON LA INSTITUCIÓN (COORDINADOR DEL PROCESO UTE):	Toala Quimí, Edison José Teléfono: +593-990-976776 E-mail: edison.toala@cu.ucsg.edu.ec		
SECCIÓN PARA USO DE BIBLIOTECA			
Nº. DE REGISTRO (en base a datos):			
Nº. DE CLASIFICACIÓN:			
DIRECCIÓN URL (tesis en la web):			