



Automatización de procesos de análisis de información y generación de reportes mediante la implementación de tableros de visualización de datos con la herramienta Power BI para el Vicerrectorado de Docencia de la Universidad de las Fuerzas Armadas ESPE.

Pazmiño Gordon, Bryan Alexander y Tituaña Toapanta, Jhonatan David

Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Trabajo de integración curricular, previo a la obtención del título de Ingeniero en Tecnologías en Información.

Ing. López López Andrea Margarita, MSc.

06 de marzo de 2024

Resultados de la herramienta de Plagio



Plagiarism and AI Content Detection Report

Trabajo UIC Pazmiño - Tituaña.docx

Scan details

Scan time: February 28th, 2024 at 19:34 UTC
 Total Pages: 58
 Total Words: 14287

Plagiarism Detection



AI Content Detection



Plagiarism Results: (34)

Implementación de un BI, que permita analizar la data del CRM Zendesk, ... 0.8%

<https://www.wiki.com/pt/Implementacion-de-un-BI-que-permita-analizar-la-data-del-CRM-Zendesk-...>

Toggle navigation es | en | pt | fr Presentation Countries Institutions Get involved Toggle navigatio...

Implementación de un BI, que permita analizar la data del CRM Zendesk, ... 0.8%

<https://www.wiki.com/pt/Implementacion-de-un-BI-que-permita-analizar-la-data-del-CRM-Zendesk-...>

Iniciar sesiónCorreo electrónicoContraseña Iniciar sesión ¿Nuevo Usuario? Pulse aquí para registrarse¿Has olvidado tu contraseña...

Power BI: la herramienta imprescindible para el análisis de datos 0.7%

<https://www.wiki.com/pt/Power-BI-la-herramienta-imprescindible-para-el-analisis-de-datos>

Ir al contenido Sobre mí Blo...

Procesos ETL | Interactive Chaos 0.7%

<https://www.wiki.com/pt/Procesos-ETL-Interactive-Chaos>

Pasar al contenido principal ...

Certified by
Copleaks

About this report
<https://www.copleaks.com>

[copleaks.com](https://www.copleaks.com)
in f @



ANDREA MARGARITA
LOPEZ LOPEZ

Ing. López López, Andrea Margarita, Msc.



Departamento de Ciencias de la Computación
Carrera de Tecnologías de la Información

Certificación

Certifico que el trabajo de integración curricular: **"Automatización de procesos de análisis de información y generación de reportes mediante la implementación de tableros de visualización de datos con la herramienta Power BI para el Vicerrectorado de Docencia de la Universidad de las Fuerzas Armadas ESPE"** fue realizado por los señores **Pazmiño Gordon, Bryan Alexander y Tituaña Toapanta Jhonatan David** el mismo que cumple con los requisitos legales, teóricos, científicos, técnicos y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, además fue revisado y analizada en su totalidad por la herramienta de prevención y/o verificación de similitud de contenidos; razón por la cual me permito acreditar y autorizar para que se lo sustente públicamente.

Sangolquí, 06 de marzo de 2024



Ing. López López, Andrea Margarita, Msc.

C. C.: 1600687733



Departamento de Ciencias de la Computación

Carrera de Tecnologías de la Información

Responsabilidad de Autoría

Nosotros, **Pazmiño Gordon, Bryan Alexander**, con cédula de ciudadanía n° **1725536732** y **Tituaña Toapanta, Jhonatan David**, con cédula de ciudadanía n° **1752733566**, declaramos que el contenido, ideas y criterios del trabajo de integración curricular: **"Automatización de procesos de análisis de información y generación de reportes mediante la implementación de tableros de visualización de datos con la herramienta Power BI para el Vicerrectorado de Docencia de la Universidad de las Fuerzas Armadas ESPE"** es de nuestra autoría y responsabilidad, cumpliendo con los requisitos legales, teóricos, científicos, técnicos, y metodológicos establecidos por la Universidad de las Fuerzas Armadas ESPE, respetando los derechos intelectuales de terceros y referenciando las citas bibliográficas.

Sangolquí, 06 de marzo de 2024

Pazmiño Gordon, Bryan Alexander

C. C.: 1725536732

Tituaña Toapanta, Jhonatan David

C. C.: 1752733566



Departamento de Ciencias de la Computación
Carrera de Tecnologías de la Información

Autorización de Publicación

Nosotros, **Pazmiño Gordon, Bryan Alexander**, con cédula de ciudadanía n° **1725536732** y **Tituaña Toapanta, Jhonatan David**, con cédula de ciudadanía n° **1752733566**, autorizamos a la Universidad de las Fuerzas Armadas ESPE publicar el trabajo de integración curricular: **"Automatización de procesos de análisis de información y generación de reportes mediante la implementación de tableros de visualización de datos con la herramienta Power BI para el Vicerrectorado de Docencia de la Universidad de las Fuerzas Armadas ESPE"** en el Repositorio Institucional, cuyo contenido, ideas y criterios son de nuestra responsabilidad.

Sangolquí, 06 de marzo de 2024

Pazmiño Gordon, Bryan Alexander

C. C.: 1725536732

Tituaña Toapanta, Jhonatan David

C. C.: 1752733566

Dedicatoria

A mi amado Dios, compañero de cada paso, en los momentos de incertidumbre y oscuridad has sido la luz que ha iluminado mi camino, guiándome con amor y sabiduría. Siempre te encuentro en las buenas personas que se cruzan en mi camino, en cada gesto de bondad y actos de generosidad te has manifestado a través de ellos, recordándome tu amor incondicional y tu misericordia infinita.

A mi querida Mamá Rosita, aunque ya no esté presente en mi vida, siempre llevo conmigo todos los valores y principios que supo transmitirme, como el regalo más grande. Dondequiera que se encuentre, sé que su amor y sabiduría continúan guiándome en cada paso que doy.

A la única persona que estuvo a mi lado en mis inviernos y tempestades, y que no dejó de amarme ni por un solo segundo: mi amada Madre. A pesar de que el mundo siempre me decía que no, ella, con una sonrisa me aseguraba que sí. Gracias por enseñarme a vivir la vida de manera tan única, por mostrarme que esta vida es para los valientes, para los que se atreven a soñar. Gracias por siempre estar ahí.

Pazmiño Gordon, Bryan Alexander

Agradecimientos

Agradezco profundamente a mi familia por estar siempre presente en cada paso que doy, motivándome a siempre hacer las cosas de la mejor manera posible. Han sido quienes me inculcaron la convicción de que el camino hacia el éxito se construye con trabajo honesto, un carácter bien definido y, sobre todo, entregando compromiso y amor a cada paso que damos. Sus enseñanzas son el faro que guía mis acciones y me impulsan para alcanzar mis metas.

A mis queridos hermanos, Mateo y Andrés, quienes desde el momento en que llegaron al mundo se convirtieron en mi mayor motivación. Su presencia en cada día de mi vida me ha inspirado a esforzarme para que puedan seguir por un camino lleno de oportunidades y valores. Su influencia ha sido un recordatorio constante de la importancia de hacer las cosas de la mejor manera posible.

A Iván, quien desde mi corta edad ha sido mi guía y mentor, infundiéndome invaluables valores y moldeando en mi un carácter sereno. Gracias a sus enseñanzas, he llegado a comprender que en la vida todo requiere un gran esfuerzo y que muchas de las ocasiones, la constancia puede vencer al talento.

Pazmiño Gordon, Bryan Alexander

Dedicatoria

A Dios, quien ha sido mi guía constante y mi fortaleza inquebrantable a lo largo de este arduo camino académico, que ha iluminado cada paso que doy con su sabiduría ayudándome a levantar una y otra vez para poder llegar hasta donde.

A mis padres, cuyo amor incondicional, apoyo constante y sacrificio desinteresado han sido el pilar fundamental de mi vida. Su sacrificio y dedicación han hecho posible este logro, han sido mi mayor inspiración.

A mi hermano, la persona más importante en mi vida, mi amigo y mi eterno motivador. Mi motor detrás de cada logro y el recordatorio constante de que no hay límites para lo que puedo alcanzar.

A mis amigos y a todas las personas que, de una u otra manera, han contribuido a mi crecimiento académico y personal, brindándome su apoyo, consejo y aliento en cada etapa de este camino.

Tituaña Toapanta, Jhonatan David

Agradecimientos

Quisiera expresar mi más profundo agradecimiento a todas las personas que han sido parte fundamental de este arduo camino. En primer lugar, a mi familia, a mi madre, Veronica, cuyo amor incondicional, apoyo constante y sabias enseñanzas han sido la luz que me ha guiado en cada paso de mi vida. Su ejemplo de bondad y sacrificio han moldeado mi carácter, gracias por ser mi inspiración y por guiarme con tu ejemplo hacia la mejor versión de mí mismo. A mi padre, Edison, por su inquebrantable apoyo y sus sabios consejos, que siempre han sido un pilar fundamental en mi desarrollo personal y académico. A mi amado hermano Santiago, quien ha sido mi compañero de vida y mi mayor confidente. No puedo imaginar un camino sin su presencia y su aliento constante.

A mi tía Mercedes, que ha sido como una segunda madre para mí, brindándome su amor y apoyo en cada momento importante de mi vida. A mis abuelitos Pedro, Ester, Luis, Perla, Segundo, Rosa y Cecilia, cuyo cariño y amor incondicional han sido mi inspiración constante. Siempre estaré agradecido por los valores que me transmitieron y por el amor que han brindado. A mis primos y tíos, por su compañía y sabios consejos a lo largo de los años, que han sido un sostén en los momentos difíciles y una fuente de alegría en los momentos felices.

A mis amigos, quienes han compartido conmigo risas, lágrimas y momentos inolvidables. Su amistad y compañía han sido invaluable en este viaje llamado vida.

Y finalmente, a todas las personas que han estado a mi lado en mis momentos más difíciles y han sido la luz que ha iluminado mi camino. Su presencia ha hecho posible este momento y les estaré eternamente agradecido.

Tituaña Toapanta, Jhonatan David

Índice de Contenidos

Dedicatoria.....	6
Agradecimientos	7
Dedicatoria.....	8
Agradecimientos	9
Resumen	16
Abstract.....	17
Capítulo I.....	18
Introducción.....	18
Antecedentes.....	18
Planteamiento del Problema	18
Justificación.....	19
Objetivo.....	20
Objetivo General.....	20
Objetivos Específicos	20
Alcance.....	21
Hipótesis	21
Capítulo II.....	22
Marco Metodológico	22
Enfoque Mixto	22
Triangulación de Datos	22
Investigación Descriptiva	22
Diseño de Campo.....	23
Enfoque Cuantitativo	23

	11
Muestra o Población de Estudio	23
Estado del arte	23
Planteamiento de la Revisión Sistemática	23
Título	23
Introducción.....	24
Preguntas de Investigación	24
Criterios de Inclusión y Exclusión.....	24
Búsqueda y Selección de Estudios	24
Análisis de Datos y Síntesis:	25
Conclusiones:	25
Elaboración y resultados del Estado de Arte.	25
Metodología.....	28
Scrum	28
Marco Teórico.....	29
Bussiness Intelligence	29
Power BI.....	29
Python.....	31
Procesos ETL	32
Predicciones.....	33
Symfony	34
Capítulo III	35
Desarrollo.....	35
Descripción de Stakeholders.....	35

	12
Modelado Predictivo.....	36
Población	36
Variables de estudio	37
Técnica e instrumentos de recolección de datos	37
Modelos predictivos	38
Elección del modelo.....	39
Arquitectura del Prototipo.....	39
Nivel de Contexto	40
Nivel de Contenedor	41
Nivel de Componente	43
Nivel de Código	46
Automatización de Procesos de Datos en Python	51
Estructura de Carpetas Reportes.....	51
Procesos Python	53
Interfaz y API con Symfony	59
Interfaz de Usuario (UI)	59
API.....	60
Integración API Symfony con Power BI.....	62
Conexión a la API:	63
Carga de datos:	63
Modelado de datos en Power BI.....	64

Creación de tablas dimensionales	64
Definición y cálculo de medidas en Power BI	66
Medidas a calcular:	66
Implementación de Tableros	67
Capítulo IV	69
Resultados	69
Resultados por Objetivos	74
Capítulo V	76
Conclusiones y Recomendaciones	76
Conclusiones	76
Recomendaciones	77
Bibliografía	78
Apéndices	80

Índice de Figuras

Figura 1 <i>Preparación y administración de los datos por medio de Power BI.</i>	31
Figura 2 <i>Construyendo Procesos ETL con Python.</i>	32
Figura 3 <i>Método Avanzado de Predicción con Python.</i>	33
Figura 4 <i>Contexto: Interacción entre el Personal del Vicerrectorado de Docencia y el Prototipo de Planificación Académica.</i>	41
Figura 5 <i>Contenedor: Flujo de Información en el Sistema de Planificación Académica.</i>	43
Figura 6 <i>Componente: Interacción entre la Interfaz de Usuario y la Aplicación Python Empaquetada.</i>	44
Figura 7 <i>Componente: Interacción entre la aplicación empaquetada y la base de datos.</i>	44
Figura 8 <i>Componente: Interacción entre API con la base de datos.</i>	45
Figura 9 <i>Componente: Interacción entre API con Power BI.</i>	46
Figura 10 <i>Código Python: Extracción y limpieza</i>	46
Figura 11 <i>Código Python: Predicción</i>	47
Figura 12 <i>Código Python: Transformación</i>	47
Figura 13 <i>Código Python: Persistencia</i>	48
Figura 14 <i>Código Symfony: Combinado</i>	49
Figura 15 <i>Código Symfony: Predicciones</i>	50
Figura 16 <i>Código Symfony: Planificación Interfaz</i>	50
Figura 17 <i>Organización de directorios.</i>	52
Figura 18 <i>Interfaz de usuario.</i>	60
Figura 19 <i>Extracción de datos – API Combinado</i>	60
Figura 20 <i>Extracción de datos – API Predicciones.</i>	61
Figura 21 <i>Modelo de Base de datos relación en Power BI.</i>	65
Figura 22 <i>Reporte de Matrículas Actuales.</i>	69

Figura 23 <i>Reporte de Matrículas Proyectadas.</i>	71
Figura 24 <i>Comparación – Matrículas Actuales vs Matrículas Proyectadas.</i>	72
Figura 25 <i>Gráfico de Líneas del Reporte Histórico de Matrículas.</i>	73

Resumen

El Vicerrectorado de docencia de la Universidad de las Fuerzas Armadas "ESPE", ha identificado la necesidad de mejorar los procesos de análisis de datos y generación de reportes más eficientes y efectivos para fortalecer la planificación académica que se realiza cada semestre. Actualmente, el personal del Vicerrectorado dedica una cantidad significativa de tiempo recopilando y procesando datos para crear reportes que son fundamentales en la toma de decisiones para la planificación previo al inicio. Se reconoce la necesidad de optimizar la eficiencia del procedimiento actual a través de una estrategia de inteligencia empresarial (Business Intelligence), proponiendo la implementación de modelos de procesos ETL y modelos matemáticos como herramientas clave para ello. Esta solución automatizará tanto la recopilación como el análisis de datos, agilizando así los procesos de generación de reportes. Además, ofrecerá una visualización más clara y accesible de los datos, lo que mejorará significativamente el acceso del personal del Vicerrectorado a la información relevante.

Para completar esta estrategia de optimización, se prevé el alojamiento de todo este procesamiento en un entorno local. Este entorno permitirá la eficiente gestión, almacenamiento y procesamiento de datos, garantizando su disponibilidad, seguridad y confidencialidad. Con el objetivo de abordar desafíos identificados y asegurar el éxito en la optimización de procesos dentro del Vicerrectorado de la Universidad de las Fuerzas Armadas "ESPE", se ha tomado la decisión de implementar la metodología ágil Scrum para la ejecución de este proyecto. Este enfoque ágil se centra en el desarrollo iterativo, fomentando la adaptabilidad y la mejora continua, esenciales para alcanzar los objetivos del proyecto con eficacia.

Palabras clave: Inteligencia Empresarial, Procesos ETL, Scrum, Planificación Académica.

Abstract

The Vice Chancellorship of Teaching at the Universidad de las Fuerzas Armadas “ESPE” has identified the need to enhance information analysis and generation of more efficient and effective reports to strengthen the academic planning that takes place each semester currently. The Vice Chancellorship's personnel spends a significant amount of time collecting and processing data to create reports that are fundamental in decision-making for pre-start planning. The need to optimize the efficiency of the current procedure through a Business Intelligence strategy is recognized, proposing the implementation of ETL process models and mathematical models as key tools for this purpose. This solution will automate both data collection and analysis, streamlining the reporting generation processes. In addition, it will provide a clearer and more accessible visualization of data, significantly improve the access of the Vice Chancellorship's staff access to relevant information.

To complete this optimization strategy, all this processing will be hosted in a local environment. This environment will allow efficient management, storage, and processing of data, ensuring its availability, security and confidentiality. In order to address the challenges identified and ensure success in the optimization processes within the Vice Chancellorship of the Universidad de las Fuerzas Armadas “ESPE”, it has been made to implement the agile methodology Scrum for the execution of this project. This agile approach focuses on iterative development, fostering adaptability and continuous improvement, essential for achieve the project objectives effectively.

Key Words: Business Intelligence, ETL Processes, Scrum, Academic Planning.

Capítulo I

Introducción

Antecedentes

La Universidad de las Fuerzas Armadas "ESPE", a través de su Vicerrectorado de Docencia, se ha enfrentado a la necesidad crítica de optimizar los procesos de análisis de datos y la generación de reportes. En la búsqueda de una planificación académica más eficaz, el personal de este sector ha dedicado una cantidad considerable de tiempo a la recopilación y procesamiento de datos para elaborar reportes de vital importancia en la toma de decisiones.

La práctica actual de recopilación y análisis de datos presenta deficiencias notables que dificultan la eficacia en la generación de reportes y la toma de decisiones estratégicas. Este proceso, que consume una cantidad sustancial de recursos humanos y temporales, no logra ofrecer una visión completa ni inmediata de la información pertinente. Estas limitaciones han generado dificultades en la identificación de tendencias y en la disponibilidad de datos oportunos para los responsables de la planificación académica.

El reconocimiento de estas limitaciones y desafíos constituye la base fundamental para la búsqueda de soluciones que permitan mejorar la eficiencia, precisión y prontitud en la generación de reportes estratégicos para la planificación académica.

Planteamiento del Problema

En la actualidad la recopilación, análisis y procesamiento de datos desempeñan un papel fundamental en la toma de decisiones, especialmente en el ámbito educativo, donde controlar las diversas actividades académicas es esencial. Sin embargo, al examinar datos históricos, se adquiere la capacidad de prever posibles problemas, lo que permite diseñar estrategias proactivas para abordarlos anticipadamente.

La capacidad de analizar datos históricos no solo proporciona una visión detallada del pasado, sino que también sirve como un indicador clave para la toma de decisiones futuras en el ámbito educativo. Al comprender y analizar tendencias pasadas, se abre la puerta para identificar patrones recurrentes y posibles áreas de mejora. Esta información, utilizada estratégicamente, permite a las instituciones educativas adaptar sus procesos, diseñar planes de acción más efectivos y optimizar sus recursos para brindar una experiencia académica más enriquecedora y eficaz a estudiantes, docentes y personal administrativo.

Justificación

El Vicerrectorado de Docencia de la universidad enfrenta un desafío fundamental en la gestión y análisis de la información relacionada con el proceso académico. La actual metodología manual utilizada para gestionar datos y elaborar reportes demanda una considerable cantidad de tiempo y recursos. Este proceso el cual se considera manual implica la recopilación de datos de diversas fuentes, sin contar con herramientas de análisis automatizado, para la preparación de reportes detallados que sustentan la toma de decisiones.

Este reto subyace en la necesidad apremiante de mejorar la eficiencia y agilidad en el Vicerrectorado de Docencia de la universidad. El enfoque manual actual en el análisis y generación de reportes requiere una inversión significativa de tiempo y recursos humanos, limitando la capacidad de respuesta ante situaciones cambiantes. La implementación de tableros de Power BI se intuye como una solución clave para optimizar este proceso. Automatizar la recopilación, integración y análisis de datos permitirá acceder a información en tiempo real, facilitará una visualización clara y accesible de los datos, y simplificará la identificación de tendencias y patrones relevantes.

Esta necesidad de mejora se destaca como un factor crucial para avanzar hacia una gestión más eficiente y ágil de la información académica en el Vicerrectorado de Docencia. Una planificación bien

elaborada permite asignar los recursos financieros de manera más efectiva, optimizando la contratación de docentes y el uso de infraestructura disponible.

Objetivo

Objetivo General

Implementar tableros de Power BI para automatizar y agilizar el proceso de análisis de datos y generación de reportes en el Vicerrectorado de Docencia de la universidad para mejorar la toma de decisiones y optimizar la gestión de la información académica.

Objetivos Específicos

- i. Identificar los requisitos y necesidades específicas a través de entrevistas y reuniones con los coordinadores del área para comprender las expectativas y requerimientos establecidos por el Vicerrectorado.
- ii. Utilizar una metodología ágil como enfoque de gestión del proyecto, facilitando la colaboración entre el equipo de desarrollo y el Vicerrectorado de docencia, y permitiendo la adaptación continua a medida que se obtiene retroalimentación y se aprende más sobre las necesidades y requisitos del proyecto
- iii. Implementar tableros de Power BI para optimizar la generación de reportes y el análisis de datos para facilitar la toma de decisiones basada en datos y mejorará la eficiencia en la gestión de la información.
- iv. Evaluar la efectividad y el impacto de la implementación de tableros de Power BI a través de la validación de resultados obtenidos, comparando reportes generados, recopilando opiniones de usuarios y realizando análisis comparativos.

Alcance

El alcance del proyecto incluirá la integración de los diversos sistemas de información utilizados en el Vicerrectorado, procesar y transformar datos relevantes, y crear tableros interactivos en Power BI para facilitar la comprensión de información por parte de los responsables del Vicerrectorado.

Se implementarán procedimientos que permitirán automatizar la actualización de los datos y la generación de reportes con el fin de ahorrar tiempo, eliminando la necesidad de realizar tareas manuales y repetitivas.

Como parte del proyecto, se brindará capacitación al personal del Vicerrectorado en el uso de Power BI, con el fin de que puedan sacar el mayor provecho de esta herramienta para realizar análisis de datos de manera independiente.

Hipótesis

Implementar tableros de Power BI en el Vicerrectorado de Docencia reducirá significativamente el tiempo dedicado a la recopilación y generación manual de reportes.

La implementación de tableros de Power BI optimizará la gestión de la información académica, facilitando la identificación de tendencias y patrones relevantes.

La integración de una opción de inteligencia empresarial facilitará la accesibilidad a la información por medio de tableros interactivos, permitiendo así que el personal del Vicerrectorado pueda comprender los datos de manera efectiva.

Analizar datos históricos mediante Python y modelos predictivos ayudará a anticiparse a problemas académicos, permitiendo el diseño de estrategias proactivas.

El acceso mejorado a datos visualizados mediante Power BI tendrá un impacto positivo en la toma de decisiones estratégicas del Vicerrectorado de Docencia.

Capítulo II

Marco Metodológico

La elección de la metodología para este proyecto se basa en la necesidad de obtener una comprensión completa y enriquecedora de los procesos y necesidades del Vicerrectorado de Docencia, específicamente en relación con las matrículas de estudiantes. Para abordar este objetivo, se empleará una metodología mixta que combine enfoques cualitativos y cuantitativos, centrándose en el análisis de datos de matrículas de los últimos 6 semestres.

Enfoque Mixto

La metodología mixta se justifica al integrar métodos cualitativos y cuantitativos. Las reuniones y entrevistas serán los métodos cualitativos que se emplearán para poder explorar las necesidades específicas del Vicerrectorado. Al mismo tiempo, se utilizarán los métodos cuantitativos que se encuentran en base al análisis de datos de matrículas en los reportes entregados de los últimos 6 semestres, los cuales proporcionan información cuantificable para respaldar el análisis general.

Triangulación de Datos

La triangulación se llevará a cabo al combinar la información cualitativa obtenida de entrevistas y reuniones con el análisis cuantitativo de los datos de matrículas. Esta estrategia fortalecerá la validez y confiabilidad de los resultados, fundamentando de manera sólida las conclusiones y las futuras predicciones.

Investigación Descriptiva

La investigación tiene un enfoque descriptivo al medir y comprender a fondo los procesos de matrícula de los estudiantes de los últimos 6 semestres para conseguir una visión más clara y detallada sobre tendencias y patrones.

Diseño de Campo

El diseño de campo se implementa al recopilar datos directamente de los reportes de matrículas en archivos Excel correspondientes a cada semestre. Este enfoque asegura la obtención de datos auténticos y relevantes desde la fuente original.

Enfoque Cuantitativo

El enfoque cuantitativo se focaliza en el análisis de datos de matrículas, permitiendo la medición precisa, la explicación de tendencias y la identificación de patrones en los registros de inscripción de los estudiantes.

Muestra o Población de Estudio

La población de estudio consistirá en las matrículas de estudiantes de los últimos 6 semestres. Estos datos, provenientes de reportes en formato Excel, serán limpiados y analizados para realizar reportes y, posteriormente, se utilizarán para realizar predicciones de matrículas para el siguiente semestre.

Este enfoque metodológico integral permitirá una comprensión profunda de las dinámicas de matrícula, facilitando la toma de decisiones informadas y la optimización de la gestión académica en el Vicerrectorado de Docencia.

Estado del arte

Planteamiento de la Revisión Sistemática

Título

Revisión Sistemática de la Utilización y Beneficios de Power BI en Conjunción con Inteligencia de Negocios, Predicción de Datos mediante Series Temporales, Uso de Python y la Integración con Power BI.

Introducción

Dentro del ámbito empresarial actual, la toma de decisiones basadas en datos es algo esencial. Esta revisión sistemática se propone analizar de manera exhaustiva la literatura existente sobre la utilización de Power BI, herramienta líder en inteligencia de negocios, y su integración con la predicción de datos mediante series temporales aplicadas en el lenguaje de programación Python. El objetivo es identificar y sintetizar los beneficios derivados de esta integración, proporcionando una visión completa de las prácticas actuales y destacando las áreas clave de aplicación.

Preguntas de Investigación

¿Cómo se ha aplicado Power BI en el contexto de la inteligencia de negocios para mejorar la toma de decisiones empresariales?

¿Cuáles son los métodos y enfoques más comunes para la utilización de series temporales en Power BI con fines predictivos?

¿Qué papel ocupa Python con respecto al análisis de datos y cómo se integra con Power BI en entornos empresariales?

Criterios de Inclusión y Exclusión

- **Inclusión:** Estudios y documentos que aborden específicamente la aplicación de Power BI, inteligencia de negocios, series temporales, Python y la integración con entornos empresariales.
- **Exclusión:** Documentos que no aborden directamente la relación entre Power BI y los temas mencionados, así como estudios no accesibles o que no estén disponibles en idioma español o inglés.

Búsqueda y Selección de Estudios

Se realizará una búsqueda exhaustiva en bases de datos académicas, revistas especializadas, y repositorios empresariales. Se aplicarán criterios de inclusión y exclusión para seleccionar estudios relevantes, considerando la calidad metodológica y la aplicabilidad a los temas de interés.

Análisis de Datos y Síntesis:

Los estudios seleccionados serán analizados de manera sistemática, extrayendo información relevante sobre la utilización de Power BI, la inteligencia de negocios, la predicción de datos mediante series temporales aplicado al lenguaje de programación Python. Por medio de la consolidación de datos, se busca identificar tendencias y patrones para obtener resultados significativos.

Conclusiones:

Se presentarán conclusiones basadas en los hallazgos obtenidos, destacando las mejores prácticas, desafíos comunes y áreas de investigación futura. Este estudio contribuirá a la comprensión integral de cómo estas tecnologías convergen para impulsar la toma de decisiones informada en entornos empresariales.

Elaboración y resultados del Estado de Arte.

(Pacci Ayala, 2017) trata sobre la implementación de una solución de inteligencia empresarial para autoservicios el cual permite respaldar la toma de decisiones en la empresa SERTRANS Z&B S.R. Ltda. en la región de Tacna, Perú. El artículo destaca la nula adopción de la inteligencia empresarial como solución dentro de las pequeñas y medianas empresas de la región, por eso, la propuesta se encuentra enfocada para los usuarios con pocos o nulos conocimientos técnicos, esperando como resultado reducir la dependencia del departamento de Tecnologías de la Información. Para eso propone la aplicación Power BI de Microsoft para integrar una solución que sea óptima para dar una solución de inteligencia empresarial.

Se detallan los pasos necesarios para construir esta solución, incluyendo la importación y transformación de datos, la creación de modelos de datos, medidas y KPIs, así como la compartición de resultados con los empleados de la empresa.

La importancia de estas soluciones radica en su potencial para permitir a las pequeñas y medianas empresas aprovechar las tendencias de la inteligencia de negocios, a pesar de la limitada

inversión en tecnología e innovación. Para obtener una ventaja competitiva, se toma como prioridad la toma de decisiones basadas en datos para las empresas y así poder agilizar los procesos.

En conclusión, la tesis resalta la factibilidad y pertinencia de esta solución para las pequeñas y medianas empresas en la región de Tacana, proporcionando un enfoque práctico y orientado en los resultados para superar las restricciones presentes en la implementación de tecnologías de la inteligencia de negocios.

(Cedeño Coello, 2023) trata sobre la implementación de BI con el fin de analizar la información de CRM Zendesk en la empresa Fintech PeiGos S.A. Pretende medir la satisfacción del cliente con el fin de mejorar la toma de decisiones del área de customer success de tal manera que los ayude a mejorar la experiencia que se brinda al usuario.

La solución propuesta en este proyecto es la implementación de un BI utilizando la herramienta Power BI y los servicios de AWS. La recolección de datos sería mediante encuestas con el fin de asegurar la calidad del proyecto

La importancia de esta solución radica en la posibilidad de medir indicadores en tiempo real y actualizarlos según la necesidad del área, lo que permite una toma de decisiones más eficiente y una mejora en la experiencia del usuario. Además, la implementación de los servicios de AWS permite una mayor seguridad y control en el manejo de la data.

En conclusión, la implementación de un BI utilizando Power BI y los servicios de AWS es una solución efectiva para medir indicadores de servicio y satisfacción al cliente en la empresa Fintech PeiGo S.A. Esto permite una mejora en la toma de decisiones y una mejor experiencia del usuario. El uso de AWS con Power BI permite una mayor seguridad y control en el manejo de la data, lo que es fundamental en cualquier proyecto de análisis de datos.

(González Benítez & Perdigón Llanes, 2020) presentaron una revisión de la literatura que se realizó entre el año 2011 y el 2020 acerca de los modelos y técnicas para predecir las producciones de leche bovina. Para esto se utilizó bases de datos como SciELO, Google Scholar y ScienceDirect donde se encontraron muchas combinaciones de términos en español e inglés que estaban relacionados con técnicas, modelos, pronósticos, predicciones y ganado bovino.

Se seleccionaron 37 fuentes bibliográficas, incluyendo artículos de revistas, conferencias y tesis de posgrado, con el fin de ofrecer una visión integral de los enfoques y métodos utilizados en la predicción de la producción lechera.

El estudio examina múltiples técnicas y modelos que se emplean en la predicción de la producción de leche bovina, los cuales toman en cuenta modelos basados en series temporales, softcomputing y métodos híbridos. Además, se abordan métodos para calcular la exactitud de los pronósticos, lo que proporciona un marco integral para evaluar la eficacia de las soluciones propuestas.

La importancia de estas soluciones radica en su potencial para mejorar la toma de decisiones en la producción lechera, permitiendo a los productores anticipar y planificar de manera más efectiva. El uso de técnicas y modelos precisos para la predicción, posibilita la optimización para la gestión de recursos y el mejoramiento en la eficiencia operativa para contribuir en un desarrollo sostenible del sector lácteo.

En conclusión, el artículo ofrece una visión detallada de los modelos y técnicas utilizados en la predicción de las producciones de leche bovina, así como métodos para evaluar la exactitud de los pronósticos. Esta revisión bibliográfica proporciona una base sólida para futuras investigaciones y aplicaciones prácticas en el campo de la producción lechera, destacando la importancia de la precisión en los pronósticos para la toma de decisiones informadas y efectivas.

(Santamaría Valero, 2021) se enfoca en la aplicación de datos abiertos para la actualización y predicción en el contexto de la isla de Tenerife. Se aborda la recopilación, transformación y análisis de

datos meteorológicos y de incidencia acumulada de COVID-19, con el objetivo de desarrollar modelos predictivos que permitan anticipar tendencias y comportamientos.

La solución propuesta en el trabajo de fin de grado se centra en la automatización y tratamiento de datos a través de la creación de un "semáforo" que evalúa la fiabilidad de la API meteorológica de GRAFCAN, así como en la implementación de modelos de predicción utilizando las librerías Prophet y Neural Prophet.

La importancia de estas soluciones radica en la capacidad de proporcionar información actualizada y fiable, así como en la posibilidad de realizar predicciones que pueden ser fundamentales para la toma de decisiones en situaciones como la gestión de la pandemia de COVID-19. Se facilitará la recolección de información valiosa por medio de la visualización de datos a través de Power BI, lo que permitirá poder anticiparse a múltiples fenómenos relevantes, automatizando previamente con modelos predictivos.

En conclusión, el trabajo de fin de grado destaca la utilidad de las predicciones con series de tiempo y el uso de herramientas como Power BI para el análisis de datos en contextos complejos como el de la isla de Tenerife. Aunque se reconoce la limitación de los modelos debido a la escasez de datos, se resalta la importancia de continuar explorando y mejorando estas técnicas para obtener resultados más concluyentes en el futuro.

Metodología

Scrum

La metodología Scrum se justifica en este proyecto debido a que permite un enfoque ágil y colaborativo en el desarrollo de software. Scrum se basa en iteraciones cortas y frecuentes, lo que facilita la adaptación continua a medida que se obtiene retroalimentación y se aprende más sobre las necesidades y requisitos del proyecto. En el caso de implementar tableros de Power BI, Scrum permitirá una mayor flexibilidad en la incorporación de funcionalidades y ajustes, así como una entrega temprana

de resultados parciales. Además, fomentará la colaboración entre el equipo de desarrollo y el Vicerrectorado de Docencia, asegurando una participación activa en el proceso y una mayor alineación con las expectativas y necesidades del cliente. En resumen, Scrum permitirá una implementación más eficiente y adaptativa de los tableros de Power BI, contribuyendo a mejorar la toma de decisiones y optimizar la gestión de la información académica.

Marco Teórico

Business Intelligence

BI o también conocida en español como Inteligencia de Negocios, llegan a ser diferentes herramientas o estrategias que te ayudan para administrar y generar conocimiento a través de un análisis de datos en diferentes contextos. Su principal objetivo llega a ser el poder comprender el funcionamiento actual de una empresa con el fin de poder anticipar diferentes problemas que se pueden presentar a futuro para poder respaldar todas las decisiones empresariales.

Este enfoque facilita la provisión de información accesible y orientada a la toma de decisiones, sin importar la fuente de datos, lo que permite a los usuarios manipular la información de manera efectiva para respaldar sus decisiones. (Marqués, 2015).

Power BI

Esta herramienta es muy potente y dinámica aplicada en la inteligencia empresarial la cual es desarrollada por Microsoft, fue desarrollada para dar un soporte a las empresas con la recopilación, análisis y visualización de los datos para poder obtener información relevante. Esta aplicación tiene una habilidad muy desarrollada para integrar información que viene de diferentes fuentes y entregar análisis dinámicos e interactivos que permiten que el usuario pueda comprender sus datos y la toma de decisiones basada en datos (Microsoft, 2023). Su funcionalidad se desglosa en varios aspectos clave:

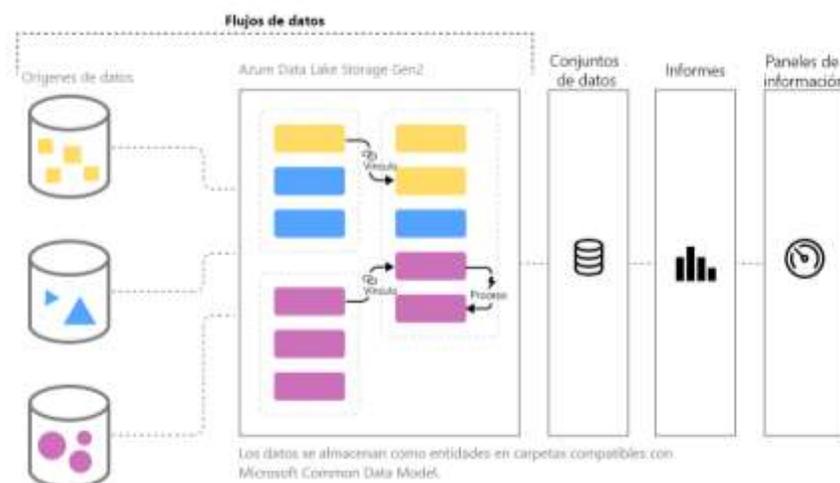
- **Conexión de datos:** tiene una diversa conexión de datos, desde base de datos almacenadas localmente hasta aplicaciones en la nube. Por lo cual, el usuario puede importar datos desde

una aplicación o implementar una conexión en tiempo real, de tal manera que se obtenga información actualizada.

- **Transformación de datos:** con Power BI, los usuarios tienen la capacidad de llevar a cabo una limpieza exhaustiva y diferentes transformaciones en los datos. Pueden aplicar filtros, modificar formatos, combinar conjuntos de datos, eliminar duplicados y realizar otras acciones para preparar los datos para el análisis.
- **Modelado de datos:** facilita la creación de modelos de datos que permiten la relación entre distintas tablas y conjuntos de información. Los usuarios pueden definir jerarquías, permitir relaciones entre datos y crear cálculos utilizando el lenguaje DAX (Data Analysis Expressions).
- **Análisis y visualizaciones:** después de que los datos son preparados, Power BI entrega una variedad de opciones para la visualización, para que los usuarios puedan generar reportes interactivos los cuales pueden llegar a ser tan simples como una tabla o tener un panel completo para analizar los datos de una forma más efectiva.
- **Colaboración y distribución:** los reportes y paneles finales generados pueden llegar a compartirse de manera sencilla con otros usuarios pertenecientes a la organización. También se pueden publicar reportes en la nube permitiendo el acceso desde cualquier lugar y dispositivo, así como la programación de actualizaciones automáticas para mantener la información al día.

Figura 1

Preparación y administración de los datos por medio de Power BI.



Nota. La extracción y preparación de datos implica limpiar, transformar y organizar los datos tal como vienen para convertirla en información útil. Tomado de Davidiseminger. (2023). Introducción a los flujos de datos. Microsoft Learn. <https://learn.microsoft.com/es-es/power-bi/transform-model/dataflows/dataflows-introduction-self-service>

Python

Python es una opción popular para los procesos ETL debido a varias razones clave. En primer lugar, ofrece una gran cantidad de bibliotecas especializadas en el manejo de datos, como Pandas, que permiten manejar conjuntos de datos y realizar fácilmente operaciones de limpieza, transformación y análisis de datos. Estas herramientas proporcionan funciones y métodos predefinidos para manipular datos, facilitando tareas como la manipulación de tablas, cálculos complejos, filtrado, agregación y más.

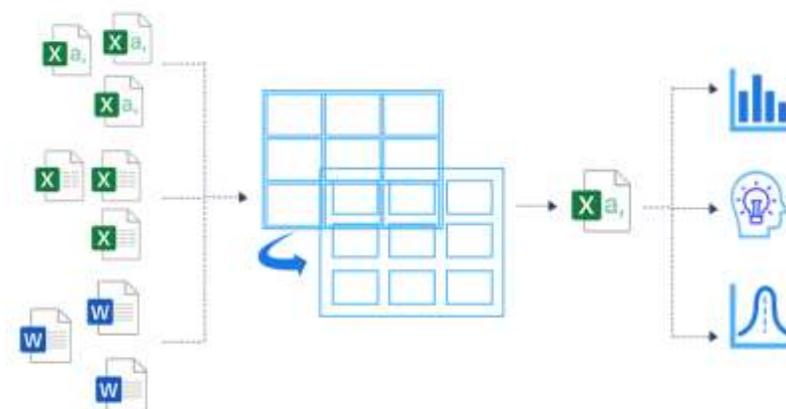
La versatilidad de Python permite construir flujos de trabajo ETL completos, desde la extracción de datos desde múltiples fuentes (bases de datos, APIs, archivos planos, etc.) hasta la transformación de estos datos según las necesidades del proyecto y finalmente, la carga en un almacén de datos o sistema de destino (Gutiérrez-Cabello, 2020)

Procesos ETL

El pilar en la inteligencia de negocios llega a ser los procesos ETL. Dichos procesos tienen la capacidad de poder estabilizar todos los datos provenientes de diferentes fuentes, desde base de datos hasta archivos planos, todos estos en un solo repositorio. Con ello podemos elevar su calidad y uniformidad, además de facilitar el acceso a la información. La fase de extracción permite recopilar datos brutos, la transformación se encarga de limpiarlos, estandarizarlos y enriquecerlos, mientras que la carga los introduce en el almacén de datos. Esta laboriosa labor de transformación garantiza que los datos sean coherentes y estén preparados para ser utilizados en análisis avanzados, creación de reportes detallados o aplicaciones de inteligencia de negocios, cumpliendo con los criterios y requisitos específicos de los usuarios finales. Así, los procesos ETL no solo son cruciales para la integración de datos, sino que también desempeñan un papel crucial en la calidad y fiabilidad de la información utilizada para la toma de decisiones empresariales (Díaz, 2010).

Figura 2

Construyendo Procesos ETL con Python.



Nota. Importancia de Python para optimizar y simplificar el flujo de procesos ETL, para mejorar la gestión y el análisis eficaz de datos en empresas. Tomado de Abhinav. (2023, 3 abril). Building ETL using Python - SarasAnalytics. <https://sarasanalytics.com/blog/etl-using-python/>

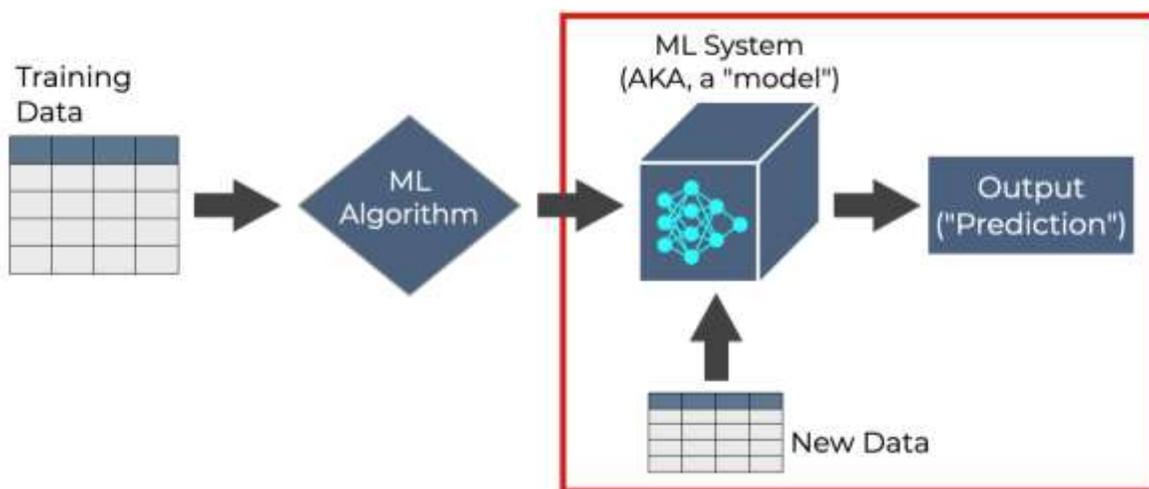
Predicciones

Los modelos predictivos en Python son herramientas que emplean datos históricos para predecir futuros resultados o estimaciones. Para poder obtener proyecciones sobre eventos futuros, dichos algoritmos se basan en el aprendizaje de diferentes patrones dentro de los datos. Su importancia radica en respaldar decisiones informadas mediante predicciones, permitiendo la planificación estratégica y la optimización de recursos.

Además, estos modelos mejoran la precisión en las predicciones, automatizan procesos y revelan conocimientos ocultos en los datos, proporcionando ideas valiosas para la toma de decisiones y el desarrollo empresarial. En Python, existen diversas bibliotecas, que ofrecen una amplia gama de algoritmos de modelado predictivo, brindando a los usuarios la capacidad de implementar y utilizar modelos según sus necesidades específicas (Korstanje, 2023).

Figura 3

Método Avanzado de Predicción con Python.



Nota. En machine learning, la predicción usa un modelo entrenado para generar resultados con nuevos datos de entrada. Tomado de Ebner, J. (2021, 24 julio). A quick introduction to machine learning - Sharp Sight. <https://www.sharpsightlabs.com/blog/introduction-to-machine-learning/>

Symfony

Si mencionamos modularidad, flexibilidad y escalabilidad podemos observar que llegan a ser características esenciales de Symfony, un framework escrito en PHP. Este framework ofrece una variedad de herramientas y componentes que permiten la simplificación para el desarrollo de aplicaciones que sean robustas y seguras (Symfony, Symfony Documentation, s.f.) .Entre sus características principales se encuentran:

- **Enrutamiento:** Permite definir rutas para las diferentes URLs de la aplicación.
- **Controladores:** Se encargan de procesar las solicitudes HTTP y generar la respuesta.
- **Plantillas:** Permiten definir la estructura y el diseño de las páginas web.
- **Formularios:** Facilitan la creación de formularios web y la validación de datos.
- **Seguridad:** Ofrece mecanismos para proteger la aplicación contra ataques web.
- **Testing:** Facilita la creación de pruebas unitarias y pruebas funcionales.

Symfony se utiliza en una amplia gama de proyectos web, desde pequeñas empresas hasta grandes corporaciones. Algunas empresas que utilizan Symfony son:

- BBC
- Facebook
- Google
- Wikipedia

Capítulo III

Desarrollo

Descripción de Stakeholders

En el marco del proyecto "Automatización de procesos de análisis de información y generación de reportes mediante la implementación de tableros de visualización de datos con la herramienta Power BI para el Vicerrectorado de Docencia de la Universidad de las Fuerzas Armadas ESPE", se identifican varios stakeholders clave cuyos roles son fundamentales para el proyecto. Estos stakeholders incluyen:

Vicerrectorado de Docencia de la Universidad de las Fuerzas Armadas ESPE: Como el principal cliente y beneficiario del proyecto, el Vicerrectorado de Docencia busca agilizar el proceso de generación de reportes sobre matrículas, número de horas y cantidad de NRC (Números de Registro de Cursos) en los diferentes períodos académicos, Para poder mejorar la gestión académica a través de la toma efectiva de decisiones. Su interés radica en obtener información precisa y oportuna, separada por campus y departamentos, para planificar eficientemente las actividades académicas.

Estudiantes de Ingeniería (Desarrolladores del Proyecto): Como los encargados directos del desarrollo e implementación del sistema de reporte académico, los estudiantes de ingeniería tienen la responsabilidad de crear una solución tecnológica efectiva que cumpla con los requisitos del Vicerrectorado de Docencia. Su objetivo es automatizar los procesos de análisis de información y generación de reportes utilizando la herramienta Power BI, asegurando la precisión, la accesibilidad y la relevancia de los datos presentados. Además, están comprometidos con la revisión semanal del sistema para corregir errores, incorporar sugerencias y adaptar el sistema a las necesidades cambiantes.

Tutora de Tesis: La tutora de tesis asignada al proyecto desempeña un papel crucial como mentora y guía para los estudiantes desarrolladores. Su función principal es proporcionar orientación técnica, metodológica y académica a lo largo de todas las etapas del proyecto. La tutora supervisa el

progreso, ofrece retroalimentación constructiva y garantiza la calidad y el cumplimiento de los objetivos establecidos.

Otros Stakeholders:

- Departamentos Académicos: Se beneficiarán de la información proporcionada por los reportes para la planificación y gestión de sus actividades.
- Comunidad Universitaria: En general, se beneficiará de la mejora en la eficiencia y transparencia de la gestión académica.

Relaciones entre Stakeholders:

- Vicerrectorado de Docencia: Es el cliente principal del proyecto y tiene una relación directa con los estudiantes de ingeniería y la tutora de tesis.
- Estudiantes de Ingeniería: Trabajan en estrecha colaboración con la tutora de tesis para desarrollar el proyecto y cumplir con las expectativas del Vicerrectorado de Docencia.
- Tutora de Tesis: Es el enlace entre el Vicerrectorado de Docencia y los estudiantes de ingeniería, y vela por el cumplimiento de los objetivos del proyecto.

Comunicación:

- Se establecerá una comunicación fluida y constante entre los stakeholders a través de reuniones presenciales, correos electrónicos, reportes y presentaciones.
- Se utilizará una plataforma online (Drive, correo electrónico) para compartir información y documentos relevantes para el proyecto

Modelado Predictivo

Población

La población de estudio está conformada por reportes de matrículas de estudiantes de los últimos 6 semestres (más el actual) en la Universidad de las Fuerzas Armadas ESPE. Se espera que la

información recopilada represente una muestra significativa de la dinámica de matrículas en la institución.

Variables de estudio

Variable Independiente:

- Campus: Ubicación física de la carrera (Matriz, Latacunga, Santo Domingo, etc.).
- Departamento: Área académica a la que pertenece la carrera (Ingeniería, Ciencias Administrativas, etc.).
- Asignatura: Materia específica dentro del plan de estudios de una carrera.

Variable Dependiente:

- Número de Matrículas: Cantidad de estudiantes inscritos en una asignatura en un periodo académico específico.
- Número de NRC: Número de Registro de Curso, código único que identifica una sección específica de una asignatura.
- Número de Horas: Carga horaria total de una asignatura.

Variables de Control:

- Semestre: Periodo académico (1er semestre, 2do semestre, etc.).
- Año: Año académico (2023, 2024, etc.).

Técnica e instrumentos de recolección de datos

Para la recolección de datos, se utilizarán principalmente reportes en formato Excel proporcionados por el personal del Vicerrectorado de Docencia perteneciente a la Universidad de las Fuerzas Armadas ESPE. Estos reportes contienen información detallada sobre las matrículas de los estudiantes, el número de horas asignadas a cada asignatura y los NRC correspondientes.

La limpieza y el análisis de los datos se llevarán a cabo utilizando herramientas de procesamiento y análisis de datos con el lenguaje de programación Python, con énfasis en la

herramienta Power BI para la generación de reportes. Se llevó a cabo un proceso de depuración de datos para asegurar la precisión y coherencia de la información recopilada.

Además, para la proyección de matrículas para el siguiente semestre, se analizarán los reportes de los 6 semestres académicos anteriores. Se emplearán técnicas de análisis estadístico y modelado predictivo para estimar el número de matrículas y NRC esperados en cada asignatura, considerando variables como campus, departamento y asignatura.

Modelos predictivos

Los modelos utilizados fueron regresión lineal, árboles de decisión y suavizamiento exponencial para la proyección de matrículas y NRC (Números de Registro de Cursos) en semestres académicos futuros. A continuación, describiré brevemente cada modelo y cómo se aplicarán en el proyecto:

Regresión Lineal:

- **Funcionamiento:** es un método estadístico que se utiliza para definir la relación que existe entre la variable dependiente y entre una o múltiples variables independientes. En este caso, la regresión lineal se aplicará para predecir el número de matrículas o NRC en una asignatura en función de variables predictoras como el semestre, el campus, el departamento, entre otros.
- **Aplicación en el proyecto:** Se recopilaron datos históricos de matrículas y NRC de asignaturas pasadas junto con las variables predictoras mencionadas anteriormente. Luego, se ajustará un modelo de regresión lineal a estos datos para predecir las matrículas y NRC esperados en semestres académicos futuros.

Árboles de Decisión:

- **Funcionamiento:** Se construyen a partir de un conjunto de datos históricos y dividen los datos en un conjunto pequeño el cual se encuentra basado en características específicas.
- **Aplicación en el proyecto:** Se utilizarán árboles de decisión para predecir las matrículas y NRC en asignaturas futuras utilizando variables como el campus, el departamento, el semestre, entre

otras. Los árboles de decisión pueden proporcionar insights sobre qué variables son más influyentes en las predicciones y ayudar a identificar patrones complejos en los datos.

Suavizamiento Exponencial:

- **Funcionamiento:** El suavizamiento exponencial es una técnica de pronóstico que requiere una base histórica de observación para poder proyectar futuros valores. Se basa en la suposición de que las tendencias y patrones en los datos pasados continuarán en el futuro.
- **Aplicación en el proyecto:** El suavizamiento exponencial será utilizado para suavizar las series temporales de matrículas y NRC en asignaturas pasadas. Este modelo permitirá proyectar las tendencias futuras en base a las observaciones históricas, lo que facilitará la estimación de las matrículas y NRC en los próximos semestres.

Elección del modelo

Múltiples factores clave interfieren al momento de elegir el modelo más adecuado, como la naturaleza de los datos, la precisión deseada y la complejidad del modelo. Se utilizarán diferentes métricas de evaluación para comparar los modelos de proyección para poder determinar cuál modelo permitirá un mejor ajuste de los datos.

Implementación en Python:

Los modelos predictivos se implementarán utilizando el lenguaje de programación Python.

NumPy, Pandas, Scikit-learn son algunas librerías utilizadas para el análisis de datos.

Arquitectura del Prototipo

Para el diseño de la arquitectura del prototipo, se optó por emplear un modelo C4 con el objetivo de facilitar una comunicación más clara con el equipo de desarrollo, tanto en las etapas iniciales de diseño como en la documentación retrospectiva de una base de código ya existente. Este modelo se compone de cuatro niveles estructuran la arquitectura del prototipo de manera organizada:

Nivel de Contexto

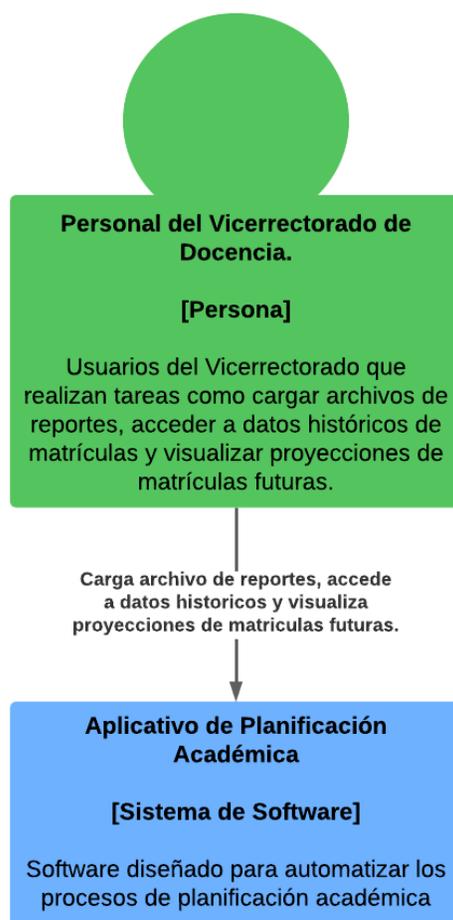
En este nivel se describe la interacción más general entre los actores (Personal del Vicerrectorado de Docencia) y el sistema (Prototipo de Planificación Académica).

Personal del Vicerrectorado de Docencia. Representa a los usuarios que desempeñan funciones en la Universidad de las Fuerzas Armadas ESPE. Este grupo de usuarios tiene la responsabilidad de llevar a cabo diversas tareas relacionadas con la planificación académica. Esto incluye actividades como cargar archivos de reportes, acceder a datos históricos de matrículas y visualizar proyecciones de matrículas futuras.

Prototipo de Planificación Académica. Es el sistema de software diseñado específicamente para automatizar los procesos de planificación académica en la Universidad de las Fuerzas Armadas ESPE. El sistema proporciona una variedad de funcionalidades que permiten cargar archivos de reportes, acceder a datos históricos de matrículas y generar proyecciones de matrículas futuras.

Figura 4

Contexto: Interacción entre el Personal del Vicerrectorado de Docencia y el Prototipo de Planificación Académica.



Nota. El diagrama representa la relación contextual entre el personal del Vicerrectorado de Docencia y el Prototipo de Planificación Académica, resaltando las tareas realizadas por los usuarios y la función del prototipo en la automatización de procesos académicos.

Nivel de Contenedor

Este nivel garantiza la eficiencia, portabilidad y escalabilidad de la aplicación. Engloba varios componentes técnicos que colaboran entre sí para proporcionar una solución integral y robusta.

Interfaz de Usuario (UI). La interfaz de usuario brinda a los usuarios del Vicerrectorado de Docencia un medio intuitivo para interactuar con la aplicación. Facilita la carga archivos de reportes y simplifica la preparación para la ejecución de la aplicación empaquetada.

Aplicación Python Empaquetada. La aplicación Python empaquetada constituye el corazón de la solución. Al empaquetarse, se garantiza la inclusión de todas las dependencias necesarias, lo que simplifica su implementación en diversos entornos sin la necesidad de preocuparse por las dependencias externas. Esto simplifica el despliegue y asegura una ejecución coherente en cualquier lugar.

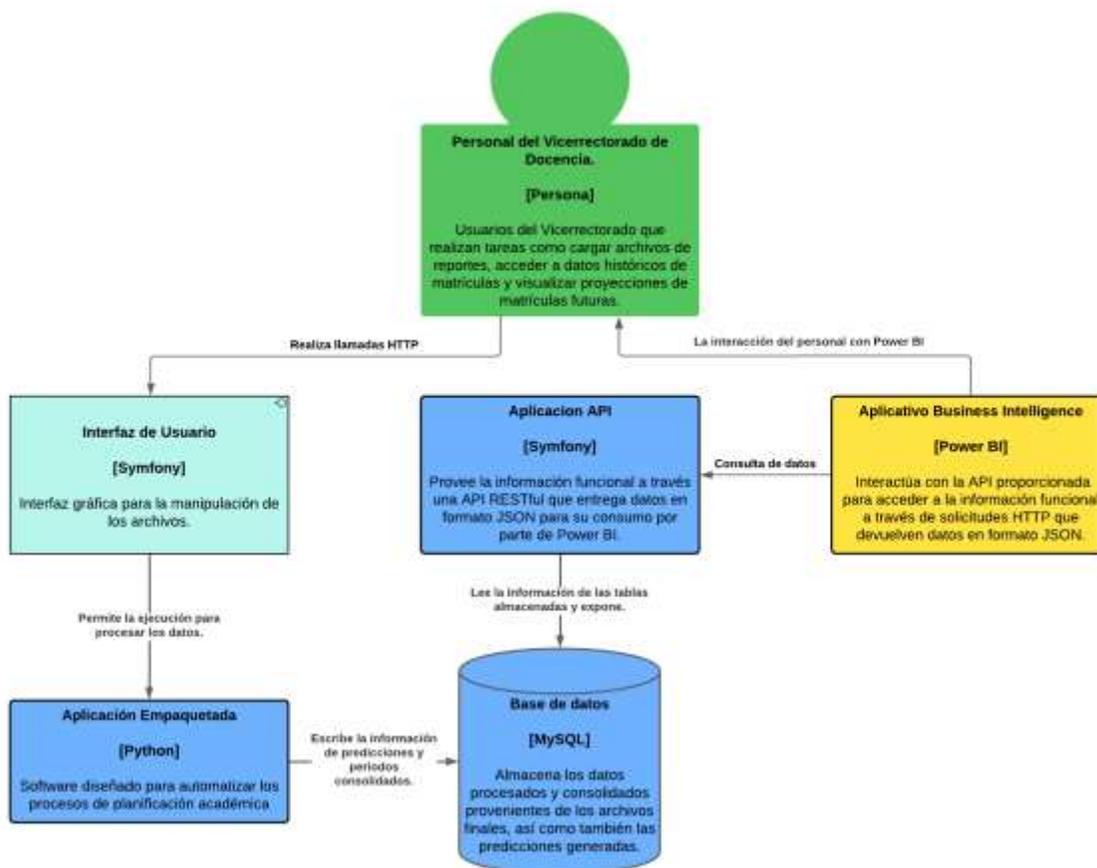
Base de Datos. La base de datos desempeña el papel fundamental de almacenar y gestionar los datos generados por la aplicación. Esto abarca reportes de matrículas y datos de predicciones para la planificación académica. Si bien cumple con el almacenamiento y gestión de datos, puede no estar diseñada de manera óptima para el análisis y la visualización de datos, ya que el diseño final se lleva a cabo en Power BI.

API Symphony. La API Symphony desempeña un papel crucial en la integración de la aplicación con otros sistemas o servicios. Ofrece una forma estructurada y segura de acceder a la funcionalidad de la aplicación mediante solicitudes HTTP. Esta API facilita la comunicación con otros componentes del sistema y permite la automatización de procesos, la extracción de datos y la integración con herramientas que permiten garantizar la interoperabilidad y la eficiencia del sistema en su conjunto.

Power BI. La aplicación permite habilitar la visualización y el análisis de los datos, para que el usuario tenga la posibilidad de crear paneles interactivos o reportes dinámicos para mejorar la toma de decisiones. Con Power BI se puede tener una capacidad muy elevada de análisis de los datos por medio de los diferentes reportes generados.

Figura 5

Contenedor: Flujo de Información en el Sistema de Planificación Académica.



Nota. El diagrama ilustra el flujo de información en el Sistema de Planificación Académica, detallando los componentes contenedores que participan en el procesamiento y gestión de los datos.

Nivel de Componente

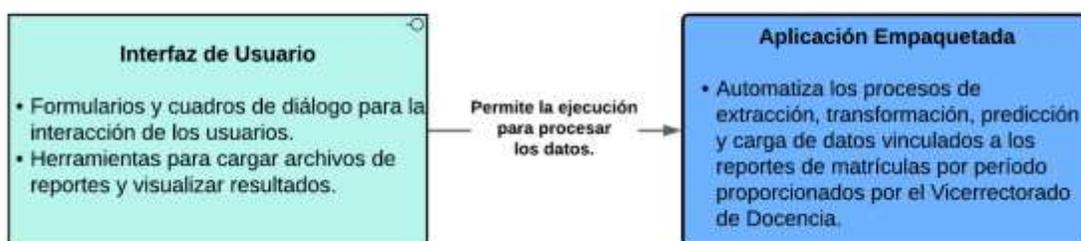
Los contenedores mencionados anteriormente desempeñan varios roles críticos que posibilitan automatizar procesos de generación de información y análisis de datos que son relevantes para la toma de decisiones en el ámbito académico. A continuación, se describe la importancia de cada uno de estos contenedores y cómo interactúan entre sí, señalando los componentes que contienen:

Contenedor 1. La interacción entre la interfaz de usuario y la aplicación Python empaquetada implica los siguientes componentes.

Figura 6

Componente: Interacción entre la Interfaz de Usuario y la Aplicación Python Empaquetada.

④ Contenedor 1: Interfaz de Usuario (Symfony) - Aplicación Empaquetada (Python)



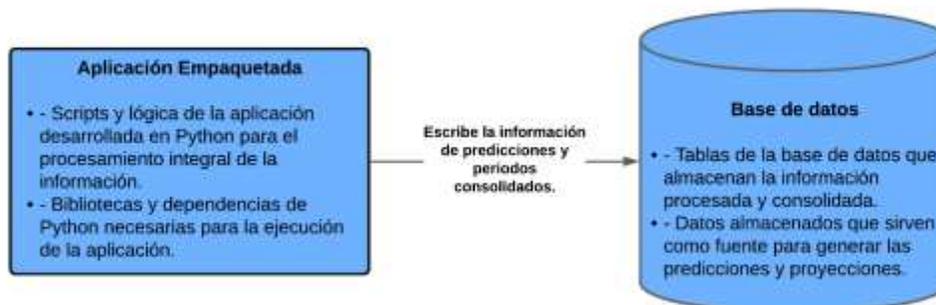
Nota. La Interfaz de Usuario (UI) desarrollada en Symfony permite a los usuarios cargar archivos de reportes, facilitando la ejecución de la Aplicación Python Empaquetada, permitiendo automatizar procesos de extracción, transformación, predicción y carga de datos, para gestionar los reportes de matrículas proporcionados por el Vicerrectorado de Docencia.

Contenedor 2. La interacción entre la aplicación empaquetada y la base de datos implica los siguientes componentes.

Figura 7

Componente: Interacción entre la aplicación empaquetada y la base de datos.

④ Contenedor 2: Aplicación Empaquetada (Python) - Base de Datos (MySQL)

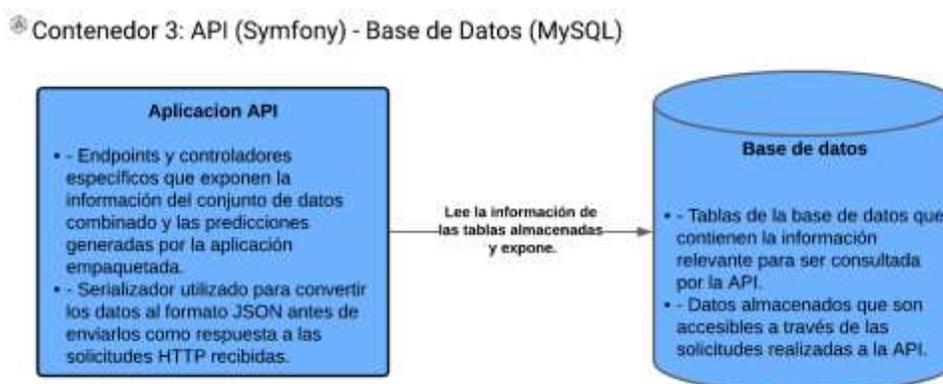


Nota. La aplicación empaquetada en Python utiliza la base de datos MySQL como su repositorio de datos. Con el fin de aprovechar el almacenamiento persistente y estructurado que nos brinda la base de datos, nuestra aplicación accede a la misma con el fin de procesar información y almacenar resultados.

Contenedor 3. La interacción entre la API con la base de datos implica los siguientes componentes.

Figura 8

Componente: Interacción entre API con la base de datos.

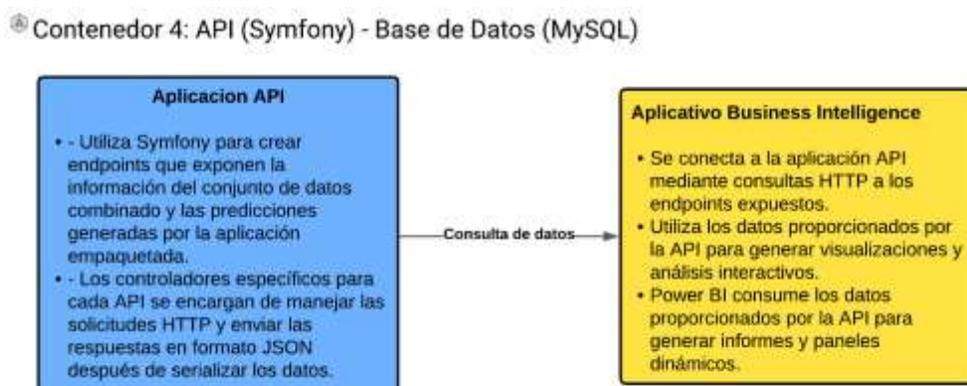


Nota. La API Symfony utiliza la base de datos MySQL como su fuente de datos. La API proporciona acceso a los datos almacenados en la base de datos a través de endpoints específicos para su consumo por aplicaciones.

Contenedor 4. La interacción entre API con Power BI implica los siguientes componentes.

Figura 9

Componente: Interacción entre API con Power BI.



Nota. La aplicación API se comunica con Power BI a través de endpoints que exponen la información del conjunto de datos combinado y las predicciones generadas por la aplicación empaquetada.

Nivel de Código

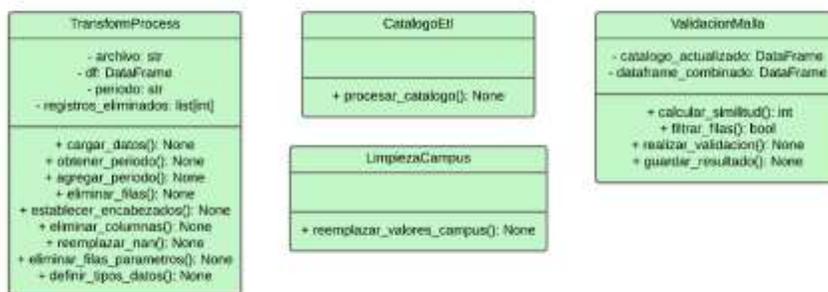
Este nivel hace referencia a la capa de implementación del software, donde se escriben cada una de las instrucciones específicas que llevan a cabo las funciones con las características deseadas. En este nivel, el equipo de desarrollo transforma los requisitos del proyecto en código ejecutable utilizando lenguajes de programación, frameworks y bibliotecas.

Python. Abarca la implementación detallada de las funcionalidades del prototipo utilizando el lenguaje para permitir comprender los diferentes componentes y herramientas que se utilizaron.

Figura 10

Código Python: Extracción y limpieza

Aplicación Empaquetada (Python) - Extracción y Limpieza

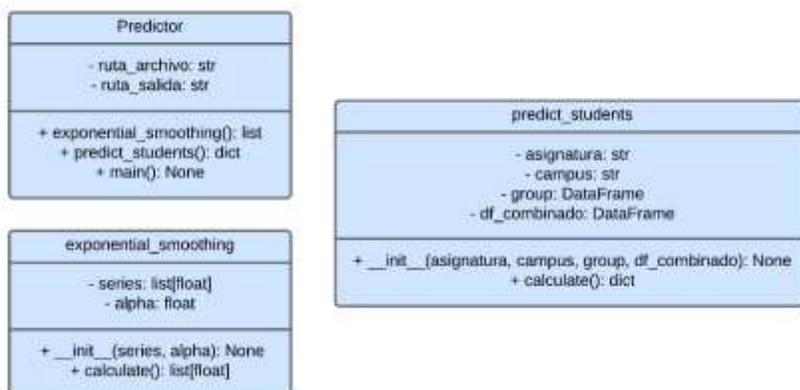


Nota. Dentro del código Python utilizado para el proyecto de planificación académica, se emplean clases dedicadas a la extracción y limpieza de datos. Estas clases son fundamentales para el procesamiento inicial de la información proporcionada por el Vicerrectorado antes de su análisis y utilización en el sistema.

Figura 11

Código Python: Predicción

Aplicación Empaquetada (Python) - Predicción

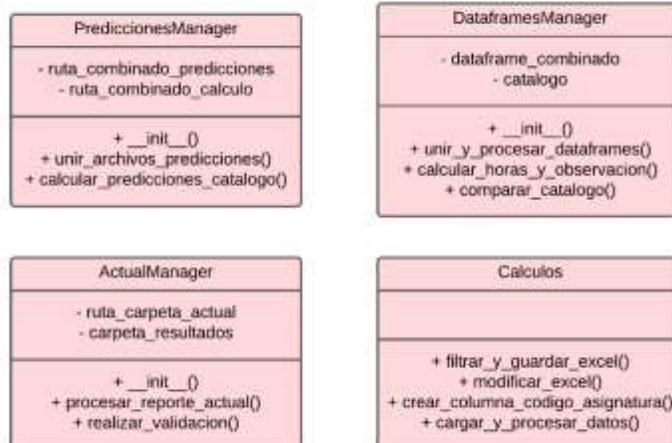


Nota. El código Python dedicado a las predicciones forma parte integral del proyecto de planificación académica. Esta sección del código se encarga de utilizar algoritmos de análisis predictivo para generar proyecciones futuras.

Figura 12

Código Python: Transformación

④ Aplicación Empaquetada (Python) - Transformación

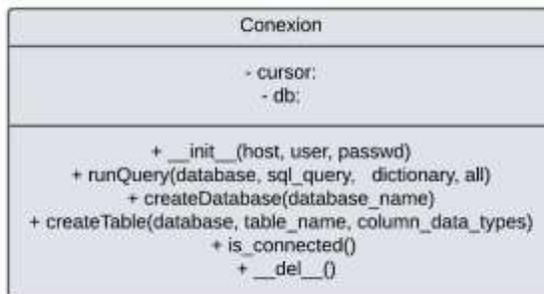


Nota. Esta sección del código se encarga de aplicar procesos de transformación y preparación a los datos extraídos, asegurando que estén en un formato adecuado para su posterior análisis y utilización en el sistema. Para optimizar la calidad y estructura de los datos, utilizaremos las funciones y métodos de transformación, de tal manera que facilitamos su interpretación.

Figura 13

Código Python: Persistencia

④ Aplicación Empaquetada (Python) - Persistencia



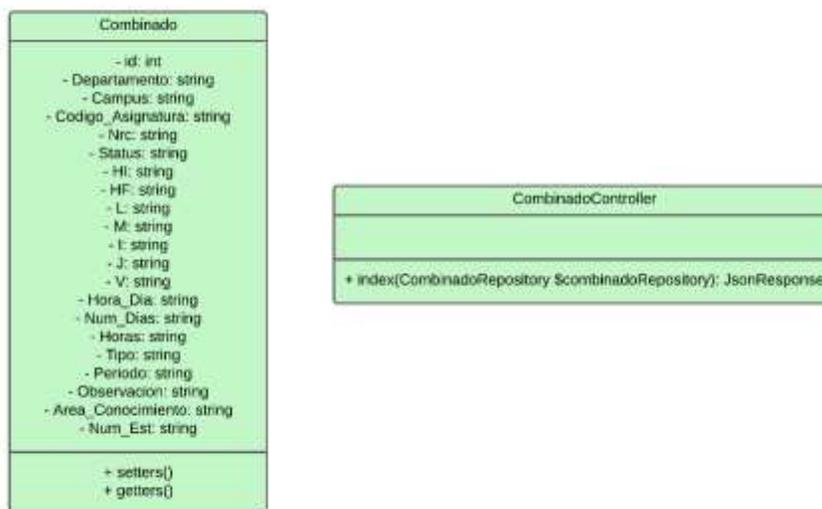
Nota. Esta sección del código se encarga de gestionar la persistencia a la base de datos. Utilizando métodos y clases específicas.

Symfony. Abarca la implementación detallada de las funcionalidades del prototipo utilizando el framework, componentes y herramientas que se utilizaron.

Figura 14

Código Symfony: Combinado

API (Symfony) - Combinado Reportes

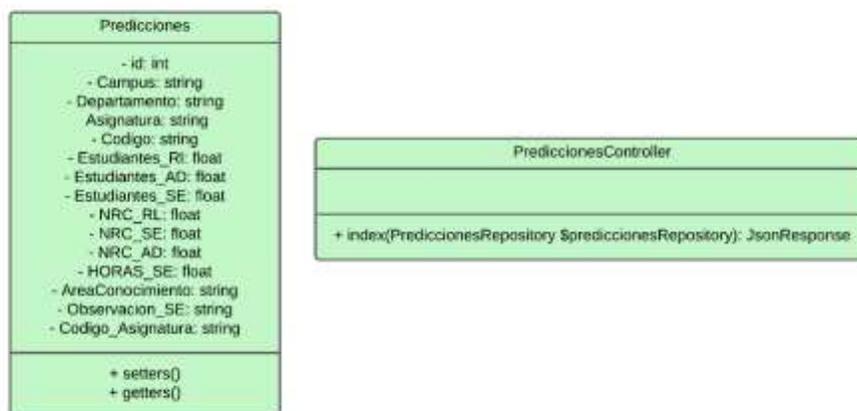


Nota. El código Symfony dedicado al componente de combinación se centra en mapear y recuperar la información utilizando el Repository correspondiente. Esta sección del código se encarga de crear un endpoint específico que permite la consulta de datos desde la base de datos. Los datos recuperados se formatean en formato JSON para su posterior impresión y consumo.

Figura 15

Código Symfony: Predicciones

API (Symfony) - Predicciones Reportes

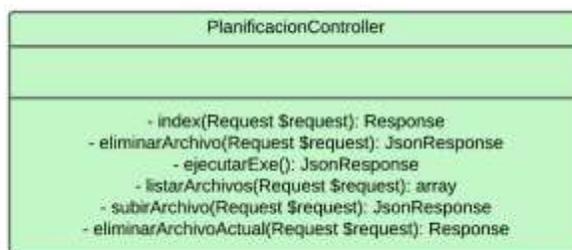


Nota. El código Symfony dedicado al componente de predicciones se encarga principalmente de la consulta de datos necesarios para realizar las predicciones. Significa que podrá acceder a la información que tengamos almacenada en la base de datos. Una vez obtenidos los datos, son procesados y formateados en formato JSON para su posterior utilización en el sistema de predicción y su consumo.

Figura 16

Código Symfony: Planificación Interfaz

API (Symfony) - Planificacion Interfaz



Nota. El código Symfony dedicado al componente de planificación de interfaz se enfoca en la implementación de métodos para el manejo y gestión de la interfaz de usuario. Estos métodos se

encargan de gestionar la interacción del usuario con el sistema, como la carga de archivos, la visualización de los reportes cargados y la ejecución del prototipo Python.

Automatización de Procesos de Datos en Python

En el marco del proyecto, se llevó a cabo un proceso de desarrollo en el lenguaje de programación Python. Este proceso se enfoca en la extracción, transformación, predicción y carga de datos vinculados a los reportes de matrículas por periodo proporcionados por el Vicerrectorado de Docencia. La relevancia de este procedimiento entrega la capacidad para automatizar tareas esenciales en el análisis y la gestión de la información académica. Por lo cual, a pesar de que nos ayuda a optimizar la utilización de los diferentes recursos disponibles podemos observar una facilidad en la toma de decisiones. A continuación, se presenta una descripción detallada del flujo de trabajo implementado, resaltando cada etapa del mismo.

Estructura de Carpetas Reportes

La estructura de directorios donde se encuentran los documentos que a ser manipulados por medio de los scripts está organizada dentro del directorio raíz llamado "Data", de la siguiente manera:

Catálogo: contiene el archivo relacionado con información sobre el catálogo de asignaturas vigentes. Este archivo es fundamental para poder llevar a cabo el análisis, ya que permite enfocarse únicamente en asignaturas actuales y evitar el desperdicio de recursos en aquellas que han sido eliminadas en los reportes de periodos anteriores.

Input: contiene los reportes de los periodos anteriores, de donde se extraerán los datos necesarios para realizar el procesamiento de análisis y proyección.

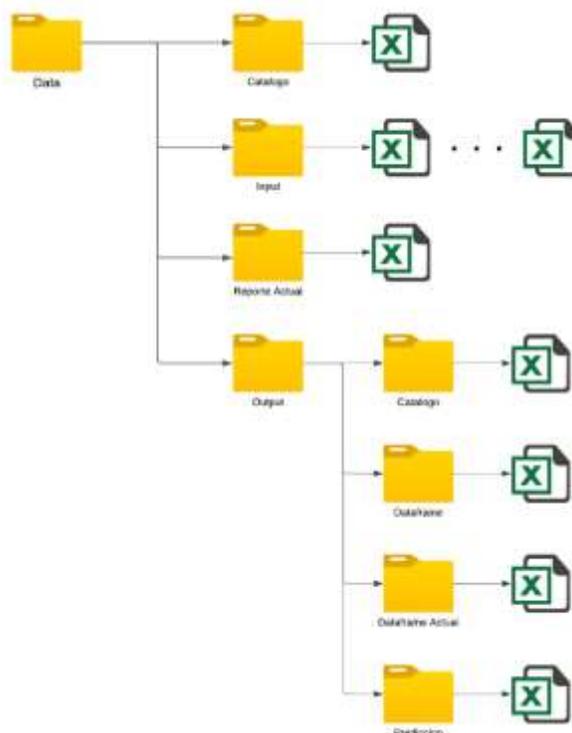
Reporte Actual: contiene el reporte que está en continuas modificaciones para el periodo actual. Este documento facilita el análisis de las proyecciones y verificar su validez.

Output: directorio principal que contendrá los resultados de los procesos automatizados:

- **Catálogo:** contiene el archivo con los campos necesarios para realizar las comparaciones y validar la existencia actual de las asignaturas.
- **Dataframe:** contiene un archivo que combina datos de múltiples archivos de entrada de periodos pasados, tras haberse realizado procesos de limpieza y transformación de los registros.
- **Dataframe Actual:** contiene el archivo después de haberse llevado a cabo la limpieza y transformación de datos para la comparación actual.
- **Predicción:** contiene un archivo que contiene las proyecciones de matrículas, NRC y cálculos de horas requeridos para el próximo periodo, basados en los datos procesados.

Figura 17

Organización de directorios.



Nota. Esta figura representa la estructura de directorios utilizada para organizar los archivos que serán procesados mediante de los scripts desarrollados en Python.

Procesos Python

Extracción. En esta fase se empleó el script “extraccion.py” para llevar a cabo una serie de procesos destinados a facilitar la extracción y limpieza de datos procedentes de los archivos ubicados tanto en el directorio que contiene los reportes de periodos pasados como en el directorio que contiene el reporte actual. Se implementaron funciones dinámicas de manera efectiva, las cuales se utilizaron en diferentes etapas del proceso general. A continuación, se describen los pasos llevados a cabo con cada reporte:

Paso 1: Extracción de datos. En primer lugar, se establece el directorio “Data/Input” para definir la ubicación de los archivos de entrada, donde se buscan todos los archivos con la extensión “.xlsx”. Estos archivos se cargan posteriormente en un Dataframe utilizando la biblioteca Pandas. Luego, se procede a iterar sobre cada uno de los archivos Excel, realizando las siguientes operaciones por cada uno de ellos:

1. Carga los datos del archivo Excel.
2. Extracción del periodo del reporte a partir del nombre del archivo.
3. Eliminación las primeras y las últimas 4 filas que tienen información no relevante para el análisis.
4. Establecimiento de encabezados de columnas siguiendo el formato de los reportes.
5. Eliminación de las columnas (14, 15 y 16) que contienen información de Edificio, Aula y Capacidad respectivamente, las cuales vienen vacías.
6. Reemplazo de los valores NaN con “X” en las columnas de la 16 a la 23, que contienen información sobre los días de clases, manteniendo aquellos que ya tienen el carácter correspondiente al día de la semana en que se imparten las clases.
7. Reemplazo de los valores que contienen “0” por “X” en la columna 4, que contiene información de los Docentes.

8. Creación de una lista llamada “columnas_eliminar” que contiene los nombres de las columnas en las que se desea eliminar filas con valores NaN, como “NIVEL FORMACION”, “HI”, “HF”.
Luego, se utiliza el método “dropna” para eliminar las filas que contienen con ese criterio, en al menos una de las columnas registradas en la lista.
9. Adición de una columna adicional para insertar el periodo del archivo del cual se está extrayendo la información.
10. Definición de los tipos de datos que tendrán las columnas para su posterior uso.

Actualización y procesamiento del catálogo de asignaturas para asegurar que la información utilizada en el análisis esté actualizada y sea precisa, para ello se realizan las siguientes operaciones:

1. Lectura de un archivo en formato “.xlsx” que se encuentra en el directorio “Data/Reporte Actual” y carga de los datos en un Dataframe.
2. Mapeo de códigos a los títulos que se tienen en los reportes que se van a analizar. Esto implica establecer equivalencias como 'COMP': 'CIENCIAS DE LA COMPUTACION', con el fin de tener la misma estructura en los archivos a analizar con el catálogo y poder realizar comparaciones.
3. Reemplazo de los valores de la columna “DPTO” con las descripciones mapeadas.
4. Renombrado del encabezado “TITULO” por “ASIGNATURA”.

Validación cruzada entre el archivo combinado y el catálogo actualizado de asignaturas para identificar y corregir asignaturas que ya no existan actualmente.

1. Carga del catálogo y el reporte combinado de los periodos en un Dataframe.
2. Validación para verificar que los códigos del catálogo sean los mismos que los presentes en el dataframe combinado.
3. Reemplazo de valores de CAMPUS solicitados en los requerimientos, en donde se solicitó integrar todos los campus asociados a Latacunga en uno solo.

Paso 2: Proyección de matrículas y NRC. El proceso de proyección de matrículas y NRC constituye una etapa crucial en la planificación académica y administrativa de la Universidad de las Fuerzas Armadas “ESPE”. A través del análisis de datos históricos recopilados en el Paso 1 y la aplicación de modelos predictivos, se persigue la anticipación de la demanda futura de los estudiantes para cada asignatura y en los distintos campus que integran la institución. Esta información anticipada permite optimizar la asignación de recursos, planificar la oferta académica y organizar una distribución adecuada de la carga académica. Para alcanzar estos objetivos, se emplea tres modelos para realizar proyecciones precisas y confiables: Regresión Lineal, Árboles de Decisión y Suavizamiento Exponencial. A continuación, se detallan los pasos que se involucran en la predicción de matrículas y NRC, siendo el mismo para ambos casos este proceso:

1. Carga de datos del archivo combinado que contiene la información consolidada de matrículas por asignatura, periodo y campus.
2. En la etapa de preprocesamiento de datos, se lleva a cabo la eliminación de los registros duplicados. Este paso es esencial para evitar contar varias veces las mismas matrículas, enfocándose en los registros que exponen el mismo número de estudiantes en el mismo NRC debido a la impartición de la asignatura en diferentes días a la semana.
3. Agrupamiento de datos por asignatura, periodo y campus. En ambos casos la agrupación se realiza para preparar los datos antes de aplicar modelos de predicción, sin embargo, existen diferencias puntuales en cada caso:
 - a. En el enfoque de las matrículas, se suman los números de estudiantes matriculados para cada combinación única de asignatura, periodo y campus.
 - b. En el enfoque de NRC, se cuentan el número de NRC para cada combinación única de asignatura, periodo y campus.

4. Modelamiento y predicción utilizando tres modelos diferentes para anticipar el número de estudiantes matriculados en el próximo periodo académico.

Árboles de Decisión: este modelo que se basa en la creación de árboles de decisión, donde múltiples arboles reciben un entrenamiento con diferentes submuestras del conjunto de datos cargado. Luego, realizan un promedio de las predicciones para mejorar la precisión. Es particularmente útil cuando existen interacciones son lineales entre las características y la variable de salida.

Regresión Lineal: basado en una relación lineal entre las variables predictoras y la variable de salida, se basa en identificar la línea recta óptima que mejor se adapte a los datos, minimizando la suma de los cuadrados de las diferencias. Es adecuado cuando las relaciones que se dan entre las variables son lineales.

Suavizamiento Exponencial: esta técnica está orientada a suavizar la serie de tiempo de matrículas, asignando un mayor peso a las observaciones más al considerar los datos históricos y el patrón de cambio reciente, resulta eficaz para capturar ciertas tendencias y estacionalidades en series de tiempo variantes.

5. Generación de resultados de predicciones para cada asignatura y campus utilizando los modelos antes mencionados.
6. Almacenamiento en dos archivos Excel con las predicciones de matrículas y NRC respectivamente.

Paso 3: Transformación y cálculos con datos. En el marco del proceso de análisis y gestión de datos para la planificación académica, el Paso 3 se presenta como una etapa importante donde se llevan a cabo una serie de operaciones cruciales para la entrega de información final. Estas operaciones son fundamentales para garantizar la precisión y la integridad de los datos, así como para habilitar el análisis y la toma de decisiones con la información recolectada.

1. Unión de los archivos generados en las predicciones, integrando dos conjuntos de datos que contienen la información sobre matrículas y NRC, consolidando toda la información relevante en un solo archivo.
2. Comparación de registros existentes en predicciones con el catálogo actualizado. Esta comparación permite validar las predicciones y realizar ajustes en función de la información más reciente y relevante disponible en el catálogo.
3. En el caso de existir un reporte actual, se realiza un procesamiento de los datos contenidos en el reporte actual, en el cual se llevan a cabo las mismas operaciones de limpieza que en los reportes encontrados en "Data/Input". Esto está incluyendo una normalización de datos y la eliminación de información redundante o no relevante.
 - a. Agrupación del dataframe actual con el archivo combinado que contiene información previamente consolidada. Esta agrupación proporciona una visión integral de todos los datos disponibles para el posterior análisis dentro de la herramienta Power BI.
4. Comparación del archivo combinado con el catálogo limpio. Para esto, se carga la información de cada archivo en dos dataframes y se realiza una comparación exhaustiva considerando los valores de Departamento, Asignatura y Código que se encuentren en los dos conjuntos de datos. Esta comparación permite identificar discrepancias para realizar ajustes necesarios en los datos combinados, asegurando coherencia y precisión más ajustada.
5. Para finalizar, se lleva a cabo una serie de cálculos y operaciones sobre los datos combinados y procesados. Esto incluye la determinación de horas mínimas y máximas de los cursos, así como la creación de una nueva columna para calcular el número de horas recibidas por asignatura para un posterior análisis y toma de decisiones.

Paso 4: Carga de la información en una base de datos. La carga de la información en una base de datos implica la transferencia de datos procesados y consolidados desde los archivos finales hacia una base de datos MySQL. Este proceso se lleva a cabo mediante el uso de funciones y clases que facilitan la lectura de archivos Excel, tanto la creación como la configuración e inserciones dentro de tablas específicas.

1. Lectura de los dos archivos procesados, el combinado y las predicciones, para convertir en dataframes, los cuales representarán la información procesada y consolidada que se va a cargar en la base de datos.
2. Por medio de un archivo config.ini se obtienen las credenciales de acceso a la base de datos, en donde se encuentran parámetros como el host, el usuario y la contraseña.
3. Creación de la conexión a la base de datos utilizando las credenciales obtenidas en el archivo de configuración mediante la librería "mysql.connector".
4. Creación de la base de datos llamada "planificación_academica" si no existe.
5. Definición de las estructuras de cada tabla que contendrán los datos a ser insertados. Cada estructura cuenta con su respectivo nombre de la tabla y las columnas dentro de la misma.
6. Inserción de datos contenidos en los dataframes en las tablas correspondientes a la base de datos. Esto se realiza por medio de la función definida para los dataframes llamada "to_sql", que permite especificar la tabla de destino y los tipos de datos de cada columna.

Paso 5: Empaquetado, Distribución y Pruebas del Ejecutable. Tras completar el desarrollo de solución en Python para el procesamiento integral de la información, llega el momento de automatizar estos procesos mediante la creación de una aplicación empaquetada con toda la lógica y bibliotecas utilizadas en los pasos previos. Con este fin, se ha empleado la librería PyInstaller, la cual facilita la creación de un archivo ejecutable capaz de llevar a cabo todos los pasos mencionados con tan solo utilizar esta simple aplicación.

Interfaz y API con Symfony

Mediante el framework Symfony, se emplea el lenguaje de programación PHP para simplificar y acelerar el proceso de desarrollo de aplicaciones web. Este framework se distingue por su arquitectura modular, compuesta por componentes reutilizables, lo que posibilita la construcción eficiente y escalable de aplicaciones robustas a largo plazo. El objetivo primordial al crear esta aplicación en Symfony es facilitar su integración con la aplicación Python empaquetada. Para lograr la integración efectiva entre la aplicación web y el ejecutable, se llevan a cabo los siguientes procesos:

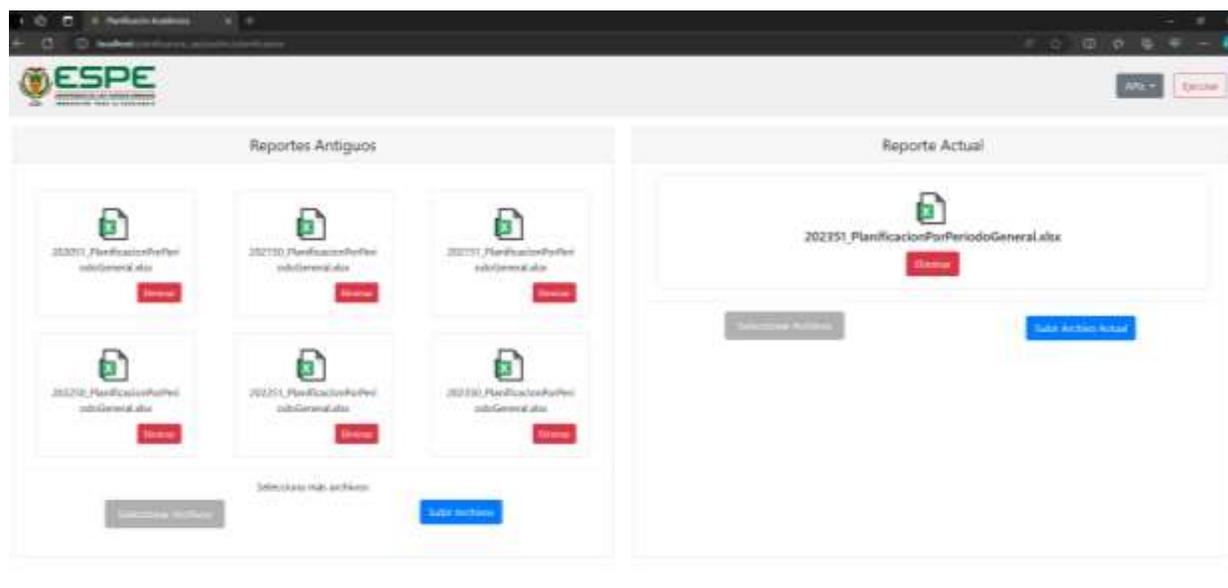
Interfaz de Usuario (UI)

Se desarrolló una interfaz de usuario interactiva utilizando Symfony. Esta interfaz permite a los usuarios del Vicerrectorado cargar archivos de reportes, lo que facilita la preparación para la ejecución de la aplicación empaquetada.

- **Uso de la Interfaz**

Figura 18

Interfaz de usuario.



Nota. La interfaz desarrollada utilizando Symfony, posibilita a los usuarios del Vicerrectorado cargar archivos de reportes, lo que simplifica la preparación para la ejecución de la aplicación empaquetada.

API

Se utilizaron los componentes disponibles en Symfony para crear endpoints que exponen la información del conjunto de datos combinado y las predicciones generadas por la aplicación empaquetada. Para lograr esto, se definieron controladores específicos para cada API y se empleó un serializador para convertir los datos al formato JSON antes de enviarlos como respuesta a las solicitudes HTTP recibidas.

- **API Combinado**

Figura 19

Extracción de datos – API Combinado.

```

1 {
2   {
3     "DEPARTAMENTO": "CIENCIAS DE ENERGIA Y MECANICA",
4     "CAMPUS": "ESPE SEDE LATACUNGA",
5     "AREA CONOCIMIENTO": "EHEC-0: PROCESOS DE MANUFACTURA",
6     "CODIGO ASIGNATURA": "ABMS - CONFORMADO MECANICO",
7     "MKC": "3105",
8     "STATUS": "A",
9     "NUM_EST": "30",
10    "HI": "12:00:00",
11    "HF": "13:59:00",
12    "L": "X",
13    "M": "X",
14    "T": "X",
15    "P": "N",
16    "V": "X",
17    "HORA_DIA": "2",
18    "NUM_DIAS": "3",
19    "HORAS": "2",
20    "TIPO": "CLAS",
21    "PERIODO": "202051",
22    "OBSERVACION": "MIN",
23    "PERIODO_FECHA": "2020-10-03"
24  },
25  {
26    "DEPARTAMENTO": "CIENCIAS DE ENERGIA Y MECANICA",
27    "CAMPUS": "ESPE SEDE LATACUNGA",
28    "AREA CONOCIMIENTO": "EHEC-0: PROCESOS DE MANUFACTURA",
29    "CODIGO ASIGNATURA": "ABMS - CONFORMADO MECANICO",
30    "MKC": "3105",
31    "STATUS": "A",
32    "NUM_EST": "30",
33    "HI": "14:00:00",
34    "HF": "15:59:00",
35    "L": "X",
36    "M": "X",
37    "T": "X",
38    "P": "N",
39    "V": "X",
40    "HORA_DIA": "2",
41    "NUM_DIAS": "3",
42    "HORAS": "2",
43    "TIPO": "CLAS",
44    "PERIODO": "202051",
45    "OBSERVACION": "MIN",
46    "PERIODO_FECHA": "2020-10-03"
47  },
48  {
49    "DEPARTAMENTO": "CIENCIAS DE ENERGIA Y MECANICA",
50    "CAMPUS": "ESPE SEDE LATACUNGA",
51    "AREA CONOCIMIENTO": "EHEC-3: DISEÑO MECANICA COMPUTACIONAL",
52    "CODIGO ASIGNATURA": "ABMS - DISEÑO MANUF E ING ASIST COMPU",
53    "MKC": "3023",
54    "STATUS": "A",
55    "NUM_EST": "20",
56    "HI": "09:30:00",
57    "HF": "11:30:00",
58    "L": "X",
59    "M": "X",
60    "T": "N",
61    "P": "N",
62    "V": "X",
63    "HORA_DIA": "2",
64    "NUM_DIAS": "2",
65    "HORAS": "4",
66    "TIPO": "CLAS",
67    "PERIODO": "202051",
68    "OBSERVACION": "MIN",
69    "PERIODO_FECHA": "2020-10-03"
70  }
71 }

```

Nota. Esta figura representa la estructura de datos obtenidos referente a reportes históricos mediante la API.

- **API Predicciones**

Figura 20

Extracción de datos – API Predicciones.

```

1  [
2  {
3    "CAMPUS": "ESPE SEDE LATACUNGA",
4    "DEPARTAMENTO": "CIENCIAS DE ENERGIA Y MECANICA",
5    "ASIGNATURA": "ING MANTENIMIENTO AUTOMOTRIZ",
6    "CODIGO": "A0010",
7    "AREA_CONOCIMIENTO": "EMEC-0 PROCESOS DE MANUFACTURA",
8    "CODIGO_ASIGNATURA": "A0010 - ING MANTENIMIENTO AUTOMOTRIZ",
9    "ESTUDIANTES_RL": 33,
10   "ESTUDIANTES_SE": 31,
11   "ESTUDIANTES_AD": 29,
12   "NRC_RL": 1,
13   "NRC_SE": 1,
14   "NRC_AD": 1,
15   "HORAS_RL": 4,
16   "HORAS_SE": 4,
17   "HORAS_AD": 4,
18   "OBSERVACION": "MIN"
19  },
20  {
21   "CAMPUS": "ESPE SEDE LATACUNGA",
22   "DEPARTAMENTO": "CIENCIAS DE ENERGIA Y MECANICA",
23   "ASIGNATURA": "MECANICA DE MATERIALES",
24   "CODIGO": "A0209",
25   "AREA_CONOCIMIENTO": "EMEC-2 MATERIALES Y MECANICA SOLIDOS",
26   "CODIGO_ASIGNATURA": "A0209 - MECANICA DE MATERIALES",
27   "ESTUDIANTES_RL": 40,
28   "ESTUDIANTES_SE": 41,
29   "ESTUDIANTES_AD": 41,
30   "NRC_RL": 1,
31   "NRC_SE": 1,
32   "NRC_AD": 1,
33   "HORAS_RL": 6,
34   "HORAS_SE": 6,
35   "HORAS_AD": 6,
36   "OBSERVACION": "MAX"
37  },
38  {
39   "CAMPUS": "ESPE SEDE LATACUNGA",
40   "DEPARTAMENTO": "CIENCIAS DE ENERGIA Y MECANICA",
41   "ASIGNATURA": "MECANICA DE MATERIALES",
42   "CODIGO": "A0210",
43   "AREA_CONOCIMIENTO": "EMEC-2 MATERIALES Y MECANICA SOLIDOS",
44   "CODIGO_ASIGNATURA": "A0210 - MECANICA DE MATERIALES",
45   "ESTUDIANTES_RL": 22,
46   "ESTUDIANTES_SE": 26,
47   "ESTUDIANTES_AD": 20,
48   "NRC_RL": 1,
49   "NRC_SE": 1,
50   "NRC_AD": 1,
51   "HORAS_RL": 4,
52   "HORAS_SE": 4,
53   "HORAS_AD": 4,
54   "OBSERVACION": "MIN"
55  },
56  {
57   "CAMPUS": "ESPE SEDE LATACUNGA"

```

Nota. Esta figura representa la estructura de datos obtenidos referente al reporte de proyecciones mediante la API.

Integración API Symfony con Power BI

Gracias a este proceso, se ofrece una automatización en la obtención de datos, logrando así una reducción en tiempo y esfuerzo al intentar la recolección de datos manualmente. Además, permite actualizaciones en tiempo real mediante programación desde la API, asegurando la actualización constante de reportes y predicciones.

Para aprovechar los datos proporcionados por la API del prototipo web desarrollado con Symfony. Dicha API ofrecerá dos conjuntos de datos:

Reportes históricos: Contiene información de los últimos 6 semestres (más el actual) sobre matrículas por carrera, campus, departamento y semestre.

Proyecciones: Incluye proyecciones para el siguiente semestre académico basadas en modelos predictivos como regresión lineal, árboles de decisión y suavizamiento exponencial.

Power BI se encargará de:

Conexión a la API:

- Utilizaremos la opción "Obtener datos desde web" en Power BI para establecer una conexión segura con la API RESTful del prototipo web.
- Configuraremos la conexión especificando la URL de la API.

Carga de datos:

Obtención de Datos de Reportes Históricos:

- Se establecerá un enlace directo desde Power BI hacia el endpoint de la API que contiene los reportes académicos de los últimos 6 semestres más el semestre actual.
- La aplicación de inteligencia empresarial realiza una petición a la API, lo que le permite recuperar datos necesarios sobre las matrículas, número de horas, NRC, entre otros datos, los cuales se encuentran divididos por Campus y Departamentos.
- Para poder generar un posterior análisis mediante visualizaciones, todo el conjunto de datos será cargado dentro de la herramienta de Power BI.

Obtención de Datos de Proyecciones:

- Se establece otro enlace desde Power BI hacia el endpoint de la API que contiene datos sobre las predicciones para el próximo periodo académico.

- Power BI va a realizar una solicitud a la API y recibirá datos con las predicciones generadas por los modelos predictivos como regresión lineal, árboles de decisión y suavizamiento exponencial.
- Estos datos de predicción se integrarán en Power BI como un conjunto de datos adicionales, que serán utilizados para mostrar proyecciones y tendencias futuras.

Modelado de datos en Power BI

Una vez obtenidos los datos históricos y las predicciones de matrículas mediante la API del prototipo web, el siguiente paso crucial es el modelado de datos en Power BI. Con el fin de mejorar el análisis, dentro de la herramienta procedemos a organizar y estructurar la información de forma eficiente.

Creación de tablas dimensionales

A partir de la tabla "Combinado" que contiene la información completa, se crearán seis tablas dimensionales:

- DEPARTAMENTO: Almacenará información sobre los departamentos académicos de la universidad.
- CAMPUS: Contendrá datos sobre los diferentes campus de la universidad.
- ÁREA DE CONOCIMIENTO: Agrupará las carreras por áreas de conocimiento.
- ASIGNATURA: Detalla las asignaturas que se ofrecen en la universidad.
- PERIODO: Registrará los semestres académicos.
- OBSERVACIÓN: Incluirá un comentario con respecto a las horas de la asignatura.

Conexiones de uno a varios:

Se establecerán relaciones de uno a varios entre las tablas dimensionales y la tabla "Combinado", también se establecerán relaciones similares entre las tablas dimensionales y la tabla "Predicciones". Esto significa que cada registro en la tabla "Combinado" y la tabla "Predicciones" estará relacionado con un solo registro en cada una de las tablas dimensionales.

Definición y cálculo de medidas en Power BI

Después de establecer el modelo de datos dentro de la aplicación, se debe definir las medidas con las cuales se van a realizar los diferentes tipos de análisis cuantitativos y obtener información valiosa.

Medidas a calcular:

1. Matrículas:

Con esta medida calculamos el número total de estudiantes matriculados en una asignatura, considerando la posibilidad de que un NRC (Número de Registro de Curso) tenga dos o más divisiones. La fórmula utilizada es la siguiente:

```

Matrículas =
SUMX(
VALUES(Combinado[NRC]),
CALCULATE(IF(MAX(Combinado[OBSERVACION]) = "Division", MAX(Combinado[NUM_EST]) / 2,
MAX(Combinado[NUM_EST])))
)

```

La observación “División” muestra que los estudiantes inscritos en una asignatura se dividen en diferentes grupos, pero no hay una distinción clara entre teoría y laboratorio en los datos.

Si la observación es “División”, se divide el número total de estudiantes por 2 para obtener el número “real” de estudiantes. De lo contrario, se suma el número total de estudiantes sin hacer ninguna división adicional.

Al dividir el número total de estudiantes por 2, se corrige el recuento para reflejar más precisamente el número real de estudiantes en cada grupo, considerando que cada estudiante va a estar presente tanto en la teoría como en el laboratorio.

Explicación de la fórmula:

- SUMX: Función que itera entre las filas de la tabla que calcula la suma de cada una de ellas.
- VALUES: Función que dentro de cada columna devolverá una lista única de valores.
- CALCULATE: Función que permite modificar el contexto de evaluación de una expresión.
- IF: Función que dependerá de una condición para determinar el valor que se va a devolver.
- MAX: Función que permite obtener el valor máximo de cada columna.

2. Total NRC:

Esta medida calcula el número total de NRC diferentes que se ofrecen en la universidad. La fórmula utilizada es la siguiente:

Total NRC = CALCULATE(DISTINCTCOUNT(Combinado[NRC]))

Explicación de la fórmula:

- DISTINCTCOUNT: Función que devuelve el número de valores únicos en una columna.
- CALCULATE: Función que permite modificar el contexto de evaluación de una expresión.

3. Campus:

La medida "Campus" en conjunto con la función SELECTEDVALUE permite crear títulos dinámicos que se adaptan a la selección del usuario. De esta manera, el título siempre refleja el contenido específico del análisis o la visualización que se está mostrando.

Campus = SELECTEDVALUE(Campus[CAMPUS], "TODAS LAS SEDES")

Implementación de Tableros

En la fase de implementación de tableros en Power BI, se visualizarán los datos procesados y calculados previamente para ofrecer una visión completa y detallada de la información académica solicitada. Estos tableros permiten al usuario final poder observar, modificar, analizar y comprender los diferentes datos de manera más dinámica, para la toma de decisiones en la Universidad de las Fuerzas Armadas ESPE.

Los elementos que se incluirán en los tableros son los siguientes:

Número de Matrículas, NRC y Horas por Departamento, Campus y Área de Conocimiento:

- Se mostrarán tablas dinámicas y gráficos de barras que presentarán el número de matrículas, NRC y horas, desglosados por departamentos, campus y área de conocimiento.

Número de Matrículas, NRC y Horas Proyectadas por Departamento, Campus y Área de**Conocimiento:**

- Se incluirán visualizaciones similares a las anteriores, pero mostrando las proyecciones para el próximo periodo académico.
- Estas proyecciones permitirán a los usuarios anticipar la demanda de matrículas y asignaturas.

Comparación del Periodo Académico Actual vs Proyectado:

- Se utilizarán tablas dinámicas y gráficos para comparar el número de matrículas, NRC y horas entre el periodo académico actual y el proyectado.

Gráfico de Líneas del Reporte Histórico de Matrículas:

- Se mostrará un gráfico de líneas que representará el reporte histórico de matrículas durante los últimos 6 periodos académicos, más el periodo actual.
- Este gráfico permitirá visualizar las tendencias en la cantidad de matrículas a lo largo del tiempo y evaluar el crecimiento o cambios en el número de matrículas.

Capítulo IV

Resultados

Anteriormente, el proceso de análisis de datos se llevaba a cabo utilizando la aplicación Excel ya que no se contaba con licencias para otras aplicaciones. Dicho proceso gastaba un tiempo considerable, donde en el mejor de los casos se requería un aproximado de dos a tres horas para realizar un análisis detallado. Sin embargo, este proceso se enfrentaba a la actualización diaria, lo que significaba que se debía volver a realizar el mismo proceso cada día mientras se iba actualizando la información.

Figura 22

Reporte de Matrículas Actuales.



Nota. En esta figura observamos el número de matrículas actuales en el Semestre 202351.

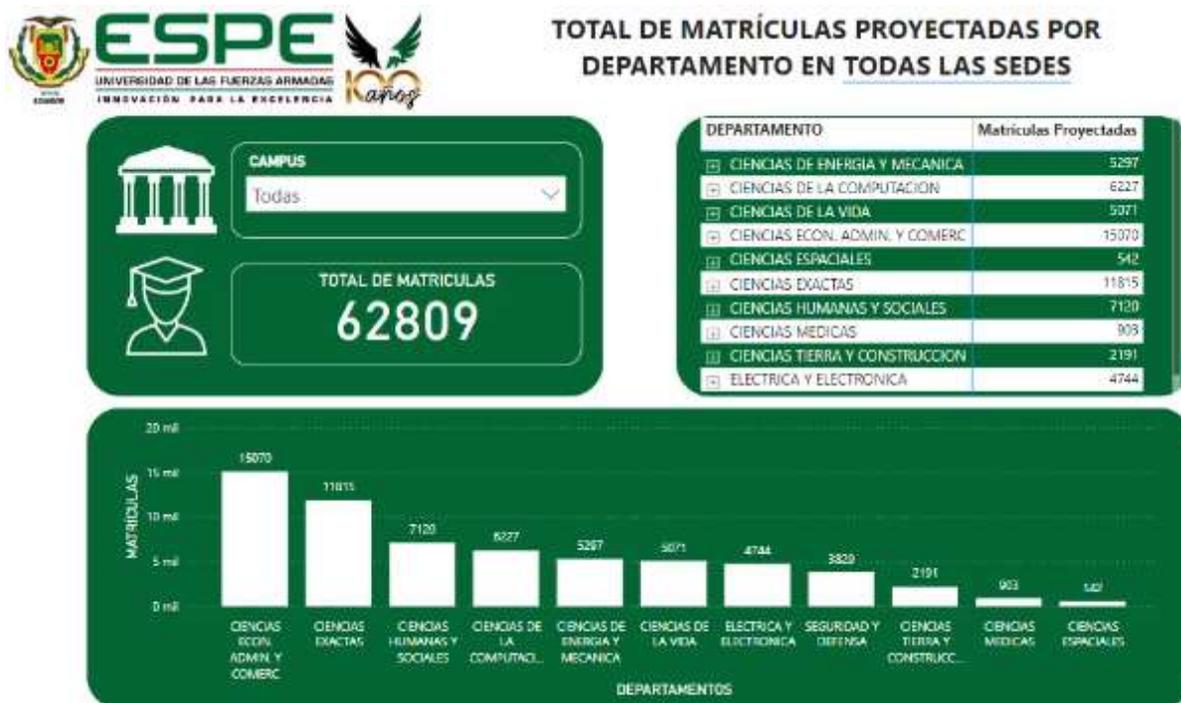
En la figura 22 podemos observar una representación clara y concisa de la distribución de las matrículas por departamento en las diferentes sedes de la Universidad de las Fuerzas Armadas ESPE. Dicha información se observa de una manera fácil de comprender. Por lo cual, el reporte generado es útil para los funcionarios del Vicerrectorado que necesitan información sobre la planificación.

Dentro de la figura se puede observar:

- Un filtro, Campus, el cual nos permite seleccionar cualquier campus o sede que tiene la ESPE, los datos irán cambiando de acuerdo a la opción que sea seleccionada dentro del filtro.
- Se tiene una tabla, en la cual se muestra el número total de matrículas por departamento, al ser una tabla dinámica podemos desglosar los departamentos y ver las asignaturas en dichos departamentos.
- La figura también incluye un gráfico de barras que muestra el número total de matrículas por departamento. El gráfico de barras ayuda a visualizar la distribución de las matrículas por departamento.
- Además, se cuenta con una etiqueta en donde muestra el número total de matrículas en todos los departamentos de las diferentes sedes.

Figura 23

Reporte de Matrículas Proyectadas.

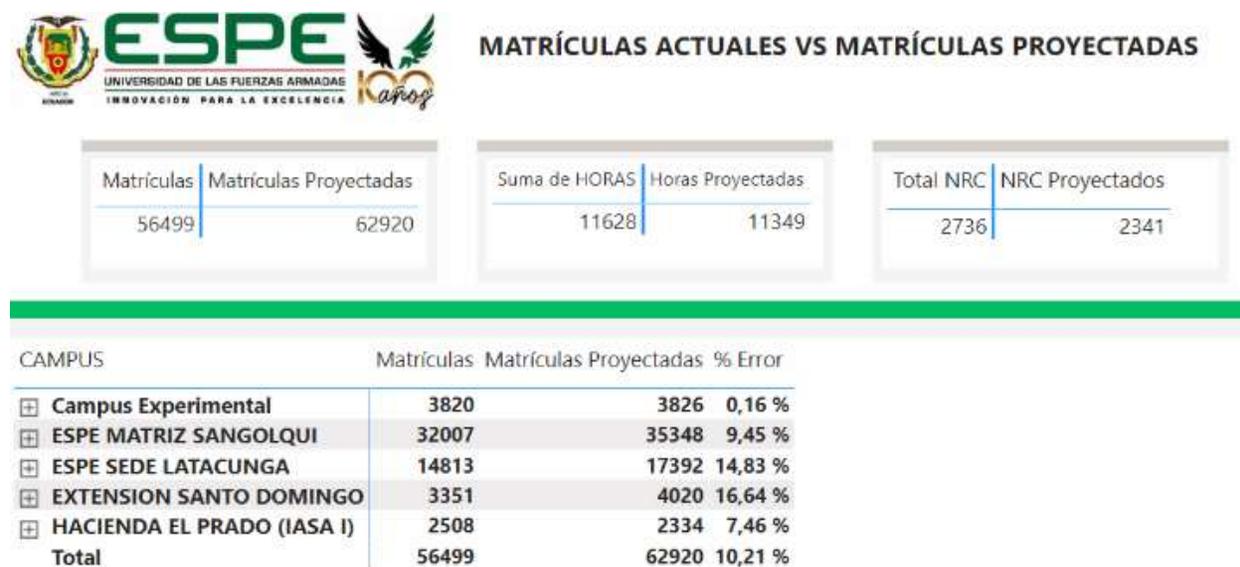


Nota. Esta figura representa el número de matrículas proyectadas para el Semestre 202450.

En la figura 23, se observa la misma distribución de información que la figura 22, sin embargo, los datos hacen referencia a las matrículas proyectadas para el siguiente periodo académico.

Figura 24

Comparación – Matrículas Actuales vs Matrículas Proyectadas.

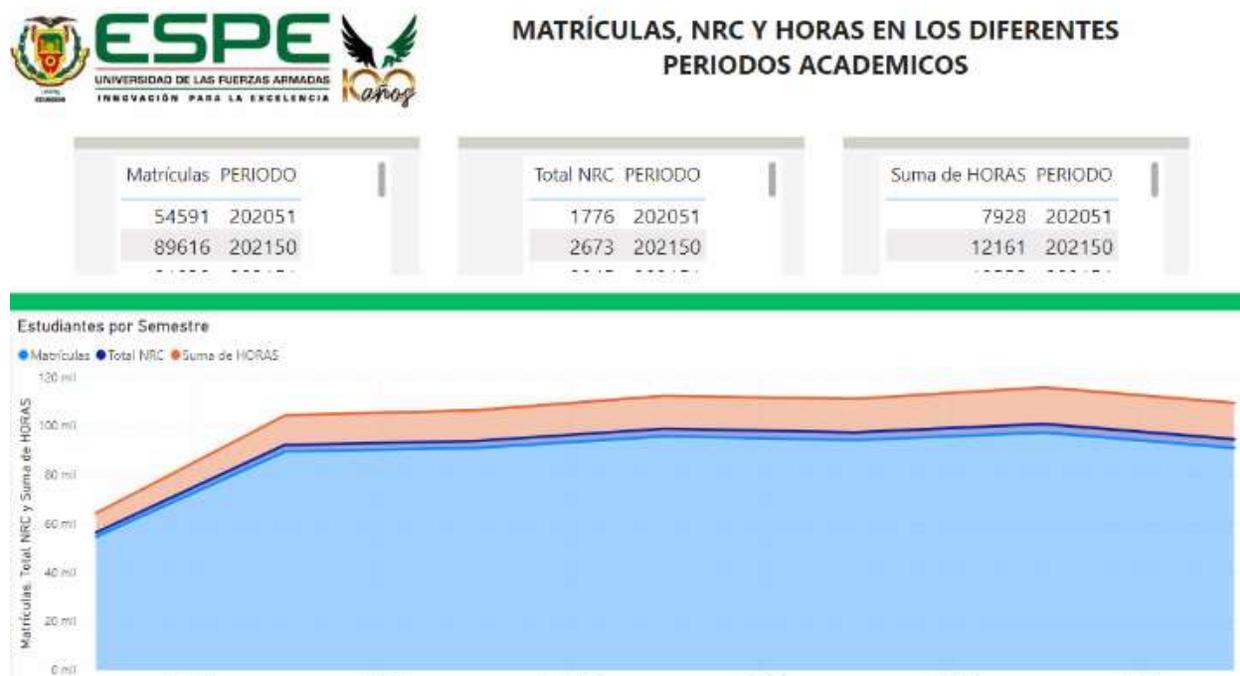


Nota. Esta figura representa la comparación directa entre el semestre actual del momento con el semestre siguiente.

Dentro de la figura 24 se establece una comparación directa entre matrículas actuales y matrículas proyectadas, esto con el fin de poder establecer la precisión del modelo de predicción. Esta comparación se realiza en una tabla dinámica, la cual nos permite observar una comparación desde el campus hasta una asignatura.

Figura 25

Gráfico de Líneas del Reporte Histórico de Matrículas.



Nota. Esta figura representa el número histórico de matrículas durante 6 semestres anteriores e incluido el actual.

En la figura 25 se observa un gráfico de líneas, donde se establece el crecimiento del número de matrículas a través de los últimos 6 semestres, esto permite ver el cambio en la cantidad de matrículas al pasar los semestres y poder ver una normalización, además, se cuenta con tablas donde se indica numéricamente el total de matrículas en cada periodo académico, al igual que el número de NRC y el total de horas.

Se ha logrado una significativa optimización gracias a la implementación del prototipo. Pues mediante una interfaz gráfica intuitiva, se carga el archivo Excel más actual y se ejecuta todo el proceso automáticamente dando un simple clic en "Ejecutar.exe". Después, al dar actualizar dentro de la herramienta Power BI, todos los reportes se actualizarán automáticamente con ello reduciendo el tiempo requerido a menos de 5 minutos.

Con la automatización de los procesos podemos observar los siguientes resultados:

- **Ahorro de Tiempo Significativo:** La automatización logró reducir drásticamente el tiempo que se necesitaba para realizar el análisis de datos. Pues, lo que antes llevaba horas, ahora se completa en minutos.
- **Actualización Automática de Datos:** Con la mínima intervención humana logramos que la información esté siempre actualizada y disponible para su análisis.
- **Mayor Eficiencia y Productividad:** Al ya no requerir las tareas manuales y repetitivas, logramos liberar tiempo y recursos que se puede dedicar a realizar análisis más detallado y mejorar la toma de decisiones.
- **Reducción de Errores Humanos:** Gracias a la automatización del proceso reducimos la intervención humana, por lo cual también se reduce la probabilidad de cometer errores humanos en la manipulación de la información, así mejorando la precisión y fiabilidad de los datos.

Resultados por Objetivos**Identificación de Requisitos y Necesidades:**

Mediante reuniones con el personal del Vicerrectorado, se identificaron los requisitos y necesidades específicas. Se comprendió las expectativas y requerimientos establecidos, lo que proporcionó una base sólida para el desarrollo del proyecto.

Metodología Ágil para la Gestión del Proyecto:

Se facilitó la colaboración entre el equipo de desarrollo y el personal del Vicerrectorado mediante la aplicación de la metodología ágil. Esto permitió una adaptación continua a medida que se obtenía retroalimentación y se aprendía más sobre las necesidades y requisitos del proyecto, logrando así una mayor flexibilidad y eficacia en la ejecución del proyecto.

Implementación de Tableros de Power BI:

Se implementaron tableros de Power BI con la finalidad de optimizar el análisis de datos y la generación de reportes. Estos tableros facilitaron la toma de decisiones basada en datos al proporcionar una visión clara y concisa de la información relevante.

Evaluación de la Efectividad y el Impacto:

Se evaluó la efectividad y el impacto de la implementación de los tableros de Power BI a través de diferentes métodos. Se compararon los reportes generados antes y después de la implementación, se recopiló opiniones del usuario sobre la usabilidad y la utilidad de los tableros. Todo esto evidenciando una gran mejora en la generación de reportes ya que se eliminó la necesidad de realizar procesos manuales y repetitivos cada que se necesitaba actualizar la información, realizando este mismo proceso de forma automática y en menor tiempo.

Capítulo V

Conclusiones y Recomendaciones

Conclusiones

- Se han cumplido todas los requisitos y necesidades indicadas por el personal del Vicerrectorado. Pues, gracias a la metodología utilizada, Scrum, mediante reuniones periódicas de sprint se logró atender y solventar los puntos críticos del proyecto.
- Se llegó a la conclusión que, el análisis de datos realizado apoya la selección de variables relevantes utilizadas en los modelos predictivos, ya que estos son cruciales para mejorar la capacidad de anticipación ante sucesos futuros y respaldarán la toma de decisiones.
- Se concluye de manera exitosa la implementación de modelos predictivos, ya que mediante un riguroso análisis y apoyados en el lenguaje de programación Python obtenemos resultados satisfactorios.
- Se ha utilizado la API desarrollada en Symfony para integración de datos con la herramienta Power BI. Gracias a esta integración se concluye que se obtiene mayor escalabilidad además de obtener una flexibilidad en la actualización de datos.
- Se llegó a la conclusión que la implementación de tableros en la herramienta de Power BI ayuda a optimizar el análisis de datos mediante reportes, por lo cual, ayuda a la toma de decisiones basada en los datos mostrados.
- Mediante la encuesta realizada se concluye una reducción significativa en los tiempos requeridos para el análisis de datos, además de mejorar el acceso a los datos actualizados y reduciendo el porcentaje de errores humanos que puede surgir durante el proceso.

Recomendaciones

- Se recomienda explorar las capacidades de la herramienta Power Bi para estar preparados para generar más informes en el futuro. Esto ayudará a tener una adaptación ágil a las nuevas necesidades para poder garantizar una respuesta eficaz a los nuevos requisitos.
- Se recomienda tener en cuenta que a medida que aumenta el número de reportes, también aumentará el tiempo de respuesta. Por lo cual será necesario priorizar los reportes que tengan mayor valor para poder evitar tiempos de respuestas más largos y poder garantizar la eficiencia en los diferentes procesos.

Bibliografía

Abhinav. (3 de Abril de 2023). *Building ETL using Python*. Obtenido de SarasAnalytics:

<https://sarasanalytics.com/blog/etl-using-python/>

Cedeño Coello, C. A. (2023). Implementación de un BI, que permita analizar la data del CRM Zendesk, mediante la herramienta Power BI, para medir indicadores de servicio y satisfacción al cliente en el área de customer success de la empresa Fintech PeiGo S.A. (*Proyecto de titulación*).

Universidad de Guayaquil, Guayaquil, Ecuador.

Davidseminger. (22 de Noviembre de 2023). *Introducción a los flujos de datos y la preparación de datos de autoservicio - Power BI*. Obtenido de <https://learn.microsoft.com/es-es/power-bi/transform-model/dataflows/dataflows-introduction-self-service>

Díaz, J. C. (2010). *Introducción al Business Intelligence*. Editorial UOC.

Ebner, J. (2 de Mayo de 2023). *How to use the Sklearn Predict Method*. Obtenido de Sharp Sight:

<https://www.sharpsightlabs.com/blog/sklearn-predict/>

González Benítez, N., & Perdigón Llanes, R. (2020). Una revisión bibliográfica sobre modelos para predecir las producciones de leche. *Revista Ingeniería Agrícola*, 69-77.

Gutiérrez-Cabello, G. S. (2020). *Modelado de datos y automatización del proceso ETL para dar apoyo a la metodología NAUTIA desarrollada por la Plataforma UPM sobre Refugiados*. Universidad Politécnica de Madrid, Madrid.

Korstanje, J. (31 de Julio de 2023). *How to select a model for your Time Series Prediction Task [Guide]*.

Obtenido de Neptune.ai: <https://neptune.ai/blog/select-model-for-time-series-prediction-task>

Marqués, M. P. (2015). *Business Intelligence. Técnicas, herramientas y aplicaciones*. Alfaomega RC Libros.

Microsoft. (2023). *Power BI para Empresas*. Obtenido de <https://powerbi.microsoft.com/es-es/enterprise/>

Pacci Ayala, C. F. (2017). Aplicación de inteligencia de negocios de autoservicio para la mejora de la toma de decisiones en una PYME en la región de Tacna. (*Tesis para optar el título profesional de Ingeniero de Sistemas*). Aplicación de inteligencia de negocios de autoservicio para la mejora de la toma de decisiones en una PYME en la región de Tacna, Tacna, Peru. Obtenido de <http://hdl.handle.net/20.500.12969/165>

Santamaría Valero, B. S. (2021). Aplicación práctica de los datos abiertos: actualización y predicción. (*Trabajo de Fin de Grado*). Universidad de La Laguna, La Laguna, España. Obtenido de <http://riull.ull.es/xmlui/handle/915/24745>

Symfony. (s.f.). *Symfony Cookbook*. Obtenido de <https://symfony.com/doc/current/cookbook/index.html>

Symfony. (s.f.). *Symfony Documentation*. Obtenido de <https://symfony.com/doc/current/index.html>

Viñals, J. T. (2011). *Ventajas y retos del Cloud Computing*. Libros de Cabecera.

Apéndices