



PLAN DE TRABAJO DEL ESTUDIANTE



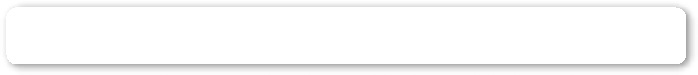
**DATOS DEL ESTUDIANTE**

|  |  |  |  |
| --- | --- | --- | --- |
| Apellidos y Nombres: | Flores Quispe Duilio Omar | ID: | 1427705 |
| Dirección Zonal/CFP: | Tacna |  |  |

Carrera: Ing. De Software con Inteligencia Artificial Semestre: IV Curso/ Mód. Formativo: Módulos y Paquetes para Machine Learning con Python

Tema de Trabajo Final: Aplicación de Módulos y Paquetes en Python para Machine Learning en un

Contexto Empresarial



**1. INFORMACIÓN**

* **Identifica la problemática del caso práctico propuesto.**

La problemática central de la empresa de análisis financiero es su ineficiencia operativa y estancamiento tecnológico al manejar grandes volúmenes de datos. Ya que su dependencia de métodos manuales para procesar, analizar y visualizar archivos CSV no solo es lenta y propensa a errores, sino que además les impide adoptar técnicas avanzadas como el Machine Learning y el Deep Learning. Esta carencia de una metodología robusta y automatizada limita severamente su capacidad para generar predicciones de mercado precisas y tomar decisiones estratégicas basadas en datos, dejándolos en una clara desventaja competitiva.

* **Identifica propuesta de solución y evidencias.**

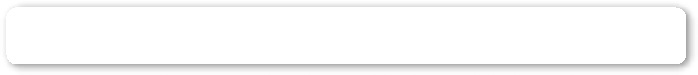
Para solucionar la problemática, se propone implementar un sistema integral de análisis predictivo en Python que automatice el flujo de trabajo, desde la ingesta de datos con Pandas hasta la visualización con Matplotlib y Seaborn.

Como evidencia, se construirá un modelo de clasificación con Scikit-learn para pronosticar tendencias del mercado, se desarrollará una red neuronal avanzada con TensorFlow y Keras para el análisis de series temporales, y se aplicarán técnicas de NLP con NLTK para el análisis de sentimiento de reportes financieros. Además, se generarán visualizaciones claras con Matplotlib y Seaborn, como gráficos comparativos y líneas de tendencia, y todo el proceso será consolidado en un informe técnico que documente los procedimientos y valide los resultados obtenidos.

* **Respuestas a preguntas guía**

**Durante el análisis y estudio del caso práctico, debes obtener las respuestas a las interrogantes:**

|  |  |
| --- | --- |
| Pregunta 01: | ¿Cómo se pueden optimizar las estructuras de datos utilizando Pandas y Numpy? |
| La optimización se logra asignando tipos de datos específicos en Pandas, como **datetime** para fechas o **category** para textos repetitivos, lo que reduce drásticamente el uso de memoria y acelera los cálculos, también que se debe realizar una limpieza para manejar valores nulos o inconsistentes que podrían ralentizar los cálculos o generar errores.  Además, se aprovecha la vectorización de NumPy para ejecutar operaciones matemáticas sobre conjuntos de datos enteros de forma casi instantánea, eliminando los bucles lentos de Python y haciendo que la preparación de datos y la ingeniería de características sean procesos altamente eficientes. | |
| Pregunta 02: | ¿Qué técnicas de Machine Learning pueden aplicarse para mejorar la predicción de tendencias de mercado? |
| Se pueden aplicar modelos de clasificación supervisada como **Random Forest** o Máquinas de Soporte Vectorial, que analizan simultáneamente múltiples variables para predecir la dirección del mercado.  La clave es combinar estos algoritmos con técnicas robustas de preprocesamiento y validación cruzada para asegurar que el modelo no solo memorice datos pasados, sino que pueda generalizar sus predicciones a condiciones de mercado futuras, resultando en pronósticos más fiables en datos futuros no vistos. | |
| Pregunta 03: | ¿Cómo se puede extraer información valiosa de reportes financieros utilizando NLP? |
| La aplicación de NLP en reportes financieros permite cuantificar el sentimiento del mercado a través de un proceso que inicia con la limpieza y preprocesamiento del texto (**tokenización**, eliminación de stopwords y normalización de palabras).  Luego, mediante léxicos financieros o modelos de NLP pre-entrenados, se asigna una puntuación de polaridad (positiva, negativa o neutra) a cada reporte, generando un valor numérico que refleja el sentimiento general. Esta variable cuantitativa puede incorporarse a modelos de Machine Learning, enriqueciendo las características y mejorando la precisión en la predicción de tendencias financieras. | |
| Pregunta 04: | ¿Cómo influye el uso de Deep Learning en la toma de decisiones empresariales? |
| El Deep Learning, a través de modelos como las redes LSTM, influye en las decisiones empresariales al ofrecer pronósticos mucho más precisos sobre series temporales complejas.  Al ser capaces de capturar patrones y dependencias a largo plazo que otros modelos ignoran, reducen la incertidumbre y permiten a la empresa tomar decisiones estratégicas más informadas y proactivas, como la optimización de inversiones y una gestión de riesgos más eficaz. | |
| Pregunta 05: | ¿Qué herramientas de visualización pueden mejorar la interpretación de datos financieros? |
| Las herramientas de visualización como Matplotlib y Seaborn mejoran drásticamente la interpretación de datos financieros al transformar tablas numéricas complejas en gráficos intuitivos.  Mediante la creación de visualizaciones como gráficos de líneas para seguir tendencias, histogramas para entender distribuciones y mapas de calor para analizar correlaciones, permiten identificar patrones y anomalías de forma rápida, facilitando una comunicación clara de los hallazgos y una toma de decisiones más ágil. | |



**2. PLANIFICACIÓN DEL TRABAJO**

* **Cronograma de actividades:**

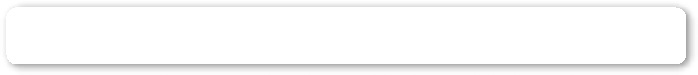
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **N°** | **ACTIVIDADES** | **CRONOGRAMA** | | | | | |
|  |  |  |  |  |  |
| 1 | PLANIFICAR | 18/08 | 22/08 |  |  |  |  |
| 2 | INFORMAR | 23/08 | 26/08 |  |  |  |  |
| 3 | DECIDIR | 27/08 | 30/08 |  |  |  |  |
| 4 | REALIZAR | 31/08 | 06/09 |  |  |  |  |
| 5 | CONTROLAR | 07/09 | 08/09 |  |  |  |  |
|  |  |  |  |  |  |  |  |

* **Lista de recursos necesarios:**

|  |  |
| --- | --- |
| **1. MÁQUINAS Y EQUIPOS** | |
| **Descripción** | **Cantidad** |
| Ordenador (PC) | 1 |
| Celular | 1 |
| Monitor (24”) | 1 |
| Usb (16 GB) | 1 |

|  |  |
| --- | --- |
| **2. HERRAMIENTAS E INSTRUMENTOS** | |
| **Descripción** | **Cantidad** |
| Lenguaje de Programación Python | Múltiple |
| Librerías de análisis de datos (Pandas, NumPy) | Múltiple |
| Framework de Machine Learning (Scikit-learn) | Múltiple |
| Frameworks de DL (TensorFlow, Keras) | Múltiple |
| Git y GitHub | 1 |
| Librerías de NLP (NLTK, SciPy) | Múltiple |
| Librerías de visualiz. (Matplotlib, Seaborn) | Múltiple |
| Fuentes de datos abiertos (Kaggle) | Múltiple |

|  |  |
| --- | --- |
| **3. MATERIALES E INSUMOS** | |
| **Descripción** | **Cantidad** |
| Lentes | 1 |
| Botella de agua | 8 |
| Cargador | 1 |
| Audífonos | 2 |



**3. DECIDIR PROPUESTA**

* **Describe la propuesta determinada para la solución del caso práctico**

**PROPUESTA DE SOLUCIÓN**

## Como parte de las estrategias para optimizar el análisis financiero y fortalecer la toma de decisiones basada en datos, se propone la implementación de un flujo de trabajo integral de Machine Learning en Python. Esta solución busca reemplazar el actual manejo manual y poco eficiente de datos en formato CSV, el cual es vulnerable a errores, consume mucho tiempo y limita la capacidad de la empresa para realizar predicciones de mercado sofisticadas y precisas.

**Esta solución está desarrollada utilizando un ecosistema de librerías especializadas como Pandas y NumPy para la manipulación de datos, Scikit-learn para la construcción de modelos predictivos y TensorFlow/Keras para el desarrollo de redes neuronales avanzadas. A través de scripts estructurados y reproducibles, se automatizará la limpieza de datos, la creación de un modelo de clasificación para predecir tendencias y un modelo de series temporales para pronósticos avanzados. Además, se integrarán técnicas de NLP con NLTK para analizar reportes financieros y se crearán visualizaciones interactivas con Matplotlib y Seaborn, generando así un sistema de análisis robusto, auditable y escalable.**

Primero, comencé estableciendo el entorno de desarrollo con Python y Google Colab, lo que me permitió un análisis interactivo y modular. Inicialicé el proyecto y procedí a la ingesta de los datasets de mercado (en formato CSV) utilizando la librería Pandas, cargando los datos en un DataFrame para su manipulación.

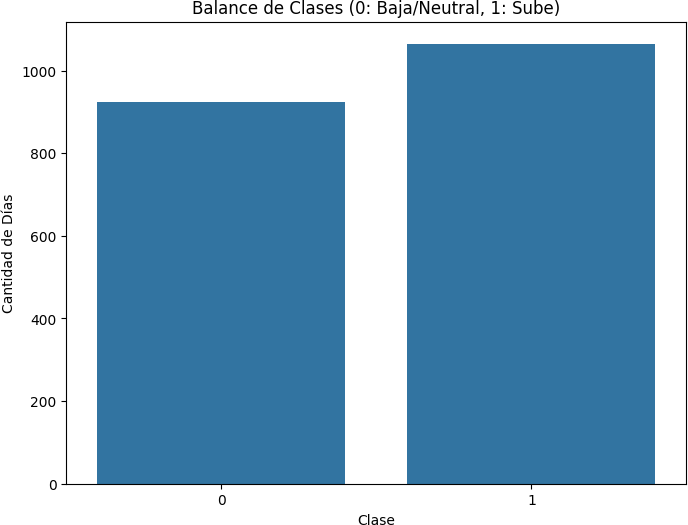
Luego, realicé el preprocesamiento y la limpieza de los datos. Me enfoqué en manejar valores nulos, convertir las columnas de fecha a un formato datetime para facilitar el análisis de series temporales y verificar la consistencia de los datos. Simultáneamente, utilicé Matplotlib y Seaborn para llevar a cabo un Análisis Exploratorio de Datos (EDA), visualizando distribuciones de precios y volúmenes para identificar patrones iniciales.

Después, implementé la ingeniería de características (feature engineering) para enriquecer el dataset. Usando NumPy y Pandas, calculé indicadores técnicos clave como medias móviles, el Índice de Fuerza Relativa (RSI) y las Bandas de Bollinger, que servirían como variables predictoras para los modelos.

A continuación, desarrollé el modelo de clasificación para predecir tendencias (alcista o bajista) con Scikit-learn. Dividí los datos en conjuntos de entrenamiento y prueba, normalicé las características y entrené un modelo de Random Forest. Evalué su rendimiento utilizando una matriz de confusión, y métricas como la precisión (accuracy) y F1-score.

Después, creé el componente de Procesamiento de Lenguaje Natural (NLP). Utilizando NLTK, procesé un conjunto de datos de reportes financieros para tokenizar el texto y aplicar análisis de sentimiento. El puntaje de sentimiento resultante se incorporó como una nueva característica en el dataset principal para enriquecer los modelos predictivos.

# 5



Luego, construí el modelo de Deep Learning con TensorFlow y Keras. Diseñé una red neuronal recurrente (LSTM) especializada en el análisis de series temporales. Preparé los datos en secuencias temporales y entrené el modelo para predecir el precio de cierre de los días siguientes, evaluando su precisión con el Error Cuadrático Medio (RMSE).

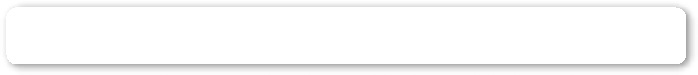
Para mejorar la interpretación de los resultados, integré visualizaciones que comparan los precios reales con las predicciones de ambos modelos. Creé gráficos de líneas de tendencia y gráficos de dispersión que facilitan la evaluación del rendimiento de los modelos.

Finalmente, realicé pruebas de todo el flujo de trabajo para asegurar su coherencia y reproducibilidad. Ajusté los hiperparámetros de los modelos para optimizar su rendimiento y consolidé todos los procedimientos, código y conclusiones en un informe técnico detallado.

PDTA: Para mí, el aspecto más complejo fue la preparación de los datos para la red neuronal LSTM, ya que la transformación de una serie temporal en secuencias supervisadas requiere una lógica cuidadosa para no filtrar información futura. La estructura de preprocesamiento y validación es conceptualmente similar para ambos modelos, pero la implementación y el ajuste fino del modelo de Deep Learning demandaron una mayor experimentación.

Dataset de Kaggle: https:/[/www.kaggle.com/datasets/aaron7sun/stocknews](http://www.kaggle.com/datasets/aaron7sun/stocknews)

Google Colab: <https://colab.research.google.com/drive/198JO6Ty1G5Fwg3f69Q5ZsnfUbArtW37L?usp=sharing>

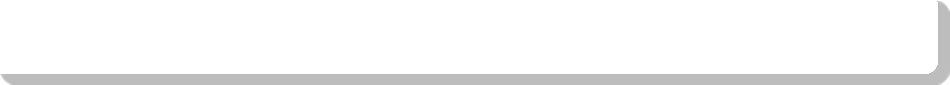


**4. EJECUTAR**

* **Resolver el caso práctico, utilizando como referencia el problema propuesto y las preguntas guía proporcionadas para orientar el desarrollo.**
* **Fundamentar sus propuestas en los conocimientos adquiridos a lo largo del curso, aplicando lo aprendido en las tareas y operaciones descritas en los contenidos curriculares.**

# **INSTRUCCIONES:** Ser lo más explícito posible. Los gráficos ayudan a transmitir mejor las ideas. Tomar en cuenta los aspectos de calidad, medio ambiente y SHI.

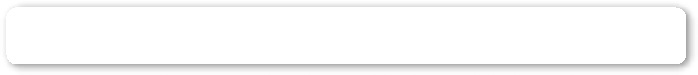
|  |  |
| --- | --- |
| **OPERACIONES / PASOS / SUBPASOS** | **NORMAS TÉCNICAS - ESTANDARES / SEGURIDAD /**  **MEDIO AMBIENTE** |
| Definición del entorno de desarrollo | Buenas prácticas de gestión de entornos |
| Carga de datos desde fuente CSV con Pandas | Metodología CRISP-DM (Entendimiento de datos) |
| Análisis Exploratorio de Datos (EDA) con Matplotlib y Seaborn | Principios de visualización de datos efectiva |
| Limpieza y preprocesamiento de los datos | Estándares de calidad de datos (consistencia) |
| División de datos en conjuntos de entrenamiento y prueba | Protocolos de Hold-out, Cross-validation |
| Entrenamiento del modelo de clasificación (Random Forest) | Metodología CRISP-DM (Modelado) |
| Evaluación del modelo con métricas (Accuracy, Matriz de Confusión) | Métricas de evaluación estándar para clasificación |
| Escalado y normalización de datos con Scikit-learn | Preprocesamiento y normalización de datos |
| Diseño de la arquitectura de la red neuronal (Keras/TensorFlow) | Buenas prácticas diseño de arquitec. neuronales |
| Evaluación del rendimiento del modelo de Deep Learning | Optimización de hiperparámetros |
| Escalado y normalización de datos con Scikit-learn | Principios de IA responsable y ética |
| Interpretación de resultados y comparación de modelos | Metodologías de reporte técnico |
| Optimización del uso de recursos computacionales (CPU/GPU en Colab) | Uso eficiente de hardware y gestión de memoria |
| Codificación y organización modular en el Notebook | Principios de codificación limpia y Clean Code |
| Documentación del código, procesos y conclusiones | Estándares de documentación de código |
|  |  |
|  |  |



**DIBUJO / ESQUEMA / DIAGRAMA DE PROPUESTA**

**(Adicionar las páginas que sean necesarias)**

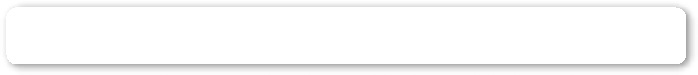
|  |  |  |
| --- | --- | --- |
|  | | |
|  | [Análisis Comparativo de Modelos de Machine Learning y Deep Learning] | |
| [Flores Quispe Duilio Omar] | [1:1] |



**5. CONTROLAR**

* + **Verificar el cumplimiento de los procesos desarrollados en la propuesta de solución del caso práctico.**

|  |  |  |
| --- | --- | --- |
| **EVIDENCIAS** | **CUMPLE** | **NO CUMPLE** |
| * ¿Se identificó claramente la problemática del caso práctico? | ☒ | ☐ |
| * ¿Se desarrolló las condiciones de los requerimientos solicitados? | ☒ | ☐ |
| * ¿Se formularon respuestas claras y fundamentadas a todas las preguntas guía? | ☒ | ☐ |
| * ¿Se elaboró un cronograma claro de actividades a ejecutar? | ☒ | ☐ |
| * ¿Se identificaron y listaron los recursos (máquinas, equipos, herramientas, materiales) necesarios para ejecutar la propuesta? | ☒ | ☐ |
| * ¿Se ejecutó la propuesta de acuerdo con la planificación y cronograma establecidos? | ☒ | ☐ |
| * ¿Se describieron todas las operaciones y pasos seguidos para garantizar la correcta ejecución? | ☒ | ☐ |
| * ¿Se consideran las normativas técnicas, de seguridad y medio ambiente en la propuesta de solución? | ☒ | ☐ |
| * ¿La propuesta es pertinente con los requerimientos solicitados? | ☒ | ☐ |
| * ¿Se evaluó la viabilidad de la propuesta para un contexto real? | ☒ | ☐ |



**6. VALORAR**

* + **Califica el impacto que representa la propuesta de solución ante la situación planteada en el caso práctico.**

|  |  |  |  |
| --- | --- | --- | --- |
| **CRITERIO DE EVALUACIÓN** | **DESCRIPCIÓN DEL CRITERIO** | **PUNTUACIÓN MÁXIMA** | **PUNTAJE CALIFICADO POR EL**  **ESTUDIANTE** |
| Identificación del problema | Claridad en la identificación del problema planteado. | 3 | 3 |
| Relevancia de la propuesta de solución | La propuesta responde adecuadamente al problema planteado y es relevante para el contexto del caso práctico. | 8 | 7 |
| Viabilidad técnica | La solución es técnicamente factible, tomando en cuenta los recursos y conocimientos disponibles. | 6 | 5 |
| Cumplimiento de Normas | La solución cumple con todas las normas técnicas de seguridad, higiene y medio ambiente. | 3 | 2 |
| **PUNTAJE TOTAL** | | **20** | **17** |

11

