

MODELO DE PRONÓSTICOS DE VENTAS PARA EMPRESA PRODUCTORA DE TANQUES ROTOMOLDEADOS

Integrantes:

Efraín Andrés Peña

Maby Jiseth Cano

Juan Esteban Pinzón

Esteban Molina

Santiago Lopez

1. Definición del Proyecto	3
1.1 Identificación del objetivo del Proyecto	3
1.2 Alcance del Proyecto	3
1.3 Metas y Resultados	4
2. Planificación del proyecto	4
2.1 Planificación del proyecto	4
2.2 Recursos	5
2.3 Presupuesto	5
2.3 Identificación de riesgos	5

1. Definición del Proyecto

1.1 Identificación del objetivo del Proyecto

Objetivo General: Desarrollar un modelo de pronóstico de ventas basado en técnicas de Machine Learning y Deep Learning para predecir con precisión la demanda futura, optimizando la producción y reduciendo costos asociados al desbalance entre oferta y demanda.

Objetivos Específicos:

- Seleccionar variables internas y externas que influyan significativamente en las ventas para incorporarlas al modelo de predicción.
- Implementar y comparar modelos de Machine Learning y Deep Learning, evaluando su rendimiento mediante métricas como MAE, RMSE y MAPE.
- Aplicar técnicas de ajuste de hiperparámetros y regularización para mejorar la precisión del modelo final.
- Desarrollar una herramienta piloto que permita realizar pronósticos de ventas y visualizar los resultados de forma interactiva.

Problema Identificado

- Métodos de pronóstico actuales de la empresa se basan en promedios y datos históricos sin considerar variaciones dinámicas externas (clima, mercado, economía, etc.).
- Baja precisión de los métodos tradicionales, generando sobreproducción, inventarios o ventas perdidas.

1.2 Alcance del Proyecto

Incluido:

1. **Análisis Exploratorio de Datos (EDA):**
 - Limpieza y normalización.
 - Análisis estadístico descriptivo.
 - Visualización de tendencias históricas.
2. **Desarrollo de Modelos:**
 - Comparación entre modelos clásicos (ARIMA, SARIMA) y avanzados (MLP, LSTM, Random Forest, XGBoost).
3. **Evaluación de Modelos:**
 - Cálculo de métricas: MAE, RMSE, MAPE, etc.
 - Validación cruzada para evitar sobreajuste.
4. **Implementación Inicial:**
 - Prototipo de interfaz gráfica para visualizar pronósticos.
 - Recomendaciones basadas en los resultados del modelo.
5. **Estudio de Factibilidad Económica:**

- Definición de costos asociados al desarrollo e implementación.
- Propuesta de modelo de negocio (suscripción, venta del modelo).

Excluido:

- Desarrollo completo de una interfaz frontend robusta.
- Personalización del modelo para sectores distintos al de la empresa.
- Escalabilidad a otros países o regiones sin datos específicos.

1.3 Metas y Resultados

Metas SMART (Específicas, Medibles, Alcanzables, Relevantes, Temporales):

- Identificar variables externas que afecten las ventas (clima, precios, tendencias de mercado, etc.) para incluir en el modelo.
- Entrenar y comparar modelos basados en Machine Learning y Deep Learning.
- Utilizar técnicas de validación cruzada y definir métricas clave de evaluación.
- Documentar todo el proyecto y proponer mejoras futuras.

2. Planificación del proyecto

2.1 Planificación del proyecto

1. Crear un plan de proyecto

a. Cronograma

- Lista de tareas y actividades

- Identificación de requisitos de datos.
- Análisis exploratorio.
- Normalización de la base de datos.
- Evaluación de factibilidad y estudio económico
- Definir CANVAS.
- Selección de técnicas de Machine Learning y Deep Learning.
- Desarrollar el modelo.
- Evaluar el rendimiento.
- Implementación de producción.

- Fechas de inicio y fin:

Inicio: 2/11/2024 - Fin: 29/11/2024

- Dependencias entre tareas:

- Identificación de requisitos de datos
- Análisis exploratorio (y_data_profiling)
- Normalización de la base de datos
- Evaluación de factibilidad y estudio económico
- Selección de técnicas de ML/DL

- Desarrollo del modelo
- Evaluación del rendimiento
- Implementación de producción

2.2 Recursos

Project Manager (Efraín Andrés Peña): Coordinación del proyecto, supervisión del cronograma y entrega de resultados.

Data Scientist (Esteban Molina): Análisis de datos y preprocesamiento.

Machine Learning Engineer (Juan Esteban Pinzón): Diseño e implementación de modelos.

Optimization Specialist (Santiago López): Ajuste y optimización de hiperparámetros.

Evaluation Analyst (Maby Jiseth Cano): Evaluación del rendimiento y generación de informes.

Requerimientos Técnicos:

- Herramientas: Python, Jupyter Notebook, TensorFlow/PyTorch, Scikit-learn.
- Visualización: Power BI, Streamlit

2.3 Presupuesto

Recurso	Costo en COP (aproximado)	Notas
Computación en la nube	2000000	Servidores con GPU (200 horas).
Salarios del equipo (5 personas)	25000000	5 integrantes durante un mes, a \$5,000,000 COP cada uno.
Otros gastos (reuniones, imprevistos)	1200000	Gastos generales para logística.

2.3 Identificación de riesgos

Riesgos Potenciales:

- Datos insuficientes o inconsistentes que puedan afectar la precisión del modelo.
- Dificultad en correr modelos complejos debido a los altos requerimientos de hardware.

- Modelos que no se ajusten correctamente a los datos, generando predicciones poco confiables.
- Por factores como falta de datos relevantes, elección incorrecta de modelo o inadecuado preprocesamiento.
- Recursos computacionales elevados que dificulten la ejecución frecuente del modelo.

Plan de Contingencia

- 1. Para la baja calidad de los datos:**
 - Realizar un análisis exhaustivo de calidad y preprocesamiento antes de comenzar el modelado.
 - Incorporar técnicas de imputación de valores faltantes o generación de datos sintéticos, si es necesario.
- 2. Para el alto costo computacional:**
 - Optimizar las arquitecturas de los modelos.
 - Implementar técnicas como early stopping y reducción del tamaño del conjunto de datos sin comprometer su representatividad.
- 3. Para la falta de precisión en las predicciones:**
 - Ajustar los hiperparámetros del modelo.
 - Probar enfoques adicionales como técnicas de ensamble (e.g., bagging o boosting).
 - Simplificar el modelo si es necesario para priorizar la precisión sobre la complejidad.
- 4. Para el problema de predicciones inexactas:**
 - Probar múltiples modelos de aprendizaje automático y profundo.
 - Evaluar diferentes configuraciones de datos y modelos para encontrar la mejor combinación.
- 5. Para el costo computacional excesivo:**
 - Utilizar servicios en la nube con opciones escalables.
 - Reducir la complejidad del modelo, priorizando aquellos que logren un buen balance entre rendimiento y eficiencia.