Predicción de Tiempos de Entrega en la Cadena de Suministro: Una Solución de Machine Learning para la Gestión Eficiente del Inventario

Seminario de Analítica y Ciencia de Datos

# Resumen Descriptivo del Proyecto

Este proyecto tiene como objetivo desarrollar un modelo predictivo que, a partir de datos históricos y técnicas de machine learning, permita estimar con precisión los tiempos de entrega de los proveedores en la cadena de suministro. Con esta solución, buscamos enfrentar el desafío de la falta de materias primas importadas, que provoca interrupciones en los procesos. El modelo tendrá impacto sobre la gestión del inventario y reducirá los tiempos de inactividad, mejorando así la eficiencia operativa.

Marco Teórico:

## Revisión de Literatura.

La predicción de tiempos de entrega en la cadena de suministro ha sido objeto de extensa investigación en el campo de la gestión de operaciones y la analítica de datos. En este proyecto, nos apoyaremos en referencias clave que aporten ideas significativas al desarrollo del mismo.

Banerjee et al. (2015) [1] se enfoca en la predicción de tiempos de entrega de proveedores en la cadena de suministro de piezas de motores de aeronaves. El estudio desarrolla un método que combina modelos supervisados como la regresión lineal y el modelo ARIMA, junto con una distribución gamma multivariante, utilizando datos históricos de órdenes de compra para identificar patrones y comportamientos en los tiempos de entrega. Además, se emplea la regresión paso a paso para identificar los factores predictivos más relevantes y se utiliza la distribución gamma multivariada para modelar tiempos de espera, mejorando la precisión del modelo. La eficacia de los modelos se evalúa mediante métricas de error de predicción, análisis de sensibilidad y especificidad. Implementar estos métodos en nuestro proyecto puede ser una estrategia efectiva para desarrollar un modelo predictivo robusto que reduzca la incertidumbre en la disponibilidad del material.

Steinberg et al. (2023) [2] propone un modelo de machine learning para predecir retrasos de proveedores en la cadena de suministro, especialmente en entornos de producción caracterizados por bajo volumen y alta variabilidad de la demanda. Para abordar los desafíos de dimensionalidad de los datos, el estudio emplea métodos de aprendizaje supervisado y técnicas de reducción de dimensionalidad. Entre los modelos utilizados se encuentran árboles de decisión, random forest y redes neuronales, comparando su precisión en distintos puntos del proceso de compra. También se implementa un control de la dimensionalidad mediante codificación binaria para reducirla sin excluir datos. Los resultados muestran que los algoritmos de boosting y random forest son los más efectivos, alcanzando altos niveles de precisión. Además, se aplica el marco CRISP-DM para proporcionar una estructura estandarizada en los proyectos de análisis de datos.

Zhou et al. (2023) [3] aborda la estimación de tiempos de entrega mediante un modelo denominado inductive graph transformer, diseñado para plataformas de comercio electrónico donde los tiempos de entrega afectan directamente al cliente. Este modelo captura interacciones complejas entre las características de los pedidos, como la dirección de origen, destino y la fecha de pago, para hacer estimaciones precisas. Utiliza una red convolucional sobre grafos para capturar relaciones estructurales en los datos, y un transformador para representar los pedidos y capturar interacciones de alto orden entre sus elementos. La capacidad del modelo para realizar predicciones complejas es valiosa en procesos con cadenas de suministro y entornos con ubicaciones cambiantes, asegurando la adaptabilidad del modelo a nuevas operaciones.

Wolter y Hanne (2024) [4] analizan la predicción de tiempos de servicio en entregas a domicilio de productos que requieren montaje e instalación, utilizando algoritmos de machine learning para reducir la dependencia del conocimiento empírico del personal logístico. El estudio emplea redes neuronales artificiales y modelos de regresión, comparando su rendimiento con máquinas de soporte vectorial y regresión lineal multivariable. También se aplican técnicas de procesamiento de datos, como normalización, manejo de valores atípicos y codificación one hot encoding para variables categóricas.

Maiti et al. (2014) [5] examina la predicción de tiempos de llegada de vehículos en tiempo real en contextos donde la disponibilidad de datos detallados es limitada. El modelo desarrollado busca predecir la llegada de autobuses en un entorno industrial en la India, donde la incertidumbre en el tráfico y la variabilidad en la velocidad de los vehículos complican la predicción. Los resultados muestran que el modelo basado en datos históricos (HD) logró una precisión del 75.56%, comparable a la de modelos más avanzados como las redes neuronales artificiales (76%) y las máquinas de soporte vectorial (71.3%). Esto sugiere que, aunque el modelo HD es más simple y menos exigente en recursos, su precisión es comparable a la de los modelos complejos.

## Modelos y Métodos ya existentes.

En el ámbito de la predicción de tiempos de entrega en la cadena de suministro, diversos estudios han explorado enfoques de machine learning para mejorar la precisión en la estimación de tiempos y reducir la incertidumbre. Estas metodologías han sido implementadas en entornos de manufactura y logística, destacándose modelos como random forest, ARIMA, redes neuronales y técnicas de optimización dimensional.

Los modelos de series temporales como ARIMA son comunes para predecir datos secuenciales en series de tiempo, permitiendo capturar patrones históricos y estacionales, especialmente en datos univariados [1]. Banerjee et al. (2015)[1] demostraron que ARIMA puede reducir significativamente los errores de predicción, lo cual lo convierte en una herramienta valiosa para visualizar patrones en datos de series de tiempo.

Random forest es utilizado cuando los datos incluyen múltiples variables y presentan alta complejidad. Este modelo, basado en árboles de decisión, mejora la precisión de la predicción [2]. Steinberg et al. (2023) [2] lo utilizaron para predecir retrasos en la entrega de proveedores en entornos de alta variabilidad. Random forest puede capturar relaciones complejas y no lineales, siendo adecuado para predecir tiempos de entrega en entornos cambiantes. Además, su menor sensibilidad a datos ruidosos y la falta de necesidad de un procesamiento extenso son ventajas importantes [3].

Las redes neuronales artificiales (ANN) son ampliamente utilizadas en la detección de patrones no lineales complejos, gracias a su capacidad de aprendizaje a partir de ejemplos históricos [2]. Varios estudios han explorado el uso de redes neuronales para mejorar la predicción de tiempos de entrega. Steinberg et al. (2023) [2] emplearon redes neuronales para prever retrasos en los proveedores, demostrando su utilidad en el análisis de los tiempos de entrega. Wolter y Hanne (2024) [4] implementaron un modelo de RNN para predecir tiempos de entrega en servicios a domicilio, logrando buenos resultados, aunque mencionaron el alto costo computacional y la necesidad de grandes volúmenes de datos. Maiti et al. (2014) [5] aplicaron redes neuronales para prever tiempos de llegada de buses en tiempo real, destacando su capacidad para adaptarse a variables dinámicas, aunque con la limitación de una interpretación compleja debido a las capas internas de las redes. Los modelos RNN ofrecen ventajas en la predicción de tiempos de entrega, pero se debe evaluar si el volumen de datos es suficiente y si los requerimientos computacionales son accesibles.

Las máquinas de soporte vectorial (SVM) son útiles en problemas de clasificación y muestran robustez ante conjuntos de datos variables, minimizando el efecto de datos aleatorios. Al enfocarse en puntos de datos relevantes, los modelos SVM logran alta precisión con tamaños de datos moderados [4]. Maiti et al. (2014) [5] emplearon SVM para abordar la variabilidad de los datos de tránsito, demostrando que es una alternativa efectiva cuando se requiere una clasificación precisa con datos limitados. En el proyecto de monografía, el modelo SVM podría ser eficaz para clasificar riesgos de retraso en entregas de materia prima, aprovechando su capacidad de procesamiento eficiente en conjuntos de datos moderados.

En la revisión de documentos del repositorio de la Universidad de Antioquia, se destaca el proyecto sobre el incumplimiento en la entrega de pedidos farmacéuticos a pacientes [6], un aspecto crítico que afecta la adherencia terapéutica. Para abordar este desafío, se desarrollaron modelos de machine learning que clasifican los pedidos según su probabilidad de cumplimiento, mejorando la eficiencia en la entrega de medicamentos. Los modelos empleados incluyen regresión logística, random forest, SVM y Naive Bayes. Además, se utilizaron metodologías de procesamiento de datos como limpieza, normalización, balanceo de datos y one-hot encoding para convertir variables categóricas a numéricas. También se empleó clustering con k-means para agrupar datos y detectar patrones que influyen en el cumplimiento. Los datos se dividieron en conjuntos de entrenamiento (30-70) y se realizó validación cruzada para evitar sobreajuste. La matriz de confusión y la curva ROC fueron utilizadas para evaluar la precisión, y random forest resultó ser el modelo más robusto en la predicción del cumplimiento de pedidos.

El proyecto "Análisis de Modelos Basados en Machine Learning para la Predicción de la Demanda de Productos" [7] aplicó modelos de machine learning como ARIMA, redes neuronales y random forest para mejorar la precisión en la predicción de demanda de productos. Este enfoque estructurado para predecir demanda e identificar patrones estacionales y tendencias también puede aplicarse al cumplimiento de entregas por parte de proveedores.

## Diferenciación del Proyecto

El proyecto propone una solución innovadora en el ámbito de la cadena de suministros, abordando un desafío crítico relacionado con la falta de materia prima importada, la cual impacta significativamente la eficiencia operativa de la empresa. Para enfrentar este reto, se plantea el desarrollo de un modelo predictivo que emplea datos históricos y técnicas avanzadas de machine learning, con el objetivo de estimar de manera precisa los tiempos de entrega y, de este modo, mejorar la eficiencia de los procesos logísticos.

A partir del análisis de estudios de referencia, se han identificado varios enfoques de modelos de machine learning que se ajustan a las características y necesidades del contexto de la cadena de suministro. Entre los modelos considerados se encuentra ARIMA, que permite capturar patrones históricos y estacionales en las series temporales; random forest, que se destaca por su capacidad para manejar datos de alta dimensionalidad; redes neuronales artificiales, que son idóneas para detectar patrones no lineales y complejos; y máquinas de soporte vectorial, que ofrecen una alta precisión en tareas de clasificación.

La integración de estos modelos, junto con la aplicación de las mejores prácticas en transformación de datos, tiene el potencial de optimizar la gestión de inventarios y minimizar las interrupciones operativas. Esta combinación de enfoques no solo contribuirá a mejorar la eficiencia de la cadena de suministro, sino que también sentará las bases para futuras aplicaciones en otros sectores logísticos y de manufactura, promoviendo una mayor robustez y adaptabilidad en la gestión de recursos.

Este proyecto, además, busca adaptar soluciones específicas a las problemáticas propias de la empresa, con el propósito de demostrar el potencial de la ciencia de datos en un entorno real. Al aprovechar de manera eficiente la información histórica almacenada, se generan mejoras significativas que resultan útiles para la operación. De esta forma, la diferencia clave de este proyecto radica en su enfoque específico, que permite explorar y aplicar soluciones adaptadas al contexto particular de la empresa, fortaleciendo así su capacidad para enfrentar desafíos logísticos.

## Referencias.

[1] A. G. Banerjee, W. Yund, D. Yang, P. Koudal, J. Carbone, y J. Salvo, “A Hybrid Statistical Method for Accurate Prediction of Supplier Delivery Times of Aircraft Engine Parts”, en *Volume 1B: 35th Computers and Information in Engineering Conference*, Boston, Massachusetts, USA: American Society of Mechanical Engineers, ago. 2015, p. V01BT02A037. doi: 10.1115/DETC2015-47605.

[2] F. Steinberg, P. Burggräf, J. Wagner, B. Heinbach, T. Saßmannshausen, y A. Brintrup, “A novel machine learning model for predicting late supplier deliveries of low-volume-high-variety products with application in a German machinery industry”, *Supply Chain Anal.*, vol. 1, p. 100003, mar. 2023, doi: 10.1016/j.sca.2023.100003.

[3] X. Zhou, J. Wang, Y. Liu, X. Wu, Z. Shen, y C. Leung, “Inductive Graph Transformer for Delivery Time Estimation”, en *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, feb. 2023, pp. 679–687. doi: 10.1145/3539597.3570409.

[4] J. Wolter y T. Hanne, “Prediction of service time for home delivery services using machine learning”, *Soft Comput.*, vol. 28, núm. 6, pp. 5045–5056, mar. 2024, doi: 10.1007/s00500-023-09220-7.

[5] S. Maiti, A. Pal, A. Pal, T. Chattopadhyay, y A. Mukherjee, “Historical data based real time prediction of vehicle arrival time”, en *17th International IEEE Conference on Intelligent Transportation Systems (ITSC)*, Qingdao, China: IEEE, oct. 2014, pp. 1837–1842. doi: 10.1109/ITSC.2014.6957960.

[6] G. A. M. Escobar, “Predicció de cumplimiento de entrega de pedidos farmacéuticos”, Monografia, Universidad de Antioquia, Antioquia Colombia.

[7] A. C. Loaiza, “Análisis de modelos basados en Machine Learning para la predicción de la demanda de productos en la empresa Dyna & Cía. S.A.”, Monografia, Universidad de Antioquia, Antioquia Colombia, 2023.