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. Dentro del marco de este proyecto, nos apoyaremos en referencias clave que aporten ideas significativas al desarrollo del proyecto.

El trabajo Banerjee et al. (2015) se enfoca en la predicción de tiempos de entrega de proveedores en la cadena de suministro de piezas de motores de aeronaves. Para abordar este problema, el estudio desarrolla un método que combina modelos supervisados tal como regresión lineal y modelo ARIMA, junto con una distribución gamma multivariante, estos modelos utilizan datos históricos de ordenes de compra completadas para aprender de patrones y comportamientos en los tiempos de entrega.

Alguno de los métodos específicos utilizados en el documento Banerjee et al. (2015) está la regresión paso a paso el cual se emplea para identificar los factores predictivos más relevantes, por otro lado, se utilizó la distribución gamma multivariada útil para modelar tiempos de espera mejorando la precisión, también se evaluó la eficacia de los modelos mediante métricas de error de predicción, análisis de sensibilidad y especificidad.

Implementar estos métodos en el proyecto de monografía puede ser una estrategia efectiva para desarrollar un modelo predictivo robusto que permita anticipar con mayor precisión los tiempos de entrega de los proveedores, reduciendo la incertidumbre en la disponibilidad del material.

Por otro lado, el estudio Steinberg et al. (2023) propone un modelo de machine learning para predecir retrasos de los proveedores en la cadena de suministro, especialmente en entornos de producción caracterizados por el bajo volumen y alta variabilidad de la demanda. Según el documento Steinberg et al. (2023). Este tipo de producción, se caracteriza por la alta variabilidad y complejidad en los componentes, el cual presenta desafíos, como problemas de dimensionalidad de los datos. Para abordar estos retos el proyecto emplea métodos de aprendizaje supervisado y de regresión, implementando técnicas de reducción de dimensionalidad para optimizar el analisis.

Alguno de los modelos utilizados en el estudio Steinberg et al. (2023) es la regresión supervisada el cual estima el retraso en días calendario, por otro lado, se evaluaron algoritmos como arboles de decisión, random forest, y redes neuronales, donde se compara su precisión en distintos puntos del proceso de compra, así mismo utiliza un control de la dimensionalidad generando codificación binaria para reducir la dimensionalidad sin excluir datos. Como resultado Este proyecto contribuye a una alta aplicabilidad en la cadena de suministro, permitiendo manejar datos de alta dimensionalidad sin comprometer la precisión, tambien, se encontró que los algoritmos de boosting y random forest son los más efectivos para el caso de estudio, alcanzando altos niveles de precisión, y por último se aplicó el marco crisp que proporciona una estructura estandarizada para los proyectos de análisis de datos.

A su vez, el estudio de Zhou et al. (2023) aborda la estimación de tiempo de entrega mediante un modelo novedoso denominado inductive graph transformer, diseñado para plataformas de comercio electrónico donde los tiempos de entrega afectan directamente al cliente y los tiempos de compra, este modelo se centra capturar en interacciones complejas entre las características de los pedidos tales como la dirección de origen, destino y la fecha de pago. Con el fin de hacer estimaciones precisas en los tiempos de entrega.

El modelo implementado en el estudio Zhou et al. (2023) emplea una red convolucional sobre grafos, el cual permite que el modelo capture relaciones estructurales en los datos, como las conexiones entre minorista y la ubicación. Además, incorpora un transformador aplicado a la representación de los pedidos con el fin de capturar interacciones de alto orden entre sus elementos, permitiendo que el modelo tome decisiones más precisas. finalmente, utiliza el transformador como función de regresión para predecir los tiempos de entrega de manera inductiva.

Además, La capacidad de la red convolucional para realizar predicciones complejas ayuda a predecir los tiempos de entrega tanto para relaciones previamente establecida como ubicaciones nuevas. Esto es valioso en procesos con cadenas de suministros o en entornos con ubicaciones cambiantes ya que asegura la adaptabilidad del modelo a operaciones nuevas.

Demanera similar, El proyecto de Wolter y Hanne (2024) analiza la predicción de tiempos de servicio en entregas a domicilio de productos, como muebles que requieren de montaje e instalación utilizando algoritmos de machine learning. El objetivo principal es reducir la dependencia del conocimiento empírico del personal logístico, optimizando los tiempos de entrega y de instalación. Para ello, El estudio emplea redes neuronales artificiales y modelos de regresión, comparando su rendimiento en la estimación real de los tiempos de servicios frente a modelos como la regresión lineal multivariable y máquinas de soporte vectorial.

En el estudio Wolter y Hanne (2024) se emplearon diversos algoritmos supervisados, entre ellos, las máquinas de soporte vectorial para comparar los resultados de clasificación y categorización de tiempos de servicio, también se utilizó la regresión lineal múltiple, que permitió analizar el rendimiento del modelo, además se emplearon tecinas de procesamiento de datos para limpiar y estructurar la información, lo cual incluye normalización, manejo de valores atípicos y codificación one hot encoding para variables categóricas.

En el estudio, también se emplearon diversos modelos, entre ellos redes neuronales artificiales y maquinas de soporte vectorial, que destacan por su capacidad para adaptarse a diferentes escenarios dentro del estudio. El modelo de red neuronal artificial permite automatizar la predicción de tiempos de entrega y montaje, utilizando patrones complejos en los datos y adaptándose a la variabilidad de diferentes escenarios.

El trabajo de Maiti et al. (2014) examina el modelo de predicción de tiempos de llegada de vehículos en tiempo real, enfocado en contextos donde la disponibilidad de datos detallados es limitada. El modelo fue desarrollado con el objetivo de 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 de los horarios de llegada.

El estudio Maiti et al. (2014) muestra que el modelo basado en datos históricos HD logro una precisión 75.56%, lo cual es notablemente cercano a los resultados obtenidos por modelos avanzados como el ANN el cual alcanzaron un 76% de precisión, mientras que las maquinas de soporte vectorial SVM lograron 71,3%, estos resultados destacan que, aun que el modelo HD es mas simple y menos exigente en recursos, su precisión es comparable a la de los modelos más complejos.

## Modelos y Métodos ya existentes.

En el ámbito de la predicción de tiempos de entrega en la cadena de suministro, diversos estudios anteriormente mencionados han explorado enfoques de modelos de machine learning con el fin de mejorar la precisión en la estimación de tiempos y reducir la incertidumbre. Estas metodologías han sido implementadas en entornos de manufactura y logística, siendo modelos como random forest, arima, redes neuronales y técnicas de optimización dimensional lo mas destacado de los estudios mencionados.

Los modelos de series temporales como ARIMA son comúnmente utilizados para predecir datos secuenciales en series de tiempo. Estos modelos permiten capturar patrones históricos y estacionales, y son particularmente eficaz en datos univariados y secuenciales como destaca (Box, Jenkins, Reinsel, & Ljung, s.f.).

En el documento de Banerjee et al. (2015) la implementación de un modelo ARIMA demostró una notable reducción de los errores de predicción, lo que resalta la capacidad para capturar patrones, por lo tanto, en el contexto del proyecto propuesto, explorar el modelo ARIMA puede proporcionar una herramienta valiosa para visualizar patrones especialmente en datos que involucren series de tiempo.

Random forest es normalmente utilizado donde los datos incluyen múltiples variables y presentan una alta variabilidad y complejidad, este modelo está basado en arboles de decisión el cual mejora la precisión de la predicción (Leo, 2001).

El documento Steinberg et al. (2023) propone predecir los retrasos en la entrega de proveedores en entornos de baja demanda y alta variabilidad por lo que utiliza algoritmos de random forest, en el contexto del proyecto el modelo puede ayudar a capturar relaciones complejas y no lineales, hace que sea adecuado predecir tiempos de entrega en entornos variables, por otro lado, random forest es menos sensible a datos ruidosos y no requiere un procesamiento amplio, lo cual es ventajoso en proyectos donde las características pueden cambiar con el tiempo (Liaw & Wiener, 2002).

Las redes neuronales artificiales ANN son ampliamente utilizadas en proyectos que requieren la detección de patrones no lineales complejos en los datos, dado que las ANN tienen la capacidad de aprender de ejemplos históricos, lo que resulta útil en situaciones con alta variabilidad (Bengio & Courville, 2016).

En los documentos revisados, varios estudios exploran el uso de redes neuronales para mejorar la predicción de tiempos de entrega en la cadena de suministros. En el documento de Steinberg et al. (2023), las redes neuronales se emplean para prever retrasos en los proveedores, mostrando ser útiles en el análisis detallado de los tiempos de entrega.

Por su parte Wolter y Hanne (2024) implementaron un modelo de RNN para predecir tiempos de entrega en servicios a domicilios, logrando buenas predicciones de tiempos de entrega y de servicios, sin embargo, estos autores mencionan desafíos como el alto costo computacional y la necesidad de grandes volúmenes de datos para optimizar el rendimiento del modelo.

También, en el estudio Maiti et al. (2014), las redes neuronales se aplican para prever tiempos de llegada de buses en tiempo real, demostrando una notable capacidad para adaptarse a variables dinámicas de tráfico, distancia y velocidad. No obstante, el documento resalta una limitación y es la complejidad de interpretar los resultados debido a las capaz internas de las redes neuronales.

Consideramos que los modelos RNN ofrecen una gran ventaja en proyectos como el planteado, que busca predecir los tiempos de entrega, la capacidad de estos modelos para detectar patrones no lineales, sin embargo, se debe revisar si el volumen de datos es suficiente y los requerimientos computacionales son accesibles.

Las máquinas de soporte vectorial son especialmente útiles en problemas de clasificación y muestran robustez ante conjuntos de datos variables, ayudando a minimizar el efecto de datos aleatorias, además, al enfocarse solo en puntos de datos más relevantes los modelos SVM logran una alta precisión con tamaños de datos moderados (Hastie, Tibshirani, & Friedman, 2009)

El estudio de Maiti et al. (2014) emplea los modelos de SVM para abordar la variabilidad de los datos de tránsito y tráfico. Esta investigación muestra que los modelos SVM ofrecen una alternativa en situaciones de clasificación precisa con acceso de datos limitados, También destaca su capacidad para manejar conjuntos de datos con alta variabilidad y proporciona una alternativa efectiva en escenarios donde se requiere una clasificación precisa de los tiempos de llegada.

A partir de esta revisión, es importante explorar en el proyecto de monografía el modelo SVM, ya que han demostrado ser eficaces donde el contexto es clasificar los riesgos de retraso en entregas de materia prima, también este tipo de modelos permitirá aprovechar la capacidad de procesamiento eficiente en conjunto de datos de tamaño moderado, lo cual reduce el requerimiento computacional.

Con el fin de ampliar la base de investigación se exploró algunos documentos en el repositorio de la universidad de Antioquia, entre los cuales se destaca el proyecto de incumplimiento en la entrega de pedidos farmacéuticos a los pacientes (Montoya Escobar, 2022), aspecto critico que afecta la adherencia terapéutica de los pacientes en Antioquia. Para abordar este desafío, se desarrollaron modelos de machine learning que clasifican los pedidos en función de su probabilidad de cumplimiento y así mejorar en la eficiencia de entregas de medicamentos, alguno de los modelos empleados incluye, regresión logística evaluando probabilidades de resultados binarios, random forest destacado en este caso por su precisión y estabilidad logrando buenos resultados, también máquinas de soporte vectorial y naive bayes. En cuanto a la trasformación de los datos el proyecto (Montoya Escobar, 2022) incluye metodologías de procesamiento de datos como limpieza, normalización y balanceo de datos, también técnicas como one-hot encoding para convertir las variables categorías a numéricas. Otra metodología utilizada es el clustering el cual utiliza el k means para agrupar los datos observando patrones en las entregas que pueden haber influido en el cumplimiento.

También en entrenamiento y validación los datos se dividieron en conjunto de entrenamiento 30-70 y se realiza una validación cruzada para evitar sobreajuste, algunas técnicas de exactitud es la matriz de confusión y la curva roc, por último, el random forest fue el modelo más robusto, logrando la mejor precisión en la predicción del cumplimiento de pedidos.

El segundo proyecto Análisis de Modelos Basados en Machine Learning para la Predicción de la Demanda de Productos (Correa Loaiza, 2023) tiene como objetivo principal analizar y aplicar modelos de machine learning para mejorar la precisión en la predicción de demanda de productos. Alguno de los modelos utilizados ARIMA, Redes Neuronales, Random Forest. Este proyecto se considera relevante el cual ofrece un enfoque estructurado de machine learning para predecir la demanda e identificar patrones estacionales y tendencias, lo que también puede aplicarse al cumplimiento de entregas por parte de proveedores.

## Diferenciación del Proyecto

El proyecto Predicción de Tiempos de Entrega en la Cadena de Suministro se presenta como una propuesta en el ámbito de la cadena de suministros, abordando un desafío critico como la falta de materia prima importada, el cual afecta significativamente la eficiencia operativa. La innovación de la propuesta está en el desarrollo de un modelo predictivo que emplea datos históricos y técnicas avanzadas de machine learning para estimar con precisión los tiempos de entrega.

A partir del análisis de estudios de referencia, se ha identificado algunos enfoques de modelos de machine learning que se ajustan a las características y necesidades del contexto planteado. Entre los modelos destacados esta ARIMA el cual ayuda a capturar patrones históricos y estacionales en las series de tiempo, random forest por la capacidad de generar datos de alta dimensionalidad, redes neuronales artificiales que ayudan a detectar patrones no lineales y complejo, maquinas de soporte vectorial para realizar clasificaciones precisas.

Este proyecto pretende integrar los modelos más utilizados para casos similares, asi mismo las mejores practicas en cuanto a la transformación de datos, la combinación de estos enfoques tiene el potencial de optimizar la gestión de inventarios y minimizar las interrupciones operativas, sentando las bases para futuras aplicaciones en otros sectores logísticos y de manufactura.

## 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.

# Rúbricas

R01: Completitud y consistencia: el estudiante realiza una entrega del momento evaluativo con todos los elementos desarrollados, manteniendo una relación coherente entre los elementos solicitados. El estudiante usa una consistencia argumentativa que incluye el uso de referencias cuando hay mérito de autoría entre otros.

R02: Línea Base: el estudiante es capaz de establecer una clara línea de trabajos previos y teorías basadas alrededor de su proyecto de clase. Esto permite evidenciar que la solución propuesta en el trabajo precio es viable y tiene una base sólida teórica detrás de su posible solución.

R02: Innovación: dentro del ámbito de la analítica y la ciencia de datos el estudiante muestra, basado en la literatura existente, una clara diferenciación de su proyecto frente las soluciones tecnologico-metodológicas encontradas. Se tiene en cuenta la inclusión de la base de datos de CESET para la rúbrica.

R04: Referencias: hay referencias válidas según la descripción del capítulo y están adecuadamente vinculadas al estilo narrativo que permite una clara expresión de los resultados de la investigación de la línea base y el marco tecnológico-metodológico.