Título del Proyecto

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 ampliamente estudiada en el ámbito de la gestión de operaciones y la analítica de datos. En marco del proyecto se encuentra como referencias relevantes que pueden aportar ideas significativas al desarrollo de este:

Banerjee et al. (2015) desarrollaron un modelo híbrido estadístico para predecir los tiempos de entrega de piezas en la industria aeronáutica. Su enfoque combina un modelo de regresión paso a paso con una distribución gamma multivariante, lo que permite una predicción robusta frente a la variabilidad en las capacidades de los proveedores. Este modelo es altamente relevante para el proyecto, ya que permite realizar predicciones precisas sobre los tiempos de entrega en contextos con largos tiempos de espera y alta variabilidad, características también presentes en cadenas de suministro de materias primas​ [1].

Steinberg et al. (2023) proponen un modelo de machine learning para predecir retrasos en la entrega de productos en entornos de manufactura de baja cantidad y alta variedad. Utilizando un enfoque de regresión, lograron predecir el grado de retraso en días calendario, lo que proporciona información crítica para tomar decisiones en fases tempranas del proceso de compra. Este estudio es altamente pertinente, ya que también aborda la predicción de retrasos en entornos industriales complejos y ofrece una solución adaptada a escenarios donde los productos son altamente personalizados​ [2].

Zhou et al. (2023) introducen un transformador gráfico inductivo (IGT) para la estimación de tiempos de entrega en grandes plataformas de comercio electrónico. Aunque se enfoca en la entrega de paquetes en el sector minorista, el uso de redes neuronales gráficas para capturar relaciones semánticas complejas entre múltiples características de los pedidos es de gran relevancia para este proyecto. El IGT podría ser adaptado para predecir tiempos de entrega en la cadena de suministro de materias primas, aprovechando su capacidad para manejar grandes volúmenes de datos y múltiples variables interrelacionadas​ [3].

Wolter y Hanne (2024) utilizaron redes neuronales artificiales (ANN) para predecir los tiempos de servicio en la entrega de muebles con ensamblaje a domicilio. Aunque el contexto es distinto, el uso de ANN para predecir la duración de procesos complejos basado en múltiples variables es una técnica valiosa para este proyecto. En particular, la capacidad de ANN para manejar incertidumbre en los tiempos de entrega basada en características específicas de los pedidos puede ser directamente aplicable [4].

Maiti et al. (2014) propusieron un modelo basado en datos históricos para la predicción en tiempo real de la llegada de vehículos. Si bien este trabajo se centra en la predicción de tiempos de llegada de autobuses, su enfoque de utilizar solo un conjunto limitado de variables clave (trayectoria del vehículo y marcas de tiempo) para realizar predicciones precisas puede inspirar una arquitectura simplificada para el modelo de predicción de tiempos de entrega en la cadena de suministro. Este enfoque es particularmente útil cuando se busca un balance entre precisión y simplicidad operativa​ [5].

## Modelos y Métodos ya existentes.

La predicción de tiempos de entrega en la cadena de suministro se ha abordado desde distintos enfoques y utilizando una variedad de modelos de machine learning. En este proyecto, se consideran metodologías previas y se evalúan sus ventajas y desventajas en el contexto de la solución propuesta.

Entre los métodos existentes se destacan los modelos híbridos, como el propuesto por Banerjee et al. (2015), que utilizan una combinación de regresión y distribuciones probabilísticas para enfrentar la variabilidad en los tiempos de entrega. Estos modelos son adecuados cuando se dispone de datos con una alta variabilidad y se requiere capturar distribuciones complejas de los tiempos de entrega. Sin embargo, pueden ser limitados en términos de escalabilidad y adaptación a escenarios más dinámicos donde la estructura de datos cambia rápidamente.

El modelo de Steinberg et al. (2023) hace uso de regresión para predecir el grado de retraso en días calendario, lo cual es especialmente relevante para nuestro objetivo de estimar el tiempo de cumplimiento o retraso de los pedidos en la cadena de suministro. Los enfoques basados en regresión permiten una predicción continua, lo cual es fundamental para evaluar la severidad de un posible retraso y tomar decisiones proactivas. Además, este modelo destaca la importancia de realizar predicciones en fases tempranas del proceso de compra, proporcionando un margen mayor para implementar contramedidas en caso de un retraso pronosticado.

Zhou et al. (2023) introducen el uso de redes neuronales gráficas mediante el transformador gráfico inductivo (IGT), el cual se presenta como una herramienta poderosa para capturar relaciones complejas en datos con múltiples variables interrelacionadas. Este enfoque podría ser particularmente útil cuando se trata de modelar cadenas de suministro que dependen de múltiples proveedores y donde las relaciones entre pedidos son críticas. Las redes neuronales gráficas tienen la ventaja de poder representar relaciones no triviales y de gran escala, aunque pueden requerir una gran capacidad computacional y ser difíciles de interpretar.

Las redes neuronales artificiales (ANN), como las utilizadas por Wolter y Hanne (2024), también han demostrado ser efectivas para predecir tiempos de entrega en escenarios donde la variabilidad es alta y la complejidad de los procesos depende de muchas variables. Las ANN son adecuadas cuando se dispone de grandes cantidades de datos históricos y se busca identificar patrones complejos. Sin embargo, su capacidad de generalización puede verse comprometida si no se dispone de un volumen suficiente de datos o si los datos contienen un alto grado de ruido.

En el ámbito de la predicción de entregas en el sector farmacéutico, el trabajo de Montoya (2022) es relevante, ya que emplea algoritmos de clasificación como Random Forest, Regresión Logística, Naive Bayes y Máquinas de Soporte de Vectores (SVM) para predecir el cumplimiento de los tiempos de entrega. Estos modelos se utilizan para realizar predicciones binarias sobre el cumplimiento o incumplimiento de los tiempos prometidos, lo cual ofrece una perspectiva interesante sobre cómo incorporar métricas que permitan la evaluación y optimización de los tiempos de entrega.

Considerando los métodos mencionados, en el proyecto se plantea la utilización de un enfoque de machine learning que permita no solo clasificar los pedidos como cumplidos o no cumplidos, sino también predecir el tiempo exacto de retraso en días calendario. Para ello, se explorarán modelos de regresión supervisada, incluyendo Random Forest Regressor y modelos basados en redes neuronales. Estos enfoques permiten trabajar directamente con la variable de interés, "tiempo de retraso o cumplimiento", y generar predicciones continuas, lo cual resulta esencial para la gestión proactiva del inventario y la optimización de la cadena de suministro. Además, se evaluará la inclusión de técnicas de reducción de dimensionalidad como PCA (Análisis de Componentes Principales) para mejorar la eficiencia computacional y evitar el problema de la "maldición de la dimensionalidad", especialmente relevante en escenarios de baja cantidad y alta variedad de datos. Finalmente, se considera la posible integración de modelos híbridos que combinen técnicas estadísticas con modelos de machine learning para aprovechar las fortalezas de ambos enfoques.

## Diferenciación del Proyecto

[Elabore una discurso basado en los capítulos anteriores que muestre por qué su proyecto es innovador, ya sea por la vía de plantear algo totalmente nuevo o por la vía de mejorar significativamente algo existente. Sea recursivo, use su propio trabajo y el de los demás]

## 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.