

CHALLENGE 3

MAESTRÍA EN CIENCIAS DE LOS DATOS

CENTRO UNIVERSITARIO DE CIENCIAS ECONÓMICO-ADMINISTRATIVAS

ANDRÉS FELIPE VÁZQUEZ VALDÉS

**“CASO DE ESTUDIO: ANÁLISIS Y PREVENCIÓN DEL SOBREINVENTARIO USANDO
HERRAMIENTAS DE CIENCIA DE DATOS APLICADO A UNA EMPRESA REVENDEDORA DE
MICROCOMPONENTES ELECTRÓNICOS”**

LÍNEA DE GENERACIÓN Y APLICACIÓN DEL CONOCIMIENTO

SMART DATA

FECHA 5 DE DICIEMBRE DE 2024

CONTEXTO DEL PROBLEMA

La empresa en la cual se va a enfocar este proyecto se dedica en su mayoría a la venta de microcomponentes electrónicos para industrias con diversos giros, ya sea automotriz, militar, aeroespacial, telecomunicaciones etcétera. Cuenta con diversos modelos de negocio dependiendo de las características y nivel de consumo de los clientes.

- *Órdenes discretas*
- *Proximity Warehouse (PW)*
- *In Plant Store (IPS)*
- *Vendor Managed Inventory (VMI) / modelo de consignación*

Para los modelos IPS, PW o VMI, el cliente manda su pronóstico de consumo (Forecast) en base a este se generan órdenes de compra al proveedor y se envía el material. La facturación no se realiza una vez que el cliente retira el material de su almacén.



PROBLEMÁTICA

Debido a la fluctuante demanda de material, cuenta con altos niveles de sobreinventario. Muchas veces este inventario no se le puede vender a otros clientes debido a que las piezas solo figuran en la demanda de uno solo, o se vuelven obsoletas.

OBJETIVO PRINCIPAL:

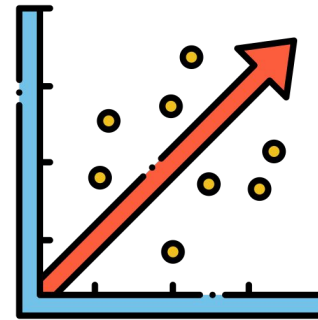
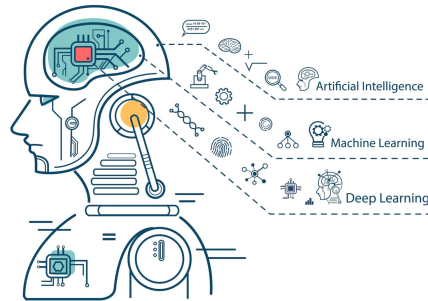
Desarrollar un modelo matemático basado en herramientas avanzadas de ciencia de datos para el análisis, prevención del sobreinventario.



CAPÍTULO II. MARCO TEÓRICO CONCEPTUAL

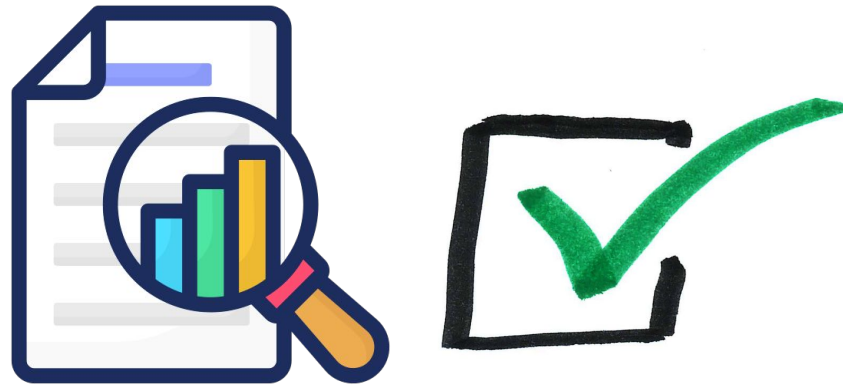
Se han desarrollado varias técnicas de pronósticos entre ellas se encuentran las **Series de Tiempo**, que se usan para pronosticar eventos futuros de acuerdo con el comportamiento observado de diferentes variables, llamadas componentes, de ocurrencia en el pasado (Leon-Álvarez, Betancur-Gómez, Jaimes-Barragán, & Grisales-Romero, 2016).

Las técnicas que se implementan en **Machine Learning** para pronosticar la demanda contemplan diversas variables de gran importancia, como variables climáticas, factores económicos, situaciones, patrones estacionales y condiciones demográficas (Khakpour, 2020). Se utilizan habitualmente varios algoritmos de machine learning, entre ellos la **Regresión Lineal**: Este algoritmo se utiliza para predecir valores numéricos, basados en una relación lineal entre diferentes valores (¿Qué Es Machine Learning (ML)? | IBM, s. f.).



CAPÍTULO III. CONTEXTO METODOLÓGICO

- Recopilación y análisis de la información a partir de bases de datos de históricos de ventas, pronósticos, recepción y envíos de material. Realizar entrevistas al personal clave para entender los procesos actuales y desafíos.
- Generar modelos basados en la regresión lineal y series de tiempo, y validar los modelos con datos reales para asegurar su efectividad.
- Evaluar y medir de los resultados, obtener porcentaje de error.



TÉCNICAS PARA LA RESOLUCIÓN DEL PROYECTO

Para este proyecto se podría implementar un modelo de predicción supervisado, porque se está tratando de predecir variables continuas o categóricas basada en datos históricos.

Entre los modelos que vimos en clase, los más adecuados para esta tarea son aquellos que son buenos para regresión (predicción de valores continuos) o clasificación (si tu objetivo es predecir categorías como “abastecer / no abastecer”).

Modelos de Aprendizaje Supervisado
Regresión Lineal
Regresión Logística
Árboles de Decisión
Máquinas de Soporte Vectorial (SVM)
Redes Neuronales
k-Vecinos Más Cercanos (k-NN)
Random Forest
Gradient Boosting (XGBoost, LightGBM)

TÉCNICAS PARA LA RESOLUCIÓN DEL PROYECTO

1. **Redes Neuronales** – especialmente redes neuronales recurrentes (RNN) o LSTM si los datos tienen una fuerte dependencia temporal.
2. **Gradient Boosting** (como XGBoost o LightGBM) – muy potentes y eficaces con datos tabulares, suelen dar excelentes resultados.
3. **Random Forest** – también es buena opción, aunque puede ser menos precisa que Gradient Boosting si no se ajusta bien.
4. **Árboles de Decisión** – útiles y explicables, pero suelen tener menor desempeño que los modelos ensemble.
5. **Máquinas de Soporte Vectorial (SVM)** – pueden funcionar si los datos no son muy grandes, pero no son los más comunes para series temporales o pronóstico.

Modelos de Aprendizaje Supervisado
Regresión Lineal
Regresión Logística
Árboles de Decisión
Máquinas de Soporte Vectorial (SVM)
Redes Neuronales
k-Vecinos Más Cercanos (k-NN)
Random Forest
Gradient Boosting (XGBoost, LightGBM)

TÉCNICAS PARA LA RESOLUCIÓN DEL PROYECTO

1. El algoritmo Gradient Boosting (Potenciación de Gradiente) es uno de los algoritmos más populares de ML. Es suficientemente poderosos para encontrar relaciones no lineales entre el target de un modelo y las características. Puede trabajar con valores faltantes, outliers y gran cantidad de valores por característica sin un tratamiento especial.

TÉCNICAS PARA LA RESOLUCIÓN DEL PROYECTO

Sería conveniente empezar por **XGBoost**, **Random Forest** o una **Red Neuronal (LSTM si tienes datos por fechas)**.

Columnas que podría tener la base de datos:

- fecha.
- material (ID o nombre del material)
- cantidad consumida (variable objetivo a predecir).
- cliente
- proveedor
- almacén (Gdl, Channdler)
- precio
- temporada (Cuarto fiscal)

Esta informacion podria ser almacenada en un csv con el nombre "consumo_material.csv"

TÉCNICAS PARA LA RESOLUCIÓN DEL PROYECTO

Se podría implementar un modelo con estas características:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from xgboost import XGBRegressor
import xgboost as xgb
from sklearn.metrics import accuracy_score, roc_curve, auc, confusion_matrix
from sklearn.metrics import mean_squared_error

# 1. Cargar los datos
df = pd.read_csv("consumo_material.csv", parse_dates=["fecha"])

train_X, test_X, train_y, test_y = train_test_split(iris.data, iris.target, test_size=0.2)
train=xgb.DMatrix(train_X,label=train_y)
test=xgb.DMatrix(test_X)
```

TÉCNICAS PARA LA RESOLUCIÓN DEL PROYECTO

```
param = {  
    'max_depth': 4,  
    'eta': 0.3,  
    'objective': 'multi:softmax',  
    'tree_method': 'auto',  
    'num_class': 3}
```

```
epochs=10
```

```
#Este puede ser con árboles de decisión
```

```
model=xgb.train(param,train,epochs)
```

```
predictions= model.predict(test)
```

```
#Se calcula el accuracy (predicciones correctas / todas las predicciones) de acuerdo a las predicciones
```

```
accuracy_score(test_y,predictions)
```

```
#Este modelo puede ser con Regresion
```

```
modelo = XGBRegressor()
```

```
modelo.fit(X_train, y_train)
```

```
#Se calcula el accuracy (predicciones correctas / todas las predicciones) de acuerdo a las predicciones
```

```
accuracy_score(test_y,predictions)
```

TÉCNICAS PARA LA RESOLUCIÓN DEL PROYECTO

El algoritmo Gradient Boosting (Potenciación de Gradiente) es uno de los algoritmos más populares de ML. Es suficientemente poderosos para encontrar relaciones no lineales entre el target de un modelo y las características. Puede trabajar con valores faltantes, outliers y gran cantidad de valores por característica sin un tratamiento especial.