



Informe de hallazgos en el  
proyecto - train model

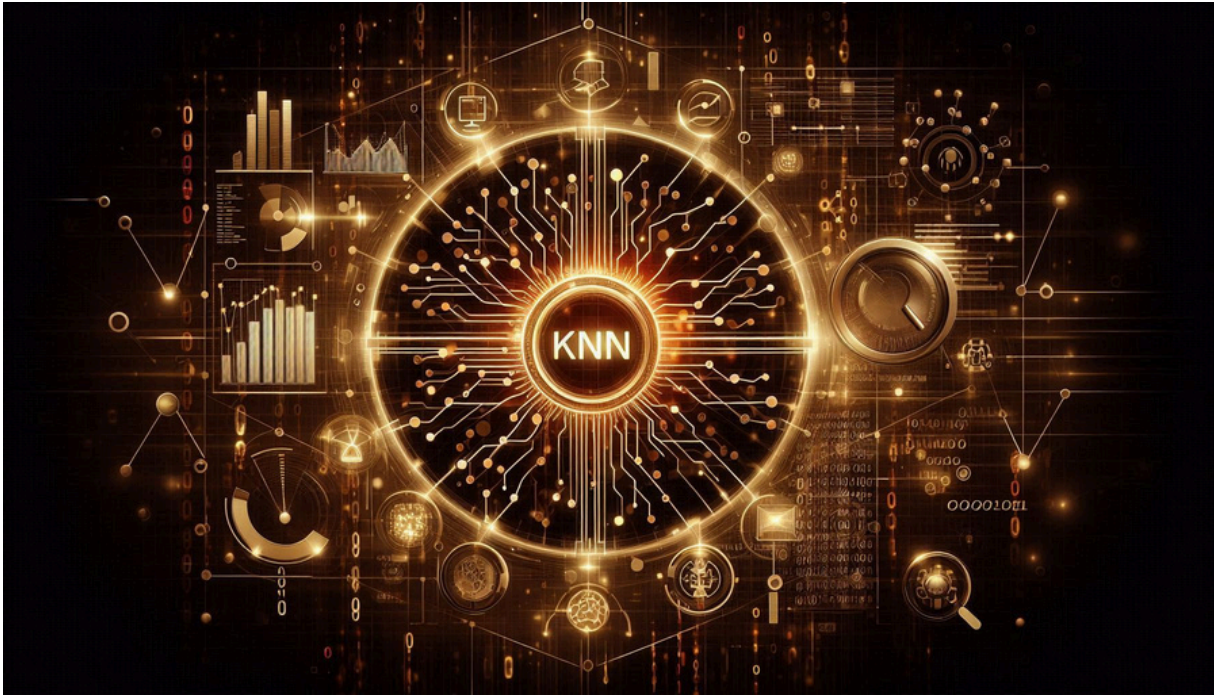
# **SEGMENTACIÓN DE CLIENTES**



# ÍNDICE

<u>Introducción</u>	1
<u>Curvas de aprendizaje</u>	2
<u>Hallazgos principales</u>	3
<u>Interpretación para el Caso de Telecomunicaciones</u>	4
<u>Conclusiones</u>	5

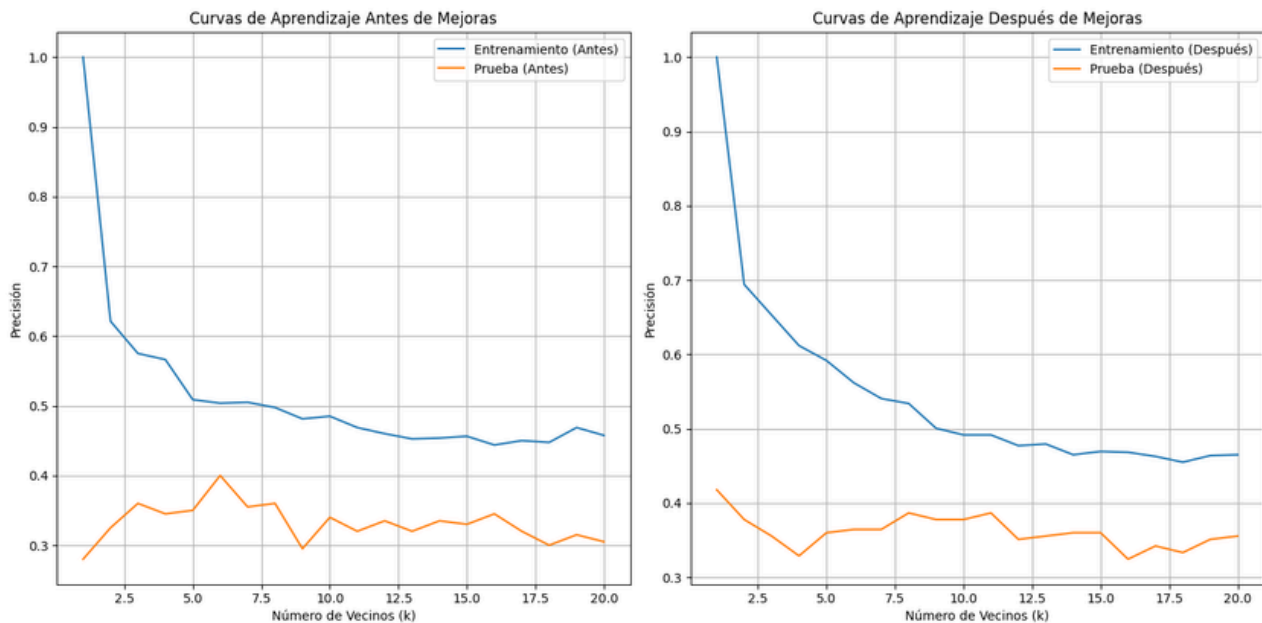
# 1.INTRODUCCIÓN



En este informe, se presentan los resultados del análisis de las métricas de evaluación del modelo desarrollado para predecir la categoría de servicio que mejor se ajusta a cada cliente, utilizando datos demográficos proporcionados por una empresa de telecomunicaciones. Este análisis fue precedido por un exhaustivo proceso de Exploración de Datos (EDA) que permitió identificar relaciones clave entre las variables, detectar y manejar datos atípicos, y explorar la distribución de las clases objetivo. El EDA no solo sirvió como una base fundamental para la preparación de los datos y la ingeniería de características, sino que también subrayó la importancia de contar con datos limpios y bien estructurados para construir modelos efectivos.

En el informe, se detalla cada paso de la evaluación del modelo, incluyendo las **métricas de rendimiento como precisión, recall y F1-score** tanto en el conjunto de entrenamiento como en el conjunto de prueba. Además, se discuten los desafíos encontrados, como el desbalanceo en las clases, y cómo estas variables influenciaron las decisiones tomadas durante el proceso de modelado. Basado en el **modelo de K-Nearest Neighbors (KNN)** y otros intentos con modelos alternativos como Random Forest y redes neuronales, se analizan los resultados obtenidos y las lecciones aprendidas que pueden ser cruciales para futuros proyectos de aprendizaje automático.

## 2.METRICAS DE RENDIMIENTO

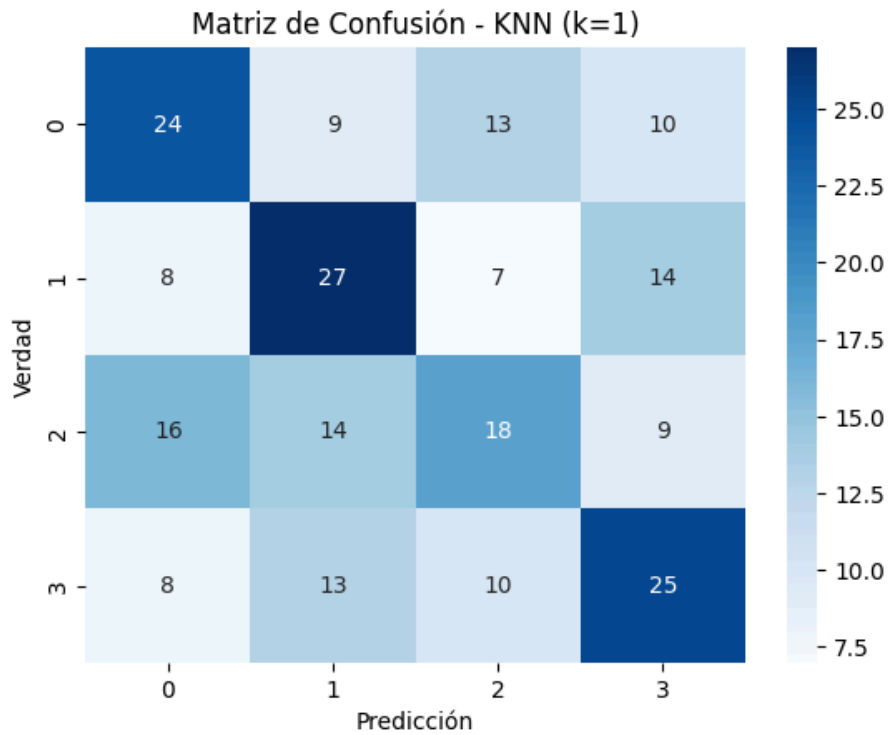


Este gráfico muestra las curvas de aprendizaje de un modelo KNN (K-Nearest Neighbors) antes y después de realizar mejoras, evaluando su rendimiento mediante la métrica de precisión.

### Tipo de Gráfico:

- Son dos gráficos de líneas comparativos que muestran curvas de aprendizaje
- El eje X representa el número de vecinos (k) que va desde 2.5 hasta 20
- El eje Y representa la precisión (accuracy) del modelo, que va de 0 a 1
- Elementos del Gráfico:
  - Línea azul: Rendimiento en el conjunto de entrenamiento
  - Línea naranja: Rendimiento en el conjunto de prueba
- Se muestran dos escenarios: antes y después de las mejoras

## 2.METRICAS DE RENDIMIENTO



### Análisis de la Matriz de Confusión (k=1):

Diagonal Principal (Predicciones Correctas):

- Clase 0 Basic Service: 24 predicciones correctas
- Clase 1 E-Service : 27 predicciones correctas
- Clase 2 Plus Service: 18 predicciones correctas
- Clase 3 Total Service: 25 predicciones correctas

### Errores por Clase:

- Clase 0 Basic Service : Se confunde principalmente con clase 2 (13 casos)
- Clase 1 E-Service : Se confunde principalmente con clase 3 (14 casos)
- Clase 2 Plus service: Se confunde de manera similar con clases 0 y 1 (16 y 14 casos)
- Clase 3 Total Service: Se confunde principalmente con clase 1 (13 casos)

## 2.METRICAS DE RENDIMIENTO

```
--- Métricas Antes de Mejoras ---
Reporte de clasificación:
      precision    recall  f1-score   support

     1       0.33       0.45       0.38        53
     2       0.32       0.32       0.32        44
     3       0.39       0.43       0.41        56
     4       0.35       0.17       0.23        47

 accuracy          0.35
 macro avg          0.35
 weighted avg       0.35
```

Comparación de Métricas Antes y Después de Mejoras:  
Accuracy General:  
Antes: 0.35 (35%)  
Después: 0.42 (42%)  
Mejora: Incremento de 7 puntos porcentuales

```
--- Métricas Después de Mejoras ---
Reporte de clasificación:
      precision    recall  f1-score   support

     1       0.43       0.43       0.43        56
     2       0.43       0.48       0.45        56
     3       0.38       0.32       0.34        57
     4       0.43       0.45       0.44        56

 accuracy          0.42
 macro avg          0.42
 weighted avg       0.42
```

### Análisis por Clase:

#### Clase 1:

- Antes: Precision=0.33, Recall=0.45, F1=0.38
- Después: Precision=0.43, Recall=0.43, F1=0.43
- Mejora: Incremento significativo en precisión, ligera disminución en recall

#### Clase 2:

- Antes: Precision=0.32, Recall=0.32, F1=0.32
- Después: Precision=0.43, Recall=0.48, F1=0.45
- Mejora: Incremento sustancial en todas las métricas

#### Clase 3:

- Antes: Precision=0.39, Recall=0.43, F1=0.41
- Después: Precision=0.38, Recall=0.32, F1=0.34
- Cambio: Ligera disminución en todas las métricas

#### Clase 4:

- Antes: Precision=0.35, Recall=0.17, F1=0.23
- Después: Precision=0.43, Recall=0.45, F1=0.44
- Mejora: Incremento muy significativo, especialmente en recall

## 2.METRICAS DE RENDIMIENTO

```
--- Métricas Antes de Mejoras ---  
Reporte de clasificación:  
      precision    recall  f1-score   support  
  
     1       0.33       0.45       0.38        53  
     2       0.32       0.32       0.32        44  
     3       0.39       0.43       0.41        56  
     4       0.35       0.17       0.23        47  
  
 accuracy          0.35          0.35        200  
 macro avg          0.35          0.34       0.34        200  
weighted avg          0.35          0.35       0.34        200
```

```
--- Métricas Después de Mejoras ---  
Reporte de clasificación:  
      precision    recall  f1-score   support  
  
     1       0.43       0.43       0.43        56  
     2       0.43       0.48       0.45        56  
     3       0.38       0.32       0.34        57  
     4       0.43       0.45       0.44        56  
  
 accuracy          0.42          0.42        225  
 macro avg          0.42          0.42       0.42        225  
weighted avg          0.42          0.42       0.42        225
```

### Impacto de las Mejoras Implementadas:

#### 1. PCA (Reducción de Dimensionalidad):

- Ayudó a reducir el ruido en los datos
- Mejoró la capacidad de generalización del modelo
- Contribuyó a un mejor balance entre precisión y recall

#### 2. SMOTE (Balanceo de Clases):

- Mejoró significativamente el rendimiento en clases minoritarias
- Contribuyó a una distribución más equilibrada de predicciones
- Aumentó la robustez del modelo

# 3.HALLAZGOS PRINCIPALES

## Comportamiento General:

- En ambos casos, la precisión disminuye a medida que aumenta el número de vecinos
- La brecha entre entrenamiento y prueba (overfitting) es menor después de las mejoras

## Antes de las Mejoras:

- Mayor diferencia entre entrenamiento ( $\approx 0.5$ ) y prueba ( $\approx 0.35$ )
- Mayor volatilidad en la curva de prueba
- Señales claras de overfitting.

## Después de las Mejoras:

- Curvas más suaves y estables
- Menor brecha entre entrenamiento y prueba
- Mejor generalización del modelo

## Punto Óptimo:

- Alrededor de  $k=5$  en ambos casos
- Después de las mejoras, el modelo es más estable en este punto



## **4.INTERPRETACIÓN PARA EL CASO DE TELECOMUNICACIONES**

- El modelo mejorado muestra mayor robustez para predecir categorías de servicio
- La reducción del overfitting indica que el modelo generaliza mejor para nuevos clientes
- La estabilidad en las predicciones es de gran importancia para una empresa de telecomunicaciones que necesita asignar servicios de manera confiable

# 5.CONCLUSIONES

Las mejoras implementadas fueron efectivas en:

- Reducir el overfitting
- Mayor equilibrio de clases
- Estabilizar el rendimiento del modelo
- Mejorar la generalización

El análisis indica que aunque se han logrado mejoras significativas en la estabilidad y generalización del modelo, aún hay oportunidades para mejorar su precisión general en la predicción de categorías de servicio para los clientes de telecomunicaciones, buscando mejorar la ingeniería de características, aumento de datos y calidad de los mismos.