

Estrategia de Solución

La estrategia para prever si un cliente continuará utilizando el servicio de una empresa proveedora de internet y telefonía se basó en un enfoque de Machine Learning utilizando modelos supervisados. A continuación, se describen los pasos seguidos para desarrollar esta solución:

1. Carga y Preprocesamiento de Datos:

- **Carga del Dataset:** Se utilizó el archivo Telco-Customer-Churn.csv que contiene información sobre los clientes y si cancelaron o no el servicio.
- **Limpieza de Datos:** Se eliminaron columnas irrelevantes (customerID y gender) y se reemplazaron valores de cadena vacíos con 'Unknown'.
- **Codificación de Variables Categóricas:** Las variables categóricas se convirtieron a valores numéricos utilizando un diccionario de mapeo específico.
- **Eliminación de Valores Faltantes:** Se eliminaron las filas con valores faltantes para asegurar la integridad del análisis.

2. Entrenamiento de Modelos:

- **Separación de Características y Etiquetas:** Se dividieron los datos en características (X) y etiquetas (y), y posteriormente en conjuntos de entrenamiento y prueba.
- **Selección de Algoritmos:** Se entrenaron dos algoritmos de aprendizaje supervisado: Random Forest y Gradient Boosting.
- **Optimización de Hiperparámetros:** Se utilizaron GridSearchCV para encontrar los mejores hiperparámetros para ambos modelos.

3. Evaluación de Modelos:

- **Métricas de Rendimiento:** Se evaluaron los modelos utilizando precisión, matriz de confusión, curvas de aprendizaje e informes de clasificación.
- **Coeficiente Kappa de Cohen:** Se calculó el coeficiente Kappa de Cohen para medir la concordancia entre los dos modelos en sus predicciones.

4. Predicción para Nuevos Usuarios:

- **Entrada de Datos del Usuario:** Se implementó un proceso interactivo para solicitar datos de un usuario específico.
- **Conversión y Predicción:** Los datos del usuario se convirtieron al formato adecuado y se utilizó el modelo de Gradient Boosting para predecir si el usuario cancelará el servicio.

Resultados Obtenidos

- **Precisión del Modelo de Random Forest:** El modelo de Random Forest alcanzó una precisión de 0.79 en el conjunto de prueba.
- **Precisión del Modelo de Gradient Boosting:** El modelo de Gradient Boosting logró una precisión de 0.80 en el conjunto de prueba.
- **Concordancia entre Modelos:** El coeficiente Kappa de Cohen fue de 0.91, indicando una concordancia casi perfecta entre las predicciones de los dos modelos.
- **Curvas de Aprendizaje:** Las curvas de aprendizaje mostraron que ambos modelos alcanzaron un buen equilibrio entre el sesgo y la varianza, con un rendimiento consistente en los conjuntos de entrenamiento y validación.

Conclusiones

- **Rendimiento de los Modelos:** Ambos modelos, Random Forest y Gradient Boosting, mostraron un buen rendimiento en términos de precisión y concordancia en la predicción de la cancelación del servicio por parte de los clientes.
- **Importancia de la Preparación de Datos:** La limpieza y transformación adecuada de los datos fueron cruciales para el rendimiento efectivo de los modelos.
- **Utilidad del Análisis Predictivo:** La capacidad de predecir la cancelación del servicio puede ser extremadamente útil para la empresa, permitiéndole identificar características de clientes leales y tomar medidas proactivas para retener a aquellos que podrían cancelar.
- **Aplicabilidad del Modelo:** El modelo de Gradient Boosting, con una precisión ligeramente superior, fue utilizado para predicciones interactivas de nuevos usuarios, demostrando su aplicabilidad práctica.

En resumen, la estrategia planteada logró desarrollar un sistema predictivo robusto que puede ayudar a la empresa a mejorar sus estrategias de retención de clientes y a identificar factores clave que influyen en la cancelación del servicio.