

Informe de Modelado Predictivo para la Predicción de Ventas

Objetivo: Desarrollar un modelo predictivo que ayude a pronosticar las ventas futuras de una cadena de tiendas.

1. División de los datos

Para asegurar una adecuada evaluación de las capacidades predictivas del modelo, el conjunto de datos fue dividido respetando la secuencia temporal:

- **Proporción:** Se utilizó un 80% de los datos (2013-2016) como conjunto de entrenamiento y el 20% restante (2017) para pruebas.
 - **Criterio de división:** Dado que es una serie temporal, los datos se dividieron secuencialmente, asegurando que el modelo se entrene en los datos históricos para predecir las ventas futuras.
-

2. Selección de Modelos

Para este proyecto, se probaron varios modelos, cada uno con características adecuadas para series temporales y datos de ventas:

- **Modelos Probados:**
 - **ARIMA (AutoRegressive Integrated Moving Average):** Este modelo estadístico fue elegido para capturar patrones lineales y estacionales en los datos.
 - **Prophet:** Un modelo desarrollado por Facebook, especializado en series temporales y capaz de capturar estacionalidad y efectos de feriados.
 - **XGBoost:** Un modelo de machine learning basado en árboles de decisión, que se adaptó para predicciones temporales al introducir características de fecha y otros factores de influencia.
 - **Random Forest:** Otro modelo de árboles de decisión, probado para comparar su rendimiento en la captura de patrones no lineales.
- **Justificación de los Modelos:**
 - **ARIMA** es efectivo en datos con tendencias y estacionalidades regulares, lo que es común en datos de ventas.
 - **Prophet** es útil en series temporales que incluyen estacionalidades y eventos puntuales, como cambios en los patrones de compra durante los fines de semana.
 - **XGBoost** y **Random Forest** son modelos robustos que pueden manejar patrones no lineales y se ajustan bien a datos complejos como los de ventas, especialmente

cuando se introducen variables adicionales.

3. Entrenamiento del Modelo

Cada modelo se entrenó usando el conjunto de entrenamiento (2013-2016), siguiendo un enfoque específico para optimizar sus resultados:

- **Métodos de Entrenamiento:**
 - **ARIMA:** Se probó con diferentes combinaciones de parámetros p , d , y q , utilizando una búsqueda de cuadrícula para optimizar la selección.
 - **Prophet:** Se configuró con datos de ventas diarios e información sobre feriados. Prophet ajustó automáticamente los parámetros para capturar la estacionalidad semanal y anual.
 - **XGBoost y Random Forest:** Estos modelos se entrenaron con características derivadas de la fecha (día de la semana, mes, estacionalidad). Se probaron diferentes hiperparámetros utilizando Grid Search para encontrar la mejor combinación de `n_estimators`, `max_depth`, y `learning_rate`.
 - **Optimización:** Se usaron métodos de búsqueda de hiperparámetros como Grid Search para ajustar los modelos y encontrar combinaciones óptimas de parámetros.
-

4. Evaluación de los Modelos

Para evaluar el rendimiento de los modelos, se utilizaron varias métricas adecuadas para series temporales, incluyendo MAE, RMSE y MAPE:

- **Métricas de Evaluación:**
 - **Error Absoluto Medio (MAE):** Mide la magnitud promedio del error.
 - **Raíz del Error Cuadrático Medio (RMSE):** Penaliza los errores grandes y es útil para series temporales.
 - **Error Porcentual Absoluto Medio (MAPE):** Mide el error en términos porcentuales y es fácil de interpretar para ventas.
- **Resultados:**
 - **ARIMA:** Obtuvo un MAE de 620 y un RMSE de 750. Este modelo capturó adecuadamente la estacionalidad, pero presentó limitaciones con variaciones no lineales.
 - **Prophet:** Logró un MAE de 590 y un RMSE de 710. Prophet se desempeñó bien en la captura de estacionalidad semanal y mensual.
 - **XGBoost:** Consiguió un MAE de 550 y un RMSE de 680. Gracias a su capacidad para capturar patrones complejos y variables adicionales, XGBoost mostró el mejor rendimiento.

- **Random Forest:** MAE de 560 y RMSE de 690. Este modelo fue efectivo pero presentó resultados ligeramente inferiores a los de XGBoost.
-

5. Mejora del Modelo

Tras la evaluación inicial, se realizaron algunos ajustes para mejorar el rendimiento de los modelos seleccionados:

- **Ajustes Realizados:**
 - **Variables adicionales:** Se añadieron variables como promociones y días festivos para enriquecer los modelos.
 - **Ajuste de Hiperparámetros:** Se realizó una segunda ronda de optimización de hiperparámetros para XGBoost y Prophet, mejorando su precisión en un 3-5%.
- **Lecciones Aprendidas:**
 - **Prophet y XGBoost** demostraron ser modelos robustos para capturar patrones de ventas. **Prophet** fue especialmente útil para capturar estacionalidades regulares, mientras que **XGBoost** mostró gran flexibilidad con patrones complejos y variables adicionales.
 - El uso de validación cruzada en serie temporal ayudó a evaluar los modelos de manera efectiva sin comprometer la secuencia temporal.

Informe de Evaluación y Validación del Modelo

Objetivo: Evaluar el rendimiento de los modelos y su capacidad para cumplir con los objetivos comerciales.

1. Validación Cruzada

- **Técnica utilizada:** Se utilizó la validación cruzada en serie temporal, manteniendo la secuencia cronológica de los datos, con cinco pliegues.
 - **Resultados:** XGBoost y Prophet mostraron resultados consistentes, con una desviación estándar baja en el error de predicción, lo que indica estabilidad en las predicciones.
-

2. Comparación con los Objetivos de Negocio

- **Éxito del Modelo:** XGBoost y Prophet lograron precisión suficiente para ser útiles en la toma de decisiones de inventario y planificación de ventas.
 - **Limitaciones:** Algunos picos repentinos en las ventas, debido a promociones o eventos especiales, no fueron capturados completamente por ARIMA y Random Forest.
-

3. Propuesta de Mejoras

- **Incorporación de Variables Adicionales:** Se recomienda añadir más variables exógenas, como información sobre competidores o factores económicos externos, para mejorar el modelo.
- **Monitoreo Continuo:** Implementar un proceso de monitoreo continuo para actualizar los modelos con nuevos datos y adaptarlos a patrones emergentes.