

Proyecto Integrado V - Línea de Énfasis (Entrega 3)

Sergio Andres Rios Gómez
Edwin Alexander Ibarra Ortiz

Refinar indicadores y conclusiones de dashboard: Taiwan Semiconductor Manufacturing Company (TSM)

Contenido

Refinar indicadores y conclusiones de dashboard: Taiwan Semiconductor
Manufacturing Company (TSM) 2

INTRODUCCIÓN..... 3

LINK DE VIDEO DE PRESENTACIÓN 3

METODOLOGÍA 3

 Enriquecimiento de datos 3

CONCLUSIONES 4

INTRODUCCIÓN

En el presente documento se precisa finiquitar la validación al análisis predictivo de los modelos entrenados, teniendo en cuenta que durante el desarrollo de las actividades se hizo uso de varios modelos entre los cuales resaltan XGBoost para clasificación binaria y Arima como propuesta de predecir el valor exacto del precio de cierre en un futuro.

LINK DE VIDEO DE PRESENTACIÓN

https://drive.google.com/file/d/1_pJ5dhhWG_pWjjsgMu2V3Fhdt7QtLTBo/view?usp=sharing

METODOLOGÍA

Enriquecimiento de datos

1. Se implementa un proceso que se encarga de generar el enriquecimiento del DataFrame, donde agregamos tres grupos de categorías (Variables temporales, Indicadores técnicos básicos y Features avanzados):

Variables temporales:

- day_of_week: Día de la semana (0-6) para capturar efectos estacionales
- month: Mes del año (1-12) para identificar patrones mensuales
- year: Año para considerar tendencias anuales
- quarter: Trimestre del año (1-4) para análisis trimestrales

Indicadores técnicos básicos:

- returns: Retornos diarios calculados como $\log(\text{precio}_t / \text{precio}_{t-1})$
- returns_prev_day: Retorno del día anterior para momentum
- volatility_20d: Volatilidad móvil de 20 días usando desviación estándar
- sma_20 y sma_50: Medias móviles simples de 20 y 50 días
- rsi: Índice de Fuerza Relativa de 14 días

Features avanzados: Para mejorar la capacidad predictiva, se desarrollaron características más sofisticadas:

- sma_ratio: Ratio precio actual/SMA-20, indicando desviación de la tendencia
- sma_cross: Señal binaria de cruce entre SMA-20 y SMA-50
- volatility_ratio: Ratio entre volatilidad actual y media móvil de 60 días
- volume_ratio: Ratio entre volumen actual y media móvil de 20 días

- rsi_oversold y rsi_overbought: Señales binarias para condiciones extremas del RSI
- returns_2d y returns_5d: Retornos acumulados de 2 y 5 días

2. Implementación de un análisis exploratorio de datos:

Incluye:

- Distribución de la variable objetivo
- Matriz de correlación entre variables predictoras
- Análisis de series temporales del precio de cierre
- Estadísticas descriptivas de todas las variables
- Visualizaciones guardadas en src/static/eda/

CONCLUSIONES

- En cuanto al comparativo de los dos modelos se puede determinar, que el modelo entrenado con XGBoost como clasificador binario, tiene muy buena precisión, pero como modelo de predicción binaria solo nos puede arrojar datos de afirmación o negación, por lo que para efectos de indicadores financieros puede ser útil pero no determinante ya que carece de la capacidad de arrojar valores de cierre en bolsa.
- El modelo entrenado con Arima, presenta un error porcentual promedio del 9.29% aproximadamente, por lo que podemos decir que es relativamente bueno, ya que, si por ejemplo decimos que el precio de cierre es de 200, podemos tender a una desviación de 18.58.
- Si en lugar de requerir que los modelos construidos con XGBoost y Arima se complementen, es decir que mientras que Arima consigue predecir el precio de cierre de la acción de TSM, nos apoyamos en XGBoost para corroborar si el precio va a subir o bajar
- El modelo construido con ARIMA nos permite estimar si el precio de la acción subirá o bajará. Aunque en la validación observamos que puede fallar en el corto plazo, especialmente en los primeros días del pronóstico, el modelo muestra una buena capacidad para anticipar correctamente la tendencia general de alza a lo largo del tiempo.
- Inicialmente el modelo con Arima no corresponde a una serie estacionaria y esto limita su capacidad predictiva, por ello es por lo que es necesario realizar transformación en los datos, en este caso aplicando diferenciación para poder

estabilizar la medida y eliminar las tendencias. Esto es imprescindible para que el modelo pueda entrenarse correctamente y generar pronósticos más precisos