

Desarrollo de una alerta al identificar un impacto en el precio de la acción de Apple en relación con el insider trading.

Autor: Santiago Rodriguez

Abril - 2025



1. OBJETIVO Y VISIÓN



- Analizar el impacto de las transacciones de compra/venta de acciones realizadas por altos ejecutivos, miembros de la junta directiva y empleados de empresas listadas en la NYSE (como posibles señales de confianza o riesgo interno) en el precio de las acciones de Apple, entre 2015 y marzo de 2025.
- El estudio busca identificar patrones, correlaciones temporales y efectos cuantificables en el valor de la compañía, con el fin de aportar datos estratégicos para decisiones de inversión basadas en el comportamiento de los *insiders*.
- En este sentido, he optado por este tema porque no he encontrado estudios o indicadores que correlacionen el insider trading con el precio de las acciones. Con el desarrollo de este proyecto, quiero crear una especie de alerta ante los movimientos de acciones por parte de los directivos y considerar ese indicador para futuras compras o ventas de acciones dentro de un portafolio de acciones

2. CONTEXTO Y ALCANCE



- El insider trading se refiere a las transacciones de compra o venta de acciones realizadas por altos ejecutivos y empleados de las compañías que cotizan en el mercado bursátil. Estas personas tienen la ventaja de acceder a información antes de que se haga pública, lo que les permitiría obtener ganancias.
- **Demanda insatisfecha:** Actualmente, los inversores analizan transacciones de insiders de forma aislada (por empresa), pero no existe un modelo consolidado que vincule estos datos multisectoriales con el desempeño de un gigante como Apple. Este vacío representa una ventana única para capturar señales tempranas de mercado
- Las empresas listadas en la NYSE representan un valor de mercado agregado de más de **\$25 billones (USD)**. Las transacciones de *insiders* en estas compañías suman miles de millones anuales, actuando como un termómetro oculto de la confianza corporativa en sectores clave vinculados a Apple (tecnología, retail, manufactura, etc.).

Impactos

- **Decisión estratégica:** Identificar patrones de compra/venta de *insiders* que históricamente hayan precedido subidas o caídas en Apple (ej.: ventas masivas en empresas proveedoras antes de una corrección en tech).
- **Gestión proactiva de riesgos:** Detectar si movimientos de ejecutivos en socios estratégicos (ej.: fabricantes de chips, distribuidores) podrían anticipar cuellos de botella o cambios en la demanda de productos.
- **Refuerzo de credibilidad:** Demostrar cómo factores externos (no financieros) impactan su valor

Limitaciones

- En este estudio se incluyen únicamente precios de acciones de Apple desde el 2015 hasta marzo 2025
- Tiempo, capacidad de procesamiento, definir más empresas clave de la industria para ampliar modelo, almacenamiento.

3. ENTENDIMIENTO DE DATOS



Fuentes de datos:

API de Yahoo Finance

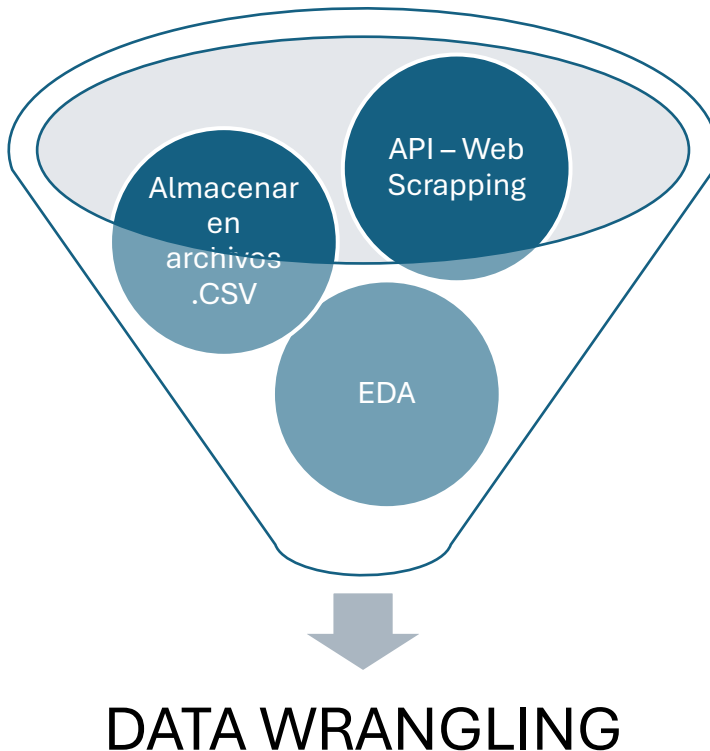
- No es oficial de Yahoo. La descontinuaron en 2017
- Obtenida a través de un desarrollador de github (1).
- Incluye ajuste de precios por split, información precisa.
- Frecuencia de actualización diaria

OPENINSIDER.COM

- Esta página recopila todos los formularios llenados por los altos ejecutivos de compañías que cotizan en bolsa de valores, extraído de la página de la SEC y los consolida en una tabla con un formato tabular. Su frecuencia de actualización es diaria.

(1) API de Yahoo Finance: GitHub - ranaroussi/yfinance: Download market data from Yahoo! Finance's API

4. PREPARACIÓN DE DATOS



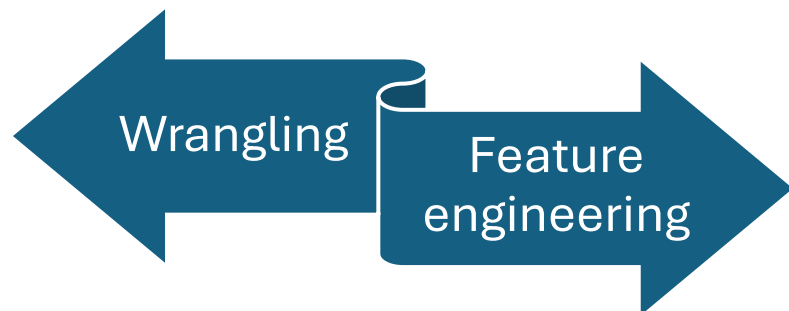
Yahoo Finance (YF)

- Precios se redondean a 2 decimales
- Crear columna de movimiento = $(PC - PA)$
- Eliminar columnas (High, Low, Volume)

OpenInsider.com (OI)

- Delimitar información desde 2015 a marzo 2025.
- Eliminar columnas (filing_date, 1d, 1w, 1m, 6m)
- Mantener un solo cargo por línea, en el caso de presentar más cargos, coloco el primero como predominante.
- Estandarizar tipos de transacción: S – Sale o P – Purchase
- Inner join con data limpia de YH - OI

4. PREPARACIÓN DE DATOS



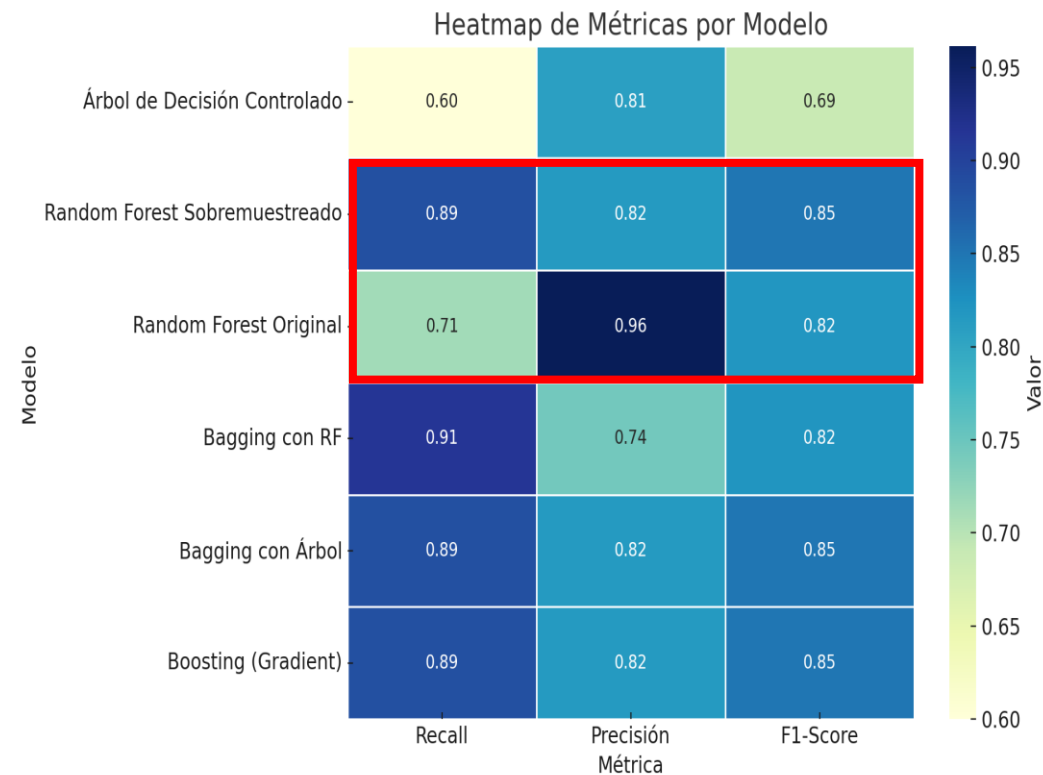
- Definir métricas en base a “delta_owned” $\geq 10\%$
- Creación de columna “transacción significativa” de tipo boolean
- Métrica en base a “movimiento” < 0
- Creación de columna “impacto negativo” de tipo boolean (**variable objetivo**)
- Dar formato a la fecha y separar por mes y año.

Resultado final para comenzar modelado:

	trade_date	ticker	insider_name	title	trade_type	quantity_of_shares	owned	delta_owned	value	Open	Close	movimiento	significant_transaction	impacto_negativo	año	mes
0	2015-01-23	AAPL	Riccio Daniel J.	SVP	S	3804	0	100	428955	24.98	25.13	0.15	1	False	2015	1
1	2015-02-18	AAPL	Jung Andrea	Dir	S	40000	14595	73	5125200	28.50	28.75	0.25	1	False	2015	2
2	2015-03-06	AAPL	Maestri Luca	CFO	S	3400	14124	19	437920	28.68	28.27	-0.41	1	True	2015	3
3	2015-03-09	AAPL	Maestri Luca	CFO	S	2800	11324	20	361116	28.58	28.39	-0.19	1	True	2015	3

5. MODELADO

- Por el problema planteado, se trabajará con modelos de clasificación
- Existió desbalanceo de clases moderado (22%) (20% - 40% sobremuestreo es opcional)
- Metodología train/test Split: entrenamiento / validación / prueba



6. ANÁLISIS DE DESEMPEÑO



¿Por qué?

- Porque me importa capturar **la mayor cantidad posible de casos donde sí habrá un impacto negativo**. Si deo pasar estos casos, el sistema no cumple su función de alerta.

Precisión es menos prioritaria (pero no irrelevante).

- Aunque no quiero alertas falsas todo el tiempo, **es más costoso dejar pasar una alerta que podría haber evitado una caída** en el precio de la acción. Pero ojo, si la precisión baja demasiado, el sistema será ignorado.

Modelo	Recall	Precision	F1-Score	Predicciones Negativas
Random Forest Sobremuestreado	0.886	0.816	0.849	38
Bagging con Random Forest	0.914	0.744	0.821	43
Random Forest Original	0.714	0.962	0.820	26
Boosting	0.886	0.816	0.849	38

6. ANÁLISIS DE DESEMPEÑO



Tabla Resumen Anexa (Detalles Técnicos)

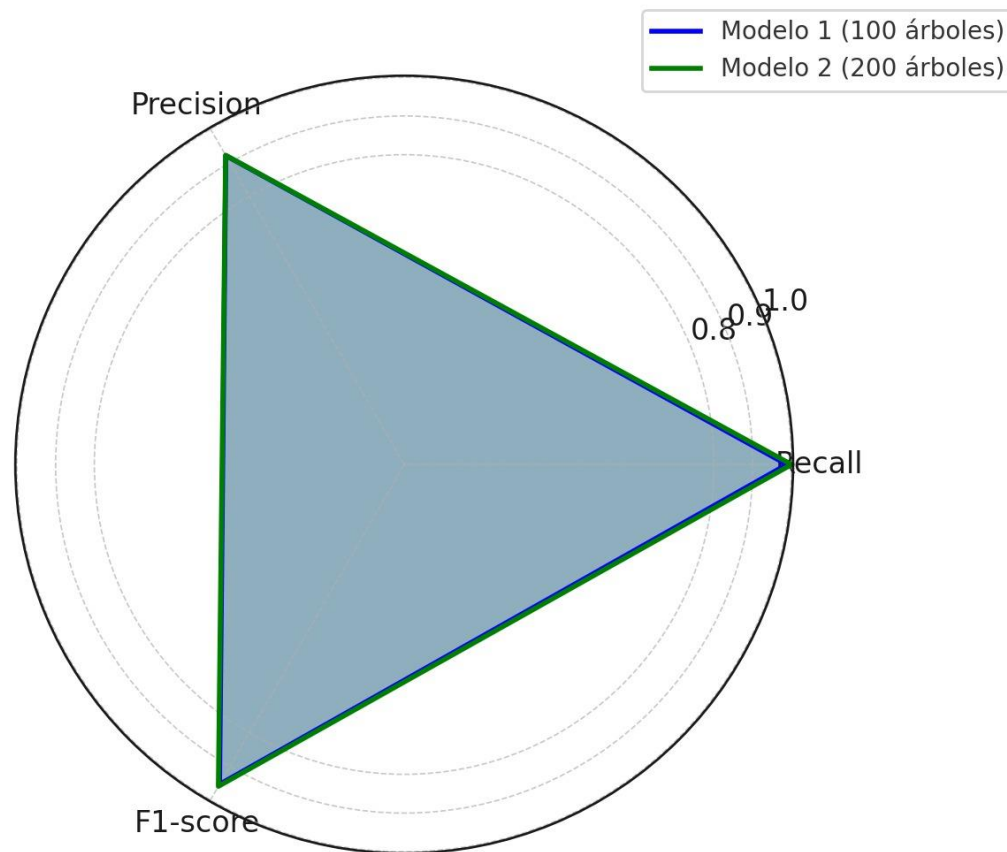
Parámetro	Modelo Original	Modelo Sobremuestreado	Modelo Optimizado
n_estimators	500	500	200
max_depth			None
min_samples_split			2
min_samples_leaf			1
max_features			sqrt

Métrica	Modelo Original	Modelo Sobremuestreado	Modelo Optimizado
Recall	0.714	0.885	1.0
Precision	0.961	0.815	0.922

```
for n_arboles in range(100, 1001, 100): # n_estimators
    for profundidad in [None, 5, 10, 20]:
        for min_split in [2, 5, 10]:
            for min_leaf in [1, 2, 4]:
                for max_feat in ['sqrt', 'log2', None]:
                    modelo_rf = RandomForestClassifier(
                        n_estimators=n_arboles,
                        max_depth=profundidad,
                        min_samples_split=min_split,
                        min_samples_leaf=min_leaf,
                        max_features=max_feat,
                        random_state=12345
```

6. ANÁLISIS DE DESEMPEÑO

Comparación de Métricas por Modelo (Radar)



7. CONCLUSIONES Y RECOMENDACIONES



Conclusiones

- El ajuste de hiperparámetros permitió un modelo que detecta el 100% de los riesgos (Recall=1.0) sin sacrificar significativamente la precisión (92.2%), superando el COSTO-BENEFICIO clásico entre sensibilidad y exactitud.
- El incremento de n_estimators (200 árboles) y la profundidad ilimitada permitieron al modelo capturar interacciones complejas en los datos de insiders, que los enfoques anteriores subestimaban.
- Este nivel de Recall garantiza que ningún evento negativo relevante pasará desapercibido en la estrategia de inversión, mitigando riesgos ocultos.
- Aunque la precisión bajó levemente (96.1% → 92.2%), el trade-off es estratégico: prefiero 8 falsas alertas por cada 100 predicciones antes que pasar por alto un riesgo real
- La profundidad ilimitada (max_depth=None) y los 200 árboles (n_estimators=200) permitieron descubrir patrones no lineales.

Recomendaciones

- Integrar el modelo en sistemas de monitoreo, priorizando alertas de proveedores clave (ej: TSMC) y sectores críticos (tech).
- Simular escenarios extremos (ej: caída del 20% en tech) para calibrar el modelo
- Realizar pruebas piloto con capital simulado (backtesting 2023-2024) para cuantificar ganancias potenciales. Ej: "Si hubiéramos actuado con las alertas del modelo en 2023, ¿cuántas pérdidas por caídas de Apple >5% se habrían evitado?"



MUCHAS GRACIAS POR SU ATENCIÓN