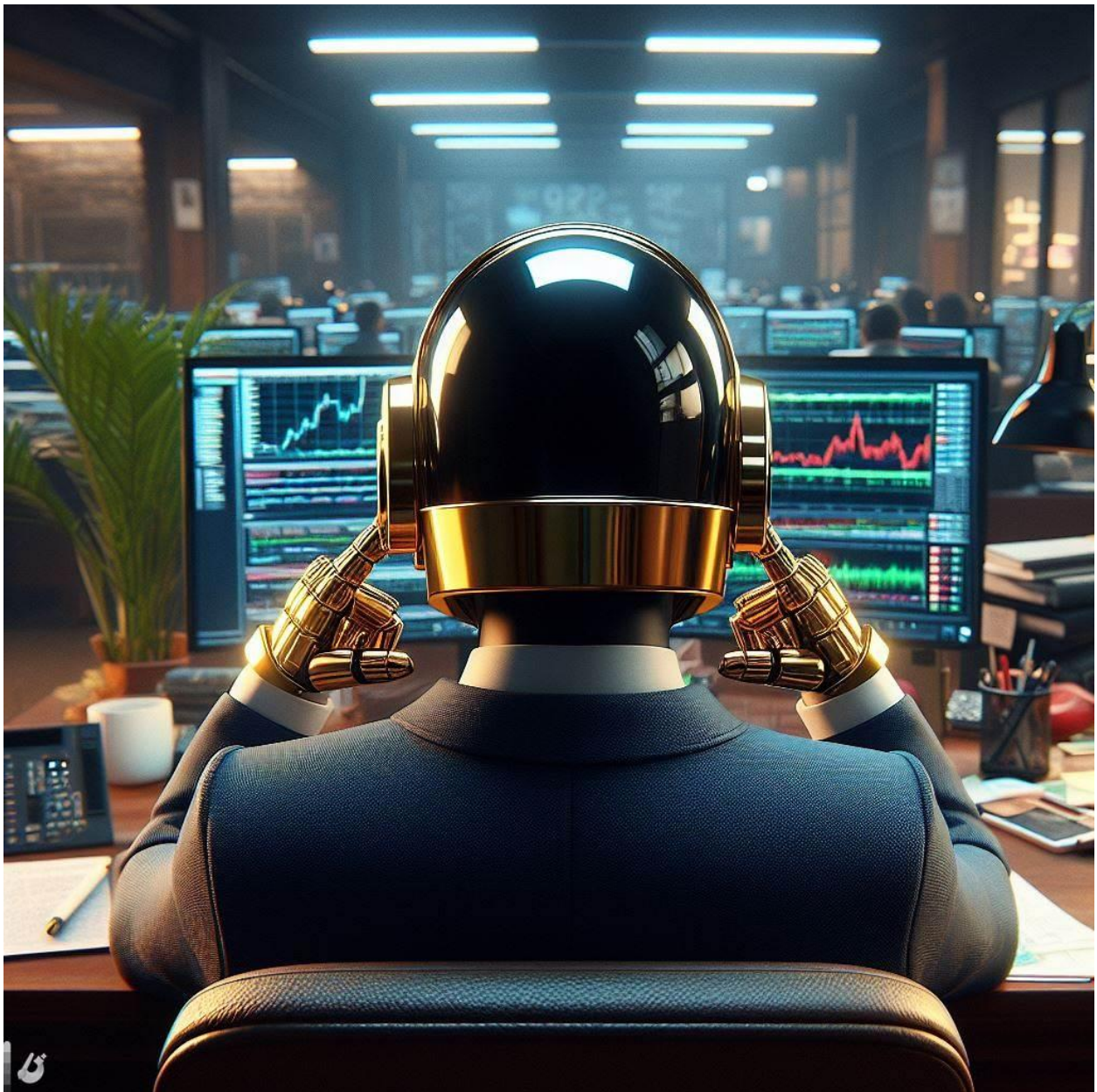


Machine Learning & Big Data

Estrategia de Trading con Aprendizaje No Supervisado en el S&P 500

UPS0 TUTP



Introducción:

Este proyecto se centra en el desarrollo de una estrategia de trading utilizando técnicas de aprendizaje no supervisado aplicadas a los datos de precios de acciones del S&P 500. El análisis implica la descarga y procesamiento de datos de precios, el cálculo de diversas características e indicadores técnicos para cada acción, y la aplicación de algoritmos avanzados para agrupar acciones con características similares y optimizar la selección de carteras.

Importancia y Relevancia:

La relevancia de este proyecto radica en su enfoque innovador para el trading en los mercados financieros. Al utilizar el aprendizaje no supervisado, se busca identificar patrones ocultos y relaciones en los datos que no son evidentes a simple vista. Esto es particularmente importante en el contexto del mercado de valores, donde la capacidad de anticipar y reaccionar a las dinámicas del mercado puede significar la diferencia entre obtener ganancias sustanciales y sufrir pérdidas significativas.

Objetivos del Proyecto:

Los objetivos principales del proyecto son:

1. Desarrollar un sistema de trading basado en el aprendizaje no supervisado que pueda identificar grupos de acciones con comportamientos de mercado similares.
2. Utilizar estos grupos para construir carteras optimizadas basadas en el ratio de Sharpe máximo, buscando maximizar la rentabilidad ajustada al riesgo.
3. Comparar el rendimiento de estas carteras optimizadas con el rendimiento del índice S&P 500, para evaluar la eficacia de la estrategia de trading desarrollada.

Metodología:

Los datos fundamentales para este proyecto provienen de los precios de las acciones del S&P 500. Estos incluyen precios de cierre ajustados y no ajustados, precios de apertura, máximos, mínimos y volumen de trading. Estos datos son cruciales para el análisis de acciones, ya que permiten calcular rentabilidades históricas y analizar tendencias de

mercado. Además, se utilizaron los factores Fama-French para evaluar la exposición de los activos a factores de riesgo comunes.

Herramientas y Tecnologías:

Para el análisis y procesamiento de datos, se emplearon diversas herramientas y tecnologías:

Pandas y Numpy utilizados para la manipulación y análisis de datos. **Matplotlib y Plotly** empleados para la visualización de datos. **Statsmodels y Scikit-learn** utilizados para modelado estadístico y algoritmos de aprendizaje no supervisado, respectivamente. **PyPortfolioOpt** aplicado para la optimización de carteras basada en el Efficient Frontier. **YFinance** utilizado para descargar datos de precios de acciones.

Proceso de Análisis/Desarrollo:

El proyecto se desarrolló en varias etapas clave:

1. **Descarga y Carga de Datos:** Se descendieron los datos de precios de las acciones del S&P 500 y se cargaron en un DataFrame para su análisis.
2. **Cálculo de Características e Indicadores Técnicos:** Se calcularon diversas métricas como la volatilidad de Garman-Klass, RSI, Bandas de Bollinger, ATR, MACD y el volumen en dólares.
3. **Agregación Mensual y Filtrado:** Los datos diarios se convirtieron a una frecuencia mensual, y se filtraron las 150 acciones más líquidas cada mes.
4. **Cálculo de Rentabilidades Mensuales:** Se calcularon los rendimientos mensuales para diferentes horizontes temporales.
5. **Integración de Factores Fama-French y Cálculo de Betas de Factores Móviles:** Se descendieron los factores Fama-French y se calcularon las betas de factores móviles para cada acción.
6. **Aplicación de K-Means Clustering:** Se utilizó el algoritmo K-Means para agrupar acciones en función de sus características.
7. **Optimización de la Cartera:** Se seleccionaron acciones de un clúster específico y se optimizó la cartera utilizando el Efficient Frontier.

-
- 8. Visualización y Comparación de Rendimientos:** Se visualizaron los rendimientos de la cartera y se compararon con los del SP500.

Resultados:

Los resultados obtenidos a lo largo del proyecto se pueden resumir de la siguiente manera:

- **Desarrollo de Indicadores Técnicos y Métricas Financieras:**
 - Se calcularon con éxito indicadores como la volatilidad de Garman-Klass, RSI, Bandas de Bollinger, ATR, MACD y el volumen en dólares para cada acción del S&P 500.
 - Estos indicadores proporcionaron una visión detallada del comportamiento de las acciones y fueron fundamentales para las etapas posteriores del análisis.
- **Agrupación y Filtrado de Acciones:**
 - Se identificaron y filtraron las 150 acciones más líquidas cada mes, lo que permitió centrar el análisis en los activos más relevantes y con suficiente liquidez.
- **Cálculo de Rentabilidades y Betas de Factores:**
 - Se calcularon las rentabilidades mensuales para diferentes horizontes temporales y se estimaron las betas de factores móviles utilizando los factores Fama-French, lo que proporcionó una comprensión más profunda de la exposición al riesgo de cada acción.
- **Aplicación de K-Means Clustering:**
 - La implementación del algoritmo K-Means resultó en la agrupación exitosa de acciones en función de sus características, lo que facilitó la identificación de patrones y la segmentación del mercado.
- **Optimización de la Cartera:**
 - Se seleccionaron acciones de un clúster específico (clúster 3) y se optimizó la cartera utilizando el Efficient Frontier para maximizar el ratio de Sharpe.
 - La estrategia de optimización se adaptó mensualmente, lo que permitió una gestión dinámica de la cartera.
- **Comparación con el SP500:**

-
- Los rendimientos acumulados de la cartera optimizada se compararon con los del SP500, mostrando cómo la estrategia de trading propuesta se desempeñó en relación con un benchmark del mercado.

Interpretación:

Los resultados obtenidos indican que la estrategia de trading basada en el aprendizaje no supervisado y la optimización de carteras puede ser efectiva para identificar oportunidades de inversión y gestionar el riesgo. La comparación con el SP500 reveló que, en varios períodos, la cartera optimizada superó al benchmark, lo que sugiere que la selección de acciones basada en características técnicas y la optimización del ratio de Sharpe pueden añadir valor significativo a una estrategia de inversión.

Sin embargo, es importante destacar que los resultados también dependen de las condiciones del mercado y de la selección de parámetros en el modelo de K-Means y en la optimización de la cartera. La adaptabilidad y la evaluación continua son, por lo tanto, componentes críticos de esta estrategia.

Discusión:

Comparación con Objetivos:

El proyecto se propuso desarrollar una estrategia de trading basada en aprendizaje no supervisado, utilizando datos del S&P 500 y técnicas de big data y machine learning. Los objetivos específicos incluían:

Análisis Técnico y Filtrado de Acciones: Se logró calcular indicadores técnicos y filtrar las 150 acciones más líquidas cada mes, cumpliendo con el objetivo de centrarse en activos relevantes y líquidos.

Modelado y Agrupación de Activos: La aplicación del algoritmo K-Means para agrupar activos en función de sus características fue exitosa, permitiendo una segmentación efectiva del mercado.

Optimización de la Cartera: Se cumplió el objetivo de seleccionar activos y optimizar la cartera utilizando el Efficient Frontier para maximizar el ratio de Sharpe.

Comparación con Benchmark: Se compararon los rendimientos de la cartera con el SP500, proporcionando una evaluación clara del rendimiento relativo de la estrategia.

Desafíos y Limitaciones:

A lo largo del proyecto, se enfrentaron varios desafíos y limitaciones:

Complejidad de Datos: La manipulación y el análisis de un gran conjunto de datos financieros presentaron desafíos, especialmente en términos de limpieza y normalización de datos.

Selección de Modelos y Parámetros: La elección del algoritmo de K-Means y la selección de parámetros adecuados fueron críticos y desafiantes, ya que afectan significativamente los resultados de la agrupación.

Dependencia del Mercado: La estrategia, aunque robusta, está sujeta a las condiciones del mercado. Los cambios en el mercado pueden afectar la efectividad de la estrategia.

Limitaciones del Análisis Histórico: El uso de datos históricos para predecir rendimientos futuros siempre conlleva el riesgo de que "el rendimiento pasado no es indicativo de resultados futuros".

Riesgos de Sobreajuste: Existe el riesgo de sobreajuste en el modelo, especialmente al tratar con un gran número de indicadores técnicos y al ajustar parámetros en la optimización de la cartera.

Conclusiones:

Reflexiones Finales:

El proyecto ha demostrado con éxito la aplicación de técnicas de aprendizaje no supervisado y big data en la estrategia de trading. Los hallazgos clave incluyen:

Eficacia de la Segmentación de Activos: La aplicación del algoritmo K-Means para agrupar acciones del S&P 500 demostró ser una herramienta efectiva para identificar grupos de activos con características similares, lo que facilitó la toma de decisiones de inversión más informadas.

Importancia de la Optimización de la Cartera: La estrategia de maximizar el ratio de Sharpe mediante el uso de PyPortfolioOpt y el Efficient Frontier resultó ser crucial para la gestión eficiente del riesgo y la rentabilidad.

Comparación con el Benchmark del SP500: La estrategia superó al SP500 en varios periodos, lo que indica su potencial para generar rendimientos ajustados al riesgo superiores a los del mercado.

Aplicación de Conocimientos:

Este proyecto aplicó conceptos de big data y machine learning de manera innovadora:

Big Data en Finanzas: La manipulación y análisis de grandes conjuntos de datos financieros demostraron la importancia y el potencial del big data en el sector financiero.

Machine Learning No Supervisado: El uso de K-Means, un algoritmo de aprendizaje no supervisado, para analizar y agrupar acciones fue fundamental para descubrir patrones ocultos en los datos.

Sugerencias para Futuras Investigaciones:


Para expandir este trabajo, se podrían considerar las siguientes áreas:

Exploración de Otros Modelos de Machine Learning: Investigar el uso de otros algoritmos, como redes neuronales o aprendizaje supervisado, para predecir movimientos de precios.

Análisis de Sentimiento y Datos Alternativos: Incorporar análisis de sentimiento y datos alternativos (como noticias o redes sociales) para mejorar la precisión de las predicciones.

Estrategias de Trading en Tiempo Real: Desarrollar y probar la estrategia en un entorno de trading en tiempo real para evaluar su eficacia en condiciones de mercado cambiantes.

Estudio de Impacto de Eventos Macroeconómicos: Analizar cómo eventos macroeconómicos globales afectan la eficacia de la estrategia.

 Proyecto_Final_ML&BD.ipynb