**4. Estado del Arte de la Investigación**

**Título del Proyecto**

**Desarrollo de un Sistema Big Data para la Visualización y Análisis de Datos Financieros de ETF**

En la actualidad, el mercado financiero se ha convertido en un entorno altamente dinámico, donde la velocidad y la precisión en la toma de decisiones son factores determinantes. La adopción de tecnologías Big Data ha facilitado el procesamiento de grandes volúmenes de información en tiempo real para optimizar estrategias de inversión. Esta investigación se enmarca en la necesidad de utilizar herramientas y metodologías que permitan mejorar la precisión en las decisiones financieras, especialmente en relación al ETF S&P 500 (SPY). Para ello, se ha estructurado la revisión del estado del arte bajo las fases clave del procesamiento de datos: ingesta, almacenamiento, procesamiento y visualización.

**Fase de Ingesta de Datos**

La ingesta de datos es la etapa inicial donde se recopila información estructurada y no estructurada desde múltiples fuentes. Una herramienta ampliamente utilizada para esta tarea es Apache NiFi, la cual permite flujos de datos automatizados y seguros. Según Apache NiFi (2024), su arquitectura basada en flujos permite integrar datos de APIs como Twitter, Reddit o plataformas financieras como Alpha Vantage.

Además, las APIs desempeñan un papel fundamental en esta etapa. Por ejemplo, Twitter Developer Platform (2024) ofrece la API v2 que facilita la recolección de opiniones en tiempo real sobre tendencias del mercado. Asimismo, Reddit proporciona acceso a discusiones especializadas de inversores minoristas a través de Reddit API Documentation (2024). Por su parte, Alpha Vantage (2024) entrega datos estructurados en tiempo real sobre precios, volumen y otras métricas bursátiles fundamentales para el análisis del ETF SPY.

**Fase de Almacenamiento de Datos**

En cuanto al almacenamiento, se prioriza el uso de bases de datos NoSQL, como MongoDB, que permiten una alta escalabilidad y flexibilidad frente a datos semi-estructurados y no estructurados. MongoDB, Inc. (2024) resalta su capacidad para gestionar colecciones de gran tamaño con baja latencia, lo que lo convierte en una opción óptima para entornos financieros que demandan disponibilidad continua.

**Fase de Procesamiento de Datos**

Una vez almacenados, los datos deben ser procesados para extraer valor. Apache Spark ha sido reconocido como uno de los motores más eficientes para el procesamiento distribuido en tiempo real. Apache Spark (2024) permite ejecutar algoritmos de análisis predictivo y aprendizaje automático a gran escala, facilitando la detección de patrones de comportamiento en activos bursátiles como el SPY.

Investigaciones recientes han demostrado el potencial del Big Data en la predicción bursátil. Por ejemplo, Balbaa et al. (2024) exploraron la analítica en tiempo real para el pronóstico del mercado financiero mediante Big Data, destacando la importancia del flujo constante de información para decisiones oportunas. Thanasas y Kampiotis (2024) también destacan el rol de la analítica Big Data en la toma de decisiones contables y financieras estratégicas.

**Fase de Visualización de Datos**

Finalmente, la visualización cumple un rol crucial para interpretar resultados analíticos. Herramientas como Metabase permiten generar dashboards interactivos que facilitan la comprensión de indicadores clave (Metabase, 2024). Saberironaghi (2025) resalta cómo la representación visual de datos no estructurados y resultados de modelos predictivos mejora la toma de decisiones financieras.

Complementariamente, la integración de técnicas de inteligencia artificial como RNN-LSTM con análisis de sentimiento también ha demostrado mejorar las predicciones bursátiles (Kasture & Shirsath, 2024). En este sentido, Dakalbab (2024) sintetiza avances en IA aplicada al trading en un análisis sistemático, subrayando la convergencia entre aprendizaje automático y análisis de datos financieros.

El análisis del estado del arte permite evidenciar una tendencia creciente en la aplicación de tecnologías Big Data en entornos financieros. Las herramientas para ingesta, almacenamiento, procesamiento y visualización se integran de manera coherente en arquitecturas modernas, permitiendo decisiones más informadas en activos como el ETF S&P 500 (SPY). La revisión bibliográfica valida la viabilidad técnica de implementar un sistema de análisis basado en estas tecnologías.

**Referencias Bibliográficas**

1. Alpha Vantage Inc. (2024). Stock APIs.<https://www.alphavantage.co/documentation/>
2. Apache NiFi. (2024). Overview.<https://nifi.apache.org/>
3. Apache Spark. (2024). Apache Spark – Unified Analytics Engine.<https://spark.apache.org/>
4. Balbaa, M. E., Tashtemirovich, A. O., & Ismailova, N. B. (2024). Real‑time analytics in financial market forecasting: A big data approach. Universidad de Economía de Tashkent.<https://www.researchgate.net/publication/380543715_Real-time_Analytics_in_Financial_Market_Forecasting_A_Big_Data_Approach>
5. Dakalbab, F. (2024). Artificial intelligence techniques in financial trading: A systematic literature review. Future Generation Computer Systems.<https://www.sciencedirect.com/science/article/pii/S1319157824001046>
6. Kasture, P., & Shirsath, K. (2024). Enhancing stock market prediction: A hybrid RNN‑LSTM framework with sentiment analysis. Indian Journal of Science and Technology, 17(29), 1–12.<https://indjst.org/articles/enhancing-stock-market-prediction-a-hybrid-rnn-lstm-framework-with-sentiment-analysis>
7. Metabase. (2024). Open source business intelligence.<https://www.metabase.com/>
8. MongoDB, Inc. (2024). What is MongoDB?<https://www.mongodb.com/what-is-mongodb>
9. Reddit API Documentation. (2024).<https://www.reddit.com/dev/api/>
10. Saberironaghi, M. (2025). Stock market prediction using machine learning and unstructured data. MDPI Open Journal, 5(3), 76.<https://www.mdpi.com/2673-9909/5/3/76>
11. Thanasas, G. L., & Kampiotis, G. (2024). The role of big data analytics in financial decision‑making and strategic accounting. Technium Business and Management, 10, 17–33.<https://www.researchgate.net/publication/386077268_The_role_of_Big_Data_Analytics_in_Financial_Decision-Making_and_Strategic_Accounting>
12. Twitter Developer Platform. (2024). Twitter API v2 Reference.<https://developer.twitter.com/en/docs/twitter-api>