Research Report: Pipeline 3 - The Transformer Timelord

Project Lead: Ayan
Date: September 15, 2025
Objective: To test the hypothesis that a state-of-the-art Transformer architecture could outperform the champion model from Pipeline 1 (V5 Directional Accuracy: 55.59%).

1. Executive Summary

This report details the development and final results of Pipeline 3, a research track focused on applying a Transformer-based deep learning model to the stock price forecasting problem. The core hypothesis was that the Transformer's self-attention mechanism could identify complex, long-range dependencies in financial time series that were missed by both the feature-based model of Pipeline 1 and the recurrent models of Pipeline 2.

Despite a ground-up re-engineering effort that included a rich feature set, advanced hyperparameter tuning with Bayesian Optimization, and a robust training regimen, the final model failed to produce a predictive edge. The best-performing version of the pipeline achieved a **Directional Accuracy of 49.75%**, a result equivalent to a random coin flip and a conclusive failure.

Conclusion: The hypothesis is **invalidated**. For this specific, noisy forecasting problem, the Transformer architecture, despite its theoretical power, proved to be a suboptimal and ineffective strategy.

2. Methodology & Evolution

The pipeline was built from the ground up, incorporating best practices from professional MLOps and lessons learned from previous pipelines.

2.1. Initial Implementation (V1)

- **Architecture:** A standard Transformer Encoder model.
- **Features:** A simple set of features including returns and basic technical indicators.
- **Results:**
  - **Runtime:** ~17 minutes
  - **RMSE:** 2.9214
  - **MAE:** 2.2466
  - **Directional Accuracy:** 50.00%
- **Analysis:** The initial run was fast but had zero predictive power, indicating severe underfitting. The model was not complex enough, nor was it given enough meaningful information to learn from.

2.2. Advanced Implementation (V2 - Final Version)

Based on the failure of V1 and a user-provided inspiration script, the pipeline was completely re-engineered.

- **Key Upgrades:**
  1. **Enhanced Feature Set:** Incorporated more powerful indicators like **MACD** and **Bollinger Bands**, and used more responsive, shorter-term windows for Google Trends features. Outliers were also clipped more aggressively to reduce noise.
  2. **Smarter Hyperparameter Tuning:** Switched from RandomSearch to the more intelligent **BayesianOptimization** tuner to find a better model architecture.
  3. **More Sophisticated Training:** Implemented a **Learning** Rate Scheduler (**ReduceLROnPlateau)** to help the model train more effectively.
  4. **More Adaptive Walk-Forward:** Increased the retraining frequency to every 30 days to adapt more quickly to recent market changes.
- **Final Results:**
  - **Runtime:** ~4 hours 5 minutes
  - **RMSE:** 24.4762
  - **MAE:** 19.8263
  - **Directional Accuracy:** 49.75%

3. Analysis of Failure

The final results are definitive. The advanced Transformer model not only failed to find a signal but demonstrated **severe instability**, as evidenced by the extremely high RMSE and MAE values.

The likely cause is **catastrophic overfitting**. The Transformer's self-attention mechanism is so powerful that, when presented with a high-noise, low-signal dataset like financial time series, it can easily begin to "memorize" the random noise in the training data. This creates a model that is perfectly tuned to the past but has zero ability to generalize to new, unseen data.

The model's complexity, which is its greatest strength in other domains (like natural language), became its greatest weakness here.

4. Final Conclusion

The research conducted in Pipeline 3 was a valuable and necessary final step in our project. We have now rigorously tested the current state-of-