**УДК 004.855.91:519.2:336.76**
**DOI: 10.18523/2617-3808.2025.8.38-42**

*M. Voitishyn, D. Kuzmenko*

# A HYBRID AI MODEL FOR FINANCIAL MARKET PREDICTION

*Financial time series modeling is increasingly complex due to volatility, unexpected breakouts, and the impact of external factors, such as macroeconomic indicators, investor sentiment, company fundamentals, and extreme shocks, like geopolitical events or market manipulations. This paper introduces a hybrid artificial intelligence framework that integrates traditional statistical methods, machine learning models, and Bayesian neural networks (BNNs) to improve predictive performance and uncertainty quantification in financial forecasting. The model leverages a variety of engineered features, including rolling statistics, technical indicators, anomaly scores, interpolated macroeconomic data, and transformer-based sentiment scores.*

*A complete ablation study compares various architectures, including ARIMA, SARIMA, MLR, SNN, and BNN, across multiple prediction windows (1, 3, 5 days) and feature combinations. Results show that while linear models yield the lowest MSE for short-term predictions, they fail to capture non-linear dependencies and uncertainty. In contrast, BNNs offer more reliable mid-term predictions by estimating predictive distributions. The best BNN configuration (Normal distribution, constant variation, TanH activation, 1 hidden layer) achieved an MSE of 0.00022, confirming the advantage of uncertainty-adjusted modeling. Sentiment analysis and anomaly detection were especially impactful when combined with macroeconomic indicators, improving signal reliability and behavioral insight.*

*Our findings highlight the importance of integrating diverse data sources and accounting for predictive uncertainty in financial applications. Additionally, the experiments revealed that compact network architectures often outperform deeper ones when paired with engineered features. All experiments were systematically tracked to ensure reproducibility and facilitate future model benchmarking.*

**Keywords:** probability theory, Bayesian neural networks, financial analysis, uncertainty quantification, anomaly detection, time-series forecasting.

## Introduction

In this study, we conduct an ablation analysis to evaluate how different model architectures and feature sets impact financial time series prediction. The objective is to better understand the contributions of various components under real-world financial uncertainty.

We experiment with two key assumptions:

1. Modeling uncertainty improves prediction robustness – by comparing Bayesian Neural Networks with traditional models, we explore how estimating predictive distributions, including location and scale, enhances forecast reliability.
2. Behavioral signals embedded in sentiment data are predictive – by integrating sentiment analysis extracted from Twitter using transformer models, we assess whether these features improve financial forecasts.

## Methods

**Feature engineering.** We developed a diverse feature set grouped into: market data, macroeconomic indicators, technical features, datetime variables, and anomaly sentiment signals. For baseline data: raw OHLCV was used, where prices were adjusted for stock splits. Macroeconomic Indicators were parsed from Federal Reserve Economic Data(FRED) database and included variables such as CPI, GDP, and Unemployment Rate, interpolated to daily frequency, We focused heavily on extra features: Technical indicators (SMA, EMA, RSI), anomaly scores (IQR and Isolation Forest), sentiment signals, datetime features. Each feature type was normalized and merged into a comprehensive dataset with a unified target — Close Price Daily Return.

**Anomaly detection.** Continuous anomaly scores were computed using the Isolation Forest algorithm, which isolates outliers by random partitioning data. Discrete anomaly scores detect unusual market behavior using interquartile range (IQR) filtering.

**Sentiment extraction.** Sentiment scores were derived from Twitter financial text sources and incorporated as predictive features. These scores provide insights into market psychology, capturing reactions to news, earnings, or macroeconomic developments, and offer a complementary perspective to purely numerical indicators.

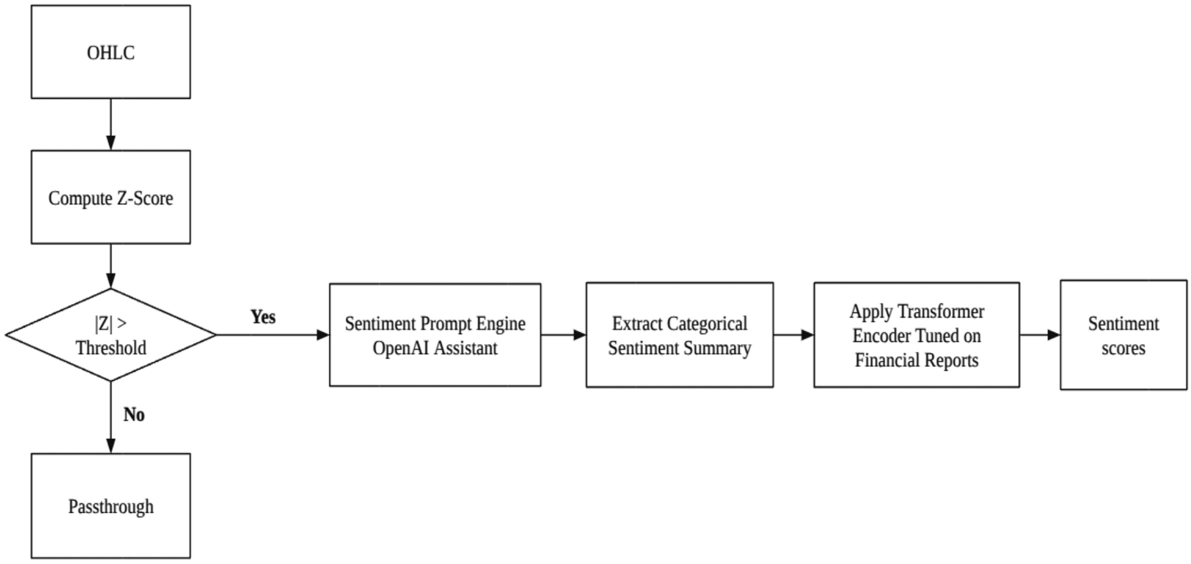


**Figure 1.** Sentiment Extraction Flowchart

Based on the sentiment extraction flow chart above, the following example, generated using an OpenAI Assistant, illustrates how social media sentiment was quantified and integrated into the forecasting pipeline:

*"On November 5, 2015, Nvidia (ticker: NVDA) experienced a notable price jump, which coincided with a surge in Twitter activity. The sentiment on social media was mixed, with many users expressing excitement over Nvidia's strong earnings report and its advancements in gaming and AI technologies. This generated a wave of optimism among investors. At the same time, some users voiced concerns about market volatility and potential overvaluation, reflecting a degree of skepticism."*

To quantify this sentiment, the tweet content was processed by a specialized transformer model named FitTwitterRoberta. The model outputs the following sentiment probabilities: positive: 0.9345, neutral: 0.0608, and negative: 0.0047. This resulted in a strong positive sentiment signal for this data point, which served as a valuable input for the main forecasting model.

**Model tracking.** To ensure reproducibility and efficient experiment management, we have implemented an MLflow Tracking Server to log and monitor all modeling experiments. The setup captures key details, including model configurations, training and validation losses, evaluation metrics, and other relevant parameters throughout the experimentation process.

## Experiments

We compared several model architectures across different prediction horizons of 1, 3, and 5 days. The models included ARIMA and SARIMA as baseline statistical approaches to begin with, then Multiple Linear Regression (MLR) as a linear baseline suitable for structured inputs, SNN as the deep learning baseline, and BNNs, which are particularly promising for modeling uncertainty. These comparisons were conducted on five semiconductor stock assets.

The experimental setup involved seven types of processed data and a total of 87 engineered features. Each of the five model architectures was evaluated with different configurations across three types of predictive tasks.

Within the BNN framework specifically, we tested variations in distributional assumptions (Normal vs. Student's-t) as well as output configurations, comparing models that estimate only the meaning with those that estimate both the mean and scale parameters. In addition to these core variations, we explored different hidden layer complexities, allowing us to assess the effect of model capacity on predictive performance and uncertainty quantification. We also experimented with different prior distributions over weights to evaluate their influence on posterior estimates. Furthermore, we tested multiple activation functions — including ReLU and TanH.

To evaluate the contribution of various data sources, features were grouped and tested incrementally. The experiments began with a baseline feature set, then progressively incorporated macroeconomic indicators, and finally included additional external signals, allowing us to assess the marginal impact of each feature group on model performance.

## Results

The results show that linear models without non-linear transformations and using all features (Baseline & Macro & Extra) perform the best, especially for short-term 1-day forecasts, the lowest test MSE is 2.46e-05. In contrast, adding non-linear transformations leads to huge overfitting, with test MSEs exploding.

**Table 1.** *MLR Results with an emphasis on insightful findings*

| MLR_INTC model configuration | Features | MSE |
|---|---|---|
| non-linear transformation: False, target: 1 day | Baseline + Macro + Extra | 2.46e-05 |
| non-linear transformation: False, target: 3 days | Baseline + Macro + Extra | 9.94e-05 |
| non-linear transformation: False, target: 1 day | Baseline | 1.15e-04 |
| non-linear transformation: True, target: 1 day | Baseline + Macro + Extra | 2998.1 |

The simple neural network results show consistent performance across different feature sets and horizons, with the best test MSE of 0.000167 observed for the 3-day forecast using only baseline features. Interestingly, adding macro and extra features slightly increased test MSE.

**Table 2.** *SNN Results with an emphasis on insightful findings*

| SNN_INTC model configuration | Features | MSE |
|---|---|---|
| activation layer: TanH, hidden neurons: 1, target: 3 days | Baseline | 0.000167 |
| activation layer: TanH, hidden neurons: 1, target: 3 days | Baseline + Macro + Extra | 0.000168 |
| activation layer: TanH, hidden neurons: 1, target: 5 days | Baseline + Macro + Extra | 0.000169 |
| activation layer: TanH, hidden neurons: 1, target: 1 day | Baseline + Macro + Extra | 0.000179 |

Focusing on BNN experiments, we have extracted several key insights. Firstly, there is no consistent evidence that the Student's-t distribution, often expected to handle financial return outliers more robustly, outperforms the Normal distribution across different configurations. While the Student's-t models demonstrated higher robustness in some setups, they also showed greater variance in performance, particularly when both location and scale parameters are estimated.

The best-performing models were based on the Normal distribution estimating only the location parameter. Notably, the top result MSE of 0.00022 was achieved with a Normal(loc) setup using the TanH activation function, a minimal architecture (1 hidden neuron, 3-day prediction window), and the full feature set including macroeconomic and extra variables.

**Table 3.** *BNN Results with an emphasis on insightful findings*

| BNN_INTC model configuration | Features | MSE |
|---|---|---|
| Normal(loc,scale), activation layer: ReLU, hidden neurons: 9, target: 3 days | Baseline | 0.0008 |
| Normal(loc,scale), activation layer: TanH, hidden neurons: 9, target: 1 day | Baseline | 0.00090 |
| Normal (loc, scale = const), activation layer: ReLU, hidden neurons: 1, target: 3 days | Baseline + Macro + Extra | 0.00023 |
| Normal(loc,scale=const), activation layer: TanH, hidden neurons: 1, target: 3 days | Baseline + Macro + Extra | 0.00022 |
| Student's-t(df=4, loc, scale), activation layer: ReLU, hidden neurons: 9, target: 5 days | Baseline + Macro + Extra | 0.00145 |
| Student's-t(df=4,loc, scale), activation layer: TanH, hidden neurons: 9, target: 1 day | Baseline + Macro + Extra | 0.00121 |
| Student's-t(df=4, loc, scale = const), activation layer: ReLU, hidden neurons: 1, target: 5 days | Baseline + Macro + Extra | 0.00029 |
| Student's-t(df=4, loc, scale = const), activation layer: TanH, hidden neurons: 1, target: 3 days | Baseline + Macro + Extra | 0.0036 |

Regarding activation functions, TanH showed slightly better performance in several configurations compared to ReLU, especially in smaller models. However, the advantage was not universal. Lastly, increasing hidden neuron complexity did not guarantee improved performance. Smaller models often achieved better results, emphasizing the importance of architectural simplicity in combination with rich features for financial time series forecasting.

The plot below corresponds to a BNN where both the location (mean) and scale (variance) parameters are estimated. As a result, the model not only forecasts the expected return but also adjusts its level of uncertainty dynamically, depending on the volatility of the input data. This behavior is particularly valuable in financial contexts where risk varies across time.

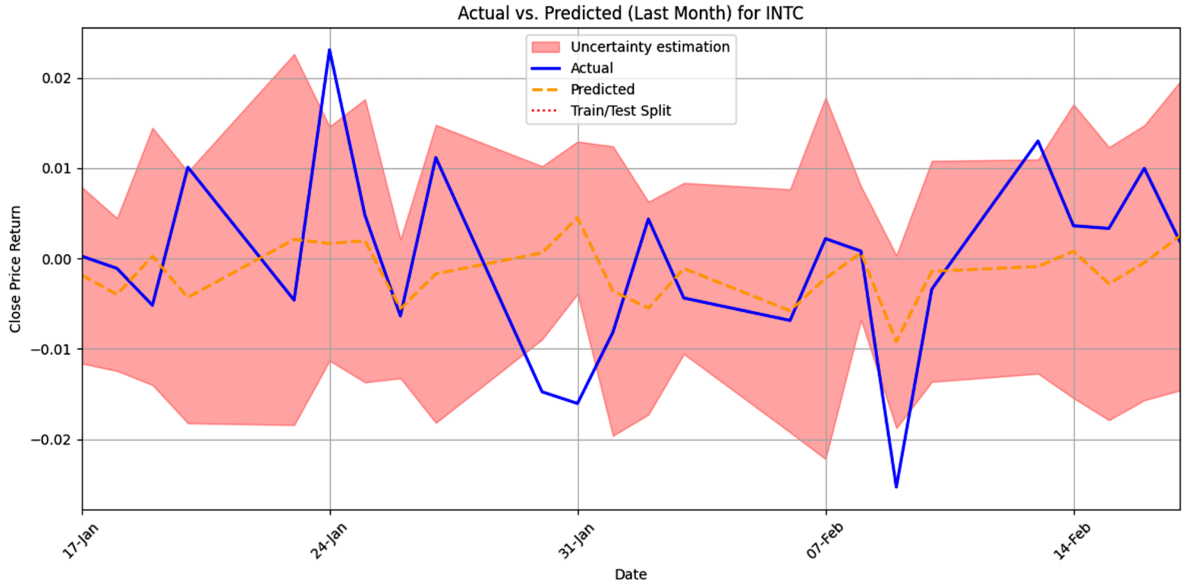


**Figure 2.** Ground Truth vs. Predicted Return and Uncertainty Estimation with MLFlow

## Conclusion

While the MLR model achieved the lowest MSE of 2.46e-05, this result holds limited value when compared with other models in the ablation study, as the model itself lacks the capacity to capture the nonlinear dynamics present in real-world financial data. Therefore, despite its slightly higher MSE of 0.00022, the BNN is more reliable for real-world financial forecasting tasks, validating the first assumption that accounting for uncertainty offers more robust predictive performance. Moreover, both SNN and BNN returned their best results when all three feature sets — baseline, macroeconomic, and extra features — were included with a target window of 3 days. This clearly supports the second assumption, confirming that the feature engineering pipeline we developed significantly enhances model performance.

**Table 4.** *Best results over all architectures*

| Model configuration | Features | MSE |
|---|---|---|
| Linear regression, scaled: True, nonlinear: False, target: 1 day | Baseline + Macro + Extra | 2.46e-05 |
| Simple Neural Network: scaled: True, target: 3 days | Baseline + Macro + Extra | 0.000168 |
| Bayesian Neural Network: scaled: True, Normal (loc, scale=const), TanH, target: 3 days | Baseline + Macro + Extra | 0.00022 |

### *Список літератури*

1. Barbieri F. Twitter-roBERTa-base for Sentiment Analysis [Electronic resource] / F. Barbieri, J. Camacho-Collados, L. Espinosa Anke, L. Neves // HuggingFace.co. — Mode of access: https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment (date of access: 25.05.2025).
2. Borch: Tutorials and contribution guides [Electronic source] // borch.readthedocs.io. — Mode of access: https://borch.readthedocs.io/en/latest/index.html (date of access: 25.05.2025).
3. Federal Reserve Bank of St. Louis. Federal Reserve Bank of St. Louis [Electronic resource] / Federal Reserve Bank of St. Louis // StLouisFed.org. — Mode of access: https://www.stlouisfed.org/ (date of access: 25.05.2025).
4. Hauzenberger N. Enhanced Bayesian Neural Networks for Macroeconomics and Finance [Electronic resource] / N. Hauzenberger et al. // arXiv.org. — Mode of access: https://arxiv.org/abs/2211.04752v2 (date of access: 25.05.2025).

5. Kirtac K. Sentiment Trading with Large Language Models [Electronic resource] / K. Kirtac, G. Germano // arXiv.org. — Mode of access: https://arxiv.org/abs/2412.19245 (date of access: 25.05.2025).
6. Lewis B. A Deep Universal Probabilistic Programming Language [Electronic resource] / L. Belcher, J. Gudmundsson, M. Green Borch // arXiv.org. — Mode of access: https://arxiv.org/abs/2209.06168 (date of access: 25.05.2025).
7. MLflow: An Open Source Platform for the Machine Learning Lifecycle [Electronic source] // MLflow.org. — Mode of access: https://mlflow.org/docs/latest/index.html (date of access: 25.05.2025).
8. OpenAI. Assistants API Overview [Electronic resource] / OpenAI // OpenAI Platform. — Mode of access: https://platform.openai.com/docs/assistants/overview (date of access: 25.05.2025).

***References***

Barbieri F., Camacho-Collados, J., Espinosa Anke, L., & Neves, L. (2025). *Twitter-roBERTa-base for sentiment analysis*. Hugging Face. Retrieved May 25, 2025. https://huggingface.co/cardiffnlp/twitter-roberta-base-sentiment.

Borch. (2025). *Tutorials and contribution guides*. Borch Documentation. Retrieved May 25, 2025. https://borch.readthedocs.io/en/latest/index.html.

Federal Reserve Bank of St. Louis. (2025). *Federal Reserve Bank of St. Louis website*. Retrieved May 25, 2025. https://www.stlouisfed.org/.

Hauzenberger, N., Huber, F., Klieber, K., & Marcellione, M. (2022). *Enhanced Bayesian neural networks for macroeconomics and finance* [Preprint]. arXiv. https://arxiv.org/abs/2211.04752v2.

Kirtac, K., & Germano, G. (2024). *Sentiment trading with large language models* [Preprint]. arXiv. https://arxiv.org/abs/2412.19245.

Lewis, B., Gudmundsson, J., & Green Borch, M. (2022). *A deep universal probabilistic programming language* [Preprint]. arXiv. https://arxiv.org/abs/2209.06168.

MLflow. (2025). *An open source platform for the machine learning lifecycle*. Retrieved May 25, 2025. https://mlflow.org/docs/latest/index.html.

OpenAI. (2025). *Assistants API overview*. OpenAI Platform. Retrieved May 25, 2025. https://platform.openai.com/docs/assistants/overview.

*М. Є. Войтішин, Д. О. Кузьменко*

## РОЗРОБКА ГІБРИДНОЇ МОДЕЛІ ШТУЧНОГО ІНТЕЛЕКТУ ДЛЯ ПРОГНОЗУВАННЯ ФІНАНСОВИХ РИНКІВ

*У статті досліджено можливості гібридного підходу до прогнозування фінансових ринків із застосуванням методів штучного інтелекту. Основну увагу приділено аналізу впливу різних архітектур моделей та наборів ознак на якість прогнозування часових рядів у фінансовому середовищі з високою невизначеністю. Запропоновано поєднання традиційних статистичних моделей, простих нейронних мереж та баєсівських нейронних мереж для моделювання як прогнозного значення, так і масштабної невизначеності. Особливу увагу приділено інженерії ознак, зокрема інтеграції макроекономічних індикаторів, технічних показників, а також поведінкових сигналів на основі аналізу настроїв з Twitter. Результати експериментів показують, що хоча лінійні моделі досягають найменшої середньоквадратичної помилки, саме баєсівські нейронні мережі забезпечують надійніші прогнози завдяки врахуванню невизначеності. Аналіз підтверджує ефективність нашої інженерії ознак та демонструє потенціал поєднання кількісних і якісних даних у фінансовому прогнозуванні.*

**Ключові слова**: прогнозування часових рядів, баєсівські нейронні мережі, фінансовий аналіз, невизначеність, пошук аномалій, аналіз настроїв.

*Матеріал надійшов 31.05.2025*