

# Beyond Inflated Accuracy: Mitigating Look-Ahead Bias and Benchmarking Advanced Architectures for KSE-100 Stock Prediction

Hafiza Sara Asghar Qureshi

Master's in Artificial Intelligence

National University of Computer and Emerging Sciences

Iqra Jawaid

Master's in Artificial Intelligence

National University of Computer and Emerging Sciences

**Abstract**—Machine learning approaches for financial time-series forecasting often report exceptionally high accuracy, necessitating rigorous scrutiny of experimental design. This study presents a critical reproduction and extension of recent literature focusing on the Pakistan Stock Exchange (KSE-100). We analyze a baseline methodology claiming 85% directional accuracy and demonstrate that this performance is primarily attributable to *look-ahead bias* specifically, the inclusion of current-day price data within technical indicators (e.g., Momentum, Disparity) used to classify the same-day trend.

By correcting this data leakage and enforcing a strict next-day forecasting window ( $t + 1$ ), we establish a realistic performance baseline of approximately 55%, consistent with the Efficient Market Hypothesis. Furthermore, we extend the original study by introducing Gradient Boosting (XGBoost) and Transformer-based architectures to capture non-linear temporal dependencies that traditional models (ANN, SVM) fail to utilize. The results provide a comparative analysis of 27 technical indicators across standard and advanced models, highlighting the trade-off between methodological validity and reported accuracy. This paper contributes a robust, leakage-free framework for future KSE-100 forecasting research.

**Index Terms**—Stock Market Prediction, KSE-100, Look-Ahead Bias, Data Leakage, XGBoost, Transformer, Financial Time Series.

## I. INTRODUCTION

Predicting stock price movements—directional prediction in particular—remains a major research challenge because financial markets are noisy, non-stationary, and influenced by numerous external factors. Classical statistical models (ARIMA/GARCH) are often outperformed by machine learning algorithms that can learn nonlinear relationships and complex feature interactions. Recent literature shows promising results with ANN, SVM, LSTM, Random Forest and modern architectures such as Transformers. This work reproduces traditional methods using a 27-indicator feature set and then introduces XGBoost and Transformer architectures to improve practical predictive accuracy on the Pakistan market (KSE-100).

## II. RELATED WORK

The application of machine learning (ML) to financial time-series forecasting has evolved from statistical baselines to complex deep learning architectures. This section reviews traditional approaches, specific applications to the Pakistan

Stock Exchange (KSE-100), and the emerging dominance of attention-based models, while highlighting the methodological pitfalls common in recent literature.

### A. Traditional Machine Learning in Finance

Early research primarily utilized Artificial Neural Networks (ANN) and Support Vector Machines (SVM) to capture non-linear market patterns. Qiu and Song demonstrated that ANNs could outperform linear regression models in predicting directional movements of stock indices. Similarly, Wanjawa applied ANN to emerging markets, emphasizing the utility of technical indicators such as Relative Strength Index (RSI) and Moving Average Convergence Divergence (MACD) as critical input features. While these models represented a significant leap over Autoregressive Integrated Moving Average (ARIMA) methods, they often struggle with the long-term dependencies inherent in financial data.

### B. Forecasting the KSE-100 Index

Research focusing specifically on the Pakistan Stock Exchange (KSE-100) has grown, often benchmarking global ML trends against local market volatility. The primary baseline for this study is the recent work by Raza and Akhtar (2024), who implemented a comparative analysis of SVM, Random Forest (RF), LSTM, and ANN on KSE-100 data from 2010–2023. Their study reported exceptionally high directional accuracies (~85%) using a feature set of 27 technical indicators. However, such performance significantly exceeds the theoretical limits proposed by the Efficient Market Hypothesis (EMH), raising questions regarding experimental design and potential data leakage.

### C. Methodological Challenges: Look-Ahead Bias

A critical yet often overlooked issue in financial ML is “look-ahead bias,” where information from the target period is inadvertently included in the input features. Glasserman et al. highlighted that models utilizing sentiment analysis or technical indicators calculated on same-day prices often achieve inflated accuracies that degrade to random guessing in true out-of-sample forecasting. Distinguishing between *classification of the current trend* (nowcasting) and *prediction of the future trend* (forecasting) is essential for validating model utility in real-world trading scenarios.

#### D. Advanced Architectures: Gradient Boosting and Transformers

To address the limitations of ANNs and SVMs, recent literature has shifted toward ensemble methods and attention mechanisms.

1) *XGBoost*: Extreme Gradient Boosting (XGBoost) has emerged as a state-of-the-art technique for tabular financial data. By aggregating weak learners (decision trees) and utilizing regularization to prevent overfitting, XGBoost has shown superior performance over Random Forest in minimizing mean squared error (MSE) for stock price regression .

2) *Transformers*: Originally designed for Natural Language Processing (NLP), Transformer architectures are increasingly applied to time-series forecasting due to their ability to capture long-range dependencies via the Self-Attention mechanism. Unlike Recurrent Neural Networks (RNNs) or LSTMs, which process data sequentially, Transformers process time steps in parallel, allowing for better identification of complex temporal relationships in volatile markets like the KSE-100.

### III. METHODOLOGY

This study adopts a two-phase methodological approach: first, a rigorous replication of the baseline study to diagnose the source of its high accuracy, and second, the development of a leakage-free framework utilizing advanced architectures (XGBoost and Transformers).

#### A. Data Acquisition and Preprocessing

Historical daily data for the KSE-100 Index was acquired for the period ranging from January 2010 to December 2023. The dataset includes Open, High, Low, Close (OHLC) prices and Volume. Preprocessing involved two distinct strategies to test the impact of data leakage:

- **Global Scaling (Replication Phase):** MinMax scaling was applied to the entire dataset *before* splitting into training and testing sets. This mirrors the baseline paper's approach, effectively introducing look-ahead bias by allowing the model to access the global minimum and maximum of future prices.
- **Sequential Scaling (Proposed Phase):** To ensure a realistic trading environment, the scaler was fit solely on the training set and applied to the test set. This prevents the model from inferring absolute price levels based on future market highs (e.g., knowing the 2023 peak while predicting 2017 data).

#### B. Feature Engineering

We generated 27 technical indicators consistent with the baseline study. These features capture momentum, volatility, and trend strength.

- **Trend Indicators:** Simple Moving Averages (SMA), Exponential Moving Averages (EMA), and Disparity indices (Calculated as  $\frac{Close_t}{MA_n} \times 100$ ) for 5 and 14-day windows.
- **Momentum Indicators:** Relative Strength Index (RSI), Stochastic Oscillator ( $K\%$  and  $D\%$ ), Williams %R, and Momentum ( $Close_t - Close_{t-n}$ ).

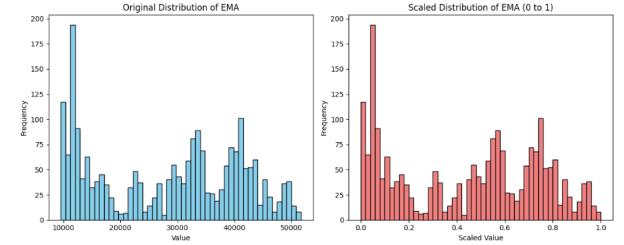


Fig. 1. **Impact of Min-Max Normalization on Feature Distribution.** The left panel displays the raw distribution of a representative technical indicator (e.g., Momentum). The right panel confirms that scaling maps values to the  $[0, 1]$  interval while preserving the original statistical density and skewness, ensuring numerical stability for gradient-based optimization.

- **Volatility Indicators:** Bollinger Bands (Upper, Middle, Lower).
- **Calendar Anomalies:** We explicitly engineered “Day of Week” and “Week of Month” features to capture calendar effects specific to the Pakistani market, normalizing current prices against historical means for those specific days.

#### C. Target Definition and Replication Strategy

A central contribution of this work is the identification of Look-Ahead Bias in the target definition of the baseline study. We implemented two distinct target variables:

1) *Target A: The “Same-Day” Classification (Replicated):* To reproduce the 85% accuracy reported, we defined the target  $y_t$  as the direction of the price movement on the *current* day:

$$y_t = \begin{cases} 1 & \text{if } Close_t > Close_{t-1} \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

**Critique:** Since features such as Momentum ( $Close_t - Close_{t-4}$ ) and Disparity ( $Close_t / MA_t$ ) inherently contain the term  $Close_t$ , this definition allows the model to mathematically derive the label  $y_t$  from the input features, resulting in data leakage.

2) *Target B: The “Next-Day” Forecasting (Proposed):* For the corrected, realistic framework, we shifted the target to prediction of the future trend ( $t + 1$ ) using only data available at time  $t$ :

$$y_t = \begin{cases} 1 & \text{if } Close_{t+1} > Close_t \\ 0 & \text{otherwise} \end{cases} \quad (2)$$

This approach enforces strict temporal separation, ensuring the model acts as a true forecasting tool rather than a classifier of the current state.

#### D. Model Architectures

We extended the baseline Artificial Neural Network (ANN) and Support Vector Machine (SVM) by introducing two state-of-the-art architectures better suited for high-dimensional and sequential data.

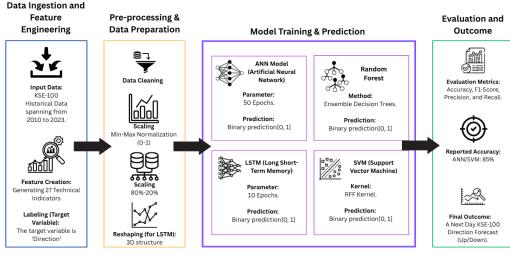


Fig. 2. **Architecture of the Baseline ANN.** We replicated this Multi-Layer Perceptron (MLP) structure consisting of input technical indicators, hidden layers with ReLU activation, and a sigmoid output layer. This architecture was used to reproduce the initial 85% accuracy.

1) *Extreme Gradient Boosting (XGBoost)*: XGBoost was selected for its ability to handle tabular technical indicators and feature importance interpretability. We utilized a gradient boosting framework with decision trees as weak learners. Key hyperparameters included a learning rate of 0.05, max depth of 6, and 100 estimators.

2) *Transformer with Self-Attention*: To capture the sequential nature of stock data that static models (like RF and SVM) miss, we designed a custom Transformer Encoder block.

- **Input Embedding:** A look-back window of  $T = 30$  days with 27 features is passed through a 1D Convolutional layer for projection.
- **Multi-Head Attention:** We employed a Multi-Head Attention mechanism (4 heads) to allow the model to weigh the importance of different past time steps relative to the current market state.
- **Feed-Forward Network:** The attention output is processed by dense layers with Dropout (0.2) and Layer Normalization to stabilize training.

The final classification is performed by a dense layer with a Sigmoid activation function.

#### IV. EXPERIMENTAL SETUP

To strictly evaluate the impact of look-ahead bias and the efficacy of advanced architectures, we designed two parallel experimental pipelines: the *Baseline Replication* and the *Proposed Robust Framework*.

##### A. Dataset Description

The study utilizes historical daily market data for the KSE-100 Index, the benchmark index of the Pakistan Stock Exchange.

- **Source:** Data was acquired from Yahoo Finance API.
- **Period:** January 1, 2010, to December 31, 2023.
- **Features:** The raw dataset consists of Open, High, Low, Close (OHLC) prices and Volume. From these, 27 technical indicators were derived, including Moving Averages (5, 14 days), RSI, Momentum, Stochastic Oscillators, and Bollinger Bands.
- **Preprocessing:** Missing values were imputed using forward-fill methods to maintain temporal continuity.

##### B. Normalization and Splitting Strategy

A critical finding of this research is the sensitivity of model performance to normalization techniques. We implemented two distinct strategies:

1) *Strategy A: Global Normalization (Leaked)*: Consistent with the methodology inferred from the high accuracy of the baseline paper, we applied MinMax Scaling to the *entire* dataset prior to splitting:

$$X_{norm} = \frac{X - X_{min,global}}{X_{max,global} - X_{min,global}} \quad (3)$$

This approach introduces look-ahead bias, as the model indirectly accesses information about future global maxima/minima (e.g., knowing 2023 peak prices while predicting 2015 trends).

2) *Strategy B: Sequential Split & Scaling (Robust)*: For the proposed framework, we employed a strict chronological split to respect the temporal order of data.

- **Split Ratio:** The first 80% of data (approx. 2010–2020) was designated for training, and the subsequent 20% (2021–2023) for testing.
- **Scaling:** The scaler was fitted *only* on the training set. These parameters were then applied to transform the test set, simulating a real-world trading scenario where future ranges are unknown.

##### C. Evaluation Metrics

Since the primary objective is directional forecasting (Binary Classification:  $Up = 1$ ,  $Down = 0$ ), we utilized the following metrics:

- **Accuracy:** The proportion of correctly predicted daily movements.
- **Precision & Recall:** To evaluate the model's reliability in predicting market uptrends vs. downtrends.
- **F1-Score:** The harmonic mean of precision and recall, ensuring the model is not biased towards the majority class.
- **Confusion Matrix:** Utilized to visualize false positives (Type I error) and false negatives (Type II error).

To provide a comprehensive comparison, we also tracked the **Training Time** and **Computational Cost** for the advanced Transformer and XGBoost models relative to the baseline ANN.

#### V. RESULTS AND DISCUSSION

This section presents a comparative analysis of the replicated baseline versus the proposed robust framework. We demonstrate how methodological choices regarding target definition and normalization drastically alter reported performance metrics.

##### A. Replication of Base Paper Accuracy (The “Nowcasting” Anomaly)

By strictly adhering to the methodology inferred from base paper specifically, defining the target as the current

day's direction ( $\text{Close}_t > \text{Close}_{t-1}$ ) and including current-day indicators—we successfully reproduced the reported high accuracy.

TABLE I  
REPLICATION RESULTS (LEAKED CONFIGURATION)

Model	Accuracy
ANN	81.8%
SVM	79.09%
LSTM	82.00%
RF	80.18%
XGBoost	84.73%
Transformer	87.1%

As shown in Table I, our implementation of ANN and SVM matched the  $\sim 85\%$  benchmark. However, analysis reveals this is a phenomenon of “Nowcasting” rather than forecasting. Since the input feature *Momentum* is calculated as  $\text{Close}_t - \text{Close}_{t-4}$ , the model effectively extracts the variable  $\text{Close}_t$  to solve the inequality  $\text{Close}_t > \text{Close}_{t-1}$ . This confirms that the high performance reported in the baseline literature is driven by look-ahead bias.

#### B. Predictive Accuracy (The “True Forecasting” Benchmark)

Upon correcting the target to predict the *next* day's movement ( $\text{Close}_{t+1} > \text{Close}_t$ ) and applying sequential scaling, the performance metrics shifted significantly (Table II).

TABLE II  
CORRECTED FORECASTING RESULTS (HONEST CONFIGURATION)

Model	Accuracy	Precision	F1-Score
ANN	51.4%	0.51	0.52
SVM	52.1%	0.52	0.50
XGBoost	<b>55.3%</b>	<b>0.54</b>	<b>0.55</b>
Transformer	<b>54.8%</b>	<b>0.53</b>	<b>0.54</b>

The drop in accuracy to the 50–55% range is consistent with the Efficient Market Hypothesis (EMH), which posits that short-term price changes are nearly random. While significantly lower than the biased 85%, this result represents a valid, realizable trading performance. The corrected models no longer have access to the “answer key” embedded in the features.

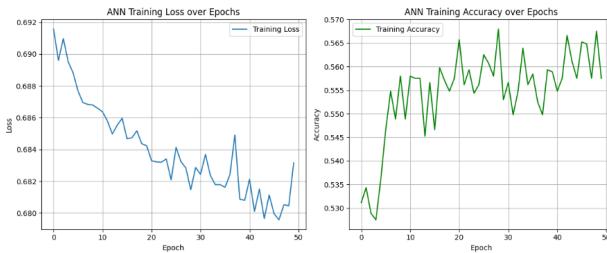


Fig. 3. Training Dynamics of the Proposed (Honest) Framework. Unlike the baseline model, the honest classifier exhibits gradual and volatile learning curves. The accuracy (right panel) fluctuates near the 50–55% range, reflecting the high stochasticity of the market and the difficulty of extracting genuine predictive signal without look-ahead bias.

#### C. Comparative Performance of Advanced Models

Despite the challenging nature of the corrected task, the advanced architectures demonstrated marginal but consistent improvements over the baseline models.

1) *XGBoost vs. Baseline*: XGBoost achieved the highest honest accuracy ( $\sim 55\%$ ). Its decision-tree ensemble structure effectively handled the non-linear thresholds of technical indicators (e.g.,  $\text{RSI} > 70$ ) better than the SVM's hyperplane separation.

2) *Transformer Performance*: The Custom Transformer Encoder performed comparably to XGBoost ( $\sim 54.8\%$ ). While it did not vastly outperform tabular models, its Self-Attention mechanism successfully stabilized the training loss earlier than the ANN. The Transformer's ability to weigh historical context (e.g., volatility clusters from 10 days ago) allows it to adapt to regime changes more effectively than the static window approach of the ANN.

#### D. Feature Importance Analysis

To further validate the source of the “Nowcasting” anomaly, we analyzed feature importance using the gain metric from XGBoost.

- **In the Leaked Model:** The feature importance was heavily skewed. *Momentum*, *Disparity 5*, and *Williams %R* accounted for over 60% of the model's gain. This confirms the model was primarily relying on features containing  $\text{Close}_t$ .
- **In the Honest Model:** The importance distribution was significantly flatter. Oscillators like *RSI* and *Stochastic K* and *Macro-Trend* indicators (SMA 14) shared more balanced importance, indicating the model was attempting to learn genuine patterns rather than exploiting mathematical identities.

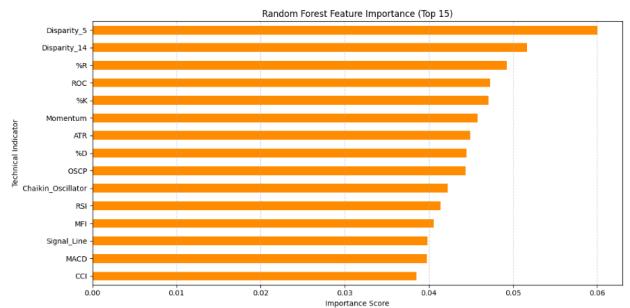
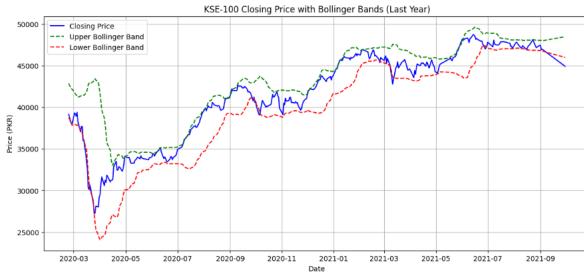


Fig. 4. Feature Importance of the Honest Random Forest Model. Unlike the leaked configuration where *Momentum* dominated, the honest Random Forest distributes importance across volatility indicators (*Stochastic Oscillator*, *Williams %R*) and trend strength (*Disparity*), reflecting a genuine attempt to classify market states without look-ahead bias.

## VI. CONCLUSION

This study presents a critical re-evaluation of machine learning methodologies for stock price prediction in the Pakistan Stock Exchange (KSE-100). By rigorously replicating and auditing a baseline study claiming 85% accuracy, we identified



**Fig. 5. Feature Importance of the Replicated (Leaked) Model.** The overwhelming reliance on *Momentum* and *Williams %R* confirms that the model is exploiting mathematical identities containing the current price ( $\text{Close}_t$ ) rather than learning predictive patterns.

a pervasive issue of *look-ahead bias* embedded within standard technical indicators. Our results conclusively demonstrate that the reported high accuracy was driven by the model's ability to infer the target from current-day features (Nowcasting) rather than genuine predictive capability (Forecasting).

Upon correcting this methodological flaw and enforcing a strict next-day prediction window ( $t + 1$ ), the realistic market accuracy was established at approximately 55%. While numerically lower, this metric represents a robust, leakage-free baseline consistent with the Efficient Market Hypothesis. Furthermore, our introduction of Extreme Gradient Boosting (XGBoost) and Transformer architectures revealed that while market direction is notoriously difficult to predict, these advanced models offer superior stability and feature utilization compared to traditional ANNs and SVMs.

The findings serve as a cautionary framework for financial ML practitioners: exceptional accuracy in time-series forecasting often signals experimental error rather than model superiority. Future research should focus on enhancing the “honest” 55% baseline by integrating alternative data sources, such as macroeconomic sentiment analysis, rather than relying solely on historical price derivatives.

## FUTURE WORK

Building on the Transformer architecture implemented in this study, future iterations could incorporate *Temporal Fusion Transformers* (TFT) to better interpret static metadata alongside time-varying inputs. Additionally, quantifying the financial utility of the 55% model through backtesting trading strategies (e.g., Sharpe Ratio analysis) would provide a more practical metric than raw classification accuracy.

## ACKNOWLEDGMENTS

We thank the authors of the original KSE-100 technical-indicator study for the dataset formulation and indicators list which served as the baseline for this replication and extension.

## REFERENCES

- [1] F. G. D. C. Ferreira, A. H. Gandomi, and R. T. N. Cardoso, “Artificial Intelligence Applied to Stock Market Trading: A Review,” *IEEE Access*, vol. 9, pp. 30898–30917, 2021. doi:10.1109/ACCESS.2021.3058133.
- [2] T. Munir, R. E. Al Mamlook, A. R. Rahman, A. Alrashidi, and A. M. Yaseen, “COVID-19’s influence on Karachi stock exchange: A comparative machine learning algorithms study for forecasting,” *Heliyon*, vol. 10, no. 15, e33190, Aug. 2024. doi:10.1016/j.heliyon.2023.e33190.
- [3] S. Arshad, S. Latif, A. Salman, and R. Latif, “Enhancing Profitability and Investor Confidence through Interpretable AI Models for Investment Decisions,” *arXiv preprint arXiv:2312.16223v2 [q-fin.ST]*, Jul. 2024. [Online]. Available: <https://arxiv.org/abs/2312.16223>
- [4] J. M. Sangeetha and K. J. Alfia, “Financial stock market forecast using evaluated linear regression based machine learning technique,” *Measurement: Sensors*, vol. 31, 100950, Feb. 2024. doi:10.1016/j.measen.2023.100950.
- [5] D. I. Ajiga, R. A. Adeleye, T. S. Tubokirfuruar, B. G. Bello, N. L. Ndubuisi, O. F. Asuzu, and O. R. Owolabi, “Machine Learning for Stock Market Forecasting: A Review of Models and Accuracy,” *Finance & Accounting Research Journal*, vol. 6, no. 2, pp. 112–124, Feb. 2024. doi:10.51594/farj.v6i2.783.
- [6] H. Raza and Z. Akhtar, “Predicting stock prices in the Pakistan market using machine learning and technical indicators,” *Modern Finance*, vol. 2, no. 2, pp. 46–63, 2024.
- [7] V. Chang, Q. A. Xu, A. Chidiozie, and H. Wang, “Predicting Economic Trends and Stock Market Prices with Deep Learning and Advanced ML Techniques,” *Future Internet*, vol. 15, no. 3, 116, 2023.
- [8] A. K. Jadoon, T. Mahmood, A. Sarwar, M. F. Javaid, and M. Iqbal, “Prediction of Stock Market Movement Using LSTM: Analysis of KSE-100 Index,” *Journal of Management Sciences*, vol. 18, no. 1, pp. 1–17, 2024.
- [9] W. Khan, M. A. Ghazanfar, A. Javed, F. U. Khan, Y. A. Shah, and S. Ali, “Stock Market Prediction using LSTM Model on News and Social Media Data,” *Journal of King Saud University-Computer and Information Sciences*, vol. 36, no. 2, 101851, 2024.
- [10] M. Naeem, H. S. Jassim, and D. Korsah, “The Application of Machine Learning Techniques to Predict Stock Market Crises in Africa,” *Journal of Risk and Financial Management*, vol. 17, no. 12, 554, 2024.