evaluation of machine learning models for stock prediction

Stock prediction studies indicate Random Forest and LSTM as primary models, achieving 78-86% classification accuracy and strong regression performance, with effectiveness varying based on market conditions and feature engineering.

Abstract

Ten studies on machine learning for stock prediction report that Random Forest models are the most frequently evaluated, appearing in every study, while Long Short-Term Memory (LSTM) networks are common for capturing time-series dependencies. For classification tasks—predicting directional movement—studies report accuracies in the range of 78–86%, with Random Forest classifiers reaching figures such as 84.89% accuracy (Khanna et al., 2022) and 86.24% (Rohit et al., 2025). For regression tasks, LSTM models tend to yield lower errors; one study reported an RMSE of 0.296 and another detailed LSTM performance with an explained variance of 0.9726 (Majumder et al., 2021).

Feature engineering emerges as a consistent element. Studies that incorporate technical indicators, lag features, and moving averages generally observe enhanced performance. In addition, model choice appears context-dependent, as simpler approaches (e.g. linear regression) perform better in less volatile markets while deep learning models excel with larger datasets and complex temporal patterns. Reporting on dataset specifics remains limited; only a few studies detail data sources and time periods, and none mention sample sizes.

Paper search

Using your research question "evaluation of machine learning models for stock prediction", we searched across over 126 million academic papers from the Semantic Scholar corpus. We retrieved the 50 papers most relevant to the query.

Screening

We screened in sources that met these criteria:

- **ML Stock Prediction Focus**: Does the study evaluate machine learning models specifically for stock price or market movement prediction?
- **Real Market Data**: Does the study use real (not simulated or synthetic) market data from any stock exchange?
- **Methodology Description**: Does the study clearly describe both the machine learning methodology and validation approach used?
- **Performance Metrics**: Does the study include quantitative performance metrics (such as RMSE, accuracy, or precision)?
- **Model Comparison**: Does the study compare multiple machine learning models?
- **Empirical Implementation**: Does the study include practical implementation and testing of the proposed methods (not purely theoretical)?
- **Financial Instrument Focus**: Does the study focus on traditional stock markets (not exclusively on cryptocurrency or other financial instruments)?

We considered all screening questions together and made a holistic judgement about whether to screen in each paper.

Data extraction

We asked a large language model to extract each data column below from each paper. We gave the model the extraction instructions shown below for each column.

- **Machine Learning Models Used**:

  List ALL machine learning models used in the study for stock prediction.

- Locate in methods or results section
- Include full model names (e.g., "Random Forest", not just "RF")
- If multiple models were compared, list all of them
- If specific variants of models were used, note those details
- If no specific model was used, write "No machine learning models used"

Example formats:

- Single model: "Random Forest"
- Multiple models: "Random Forest, Logistic Regression, Support Vector Machine, Gradient Descent"

- **Data Source and Characteristics**:

  Describe the stock market data used in the study:

- Specific stock(s) or index/indices analyzed
- Time period of historical data (start and end dates)
- Total number of data points/observations
- Source of data (e.g., Kaggle, financial database)
- Any specific preprocessing or feature engineering performed

If information is incomplete, note "Insufficient details provided"

- **Performance Evaluation Metrics**:

  List ALL performance metrics used to evaluate machine learning models:

- Accuracy percentage
- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Root Mean Squared Error (RMSE)
- F1 Score
- Explained Variance Score
- R-squared Score

Instructions:

- Extract exact numerical values if provided
- If metric was calculated but no value given, note "Metric calculated but value not reported"
- Ensure metrics match those mentioned in the study

- **Best Performing Model**:

  Identify the top-performing machine learning model:

- Name of the model

- Its specific performance metric(s)
- Reason for superior performance (if explained)

If multiple models performed similarly or no clear best model, note ”No definitive best model identified”

- **Key Research Insights**:

  Summarize the primary insights or conclusions of the study:

- Main findings about stock prediction using machine learning
- Limitations identified
- Recommendations for future research
- Any novel techniques or approaches proposed

Aim for a concise 2-3 sentence summary that captures the core contribution of the study.

Results

Characteristics of Included Studies

Study	Study Design	Machine Learning Models Used	Prediction Type	Dataset Characteristics	Full text retrieved
Ampomah et al., 2020	Comparative evaluation of tree-based ensemble models	Random Forest Classifier, Extreme Gradient Boosting (XGBoost) Classifier, Bagging Classifier, Adaptive Boosting (AdaBoost) Classifier, Extra Trees Classifier, Voting Classifier	Direction of stock price movement (classification)	8 stocks/indices (Bank of America, Exxon Mobil, S&P 500 Index, Microsoft, Dow Jones Industrial Average, CarMax, Tata Steel, HCL Technologies); Yahoo Finance Application Programming Interface; 45 predictors; data cleaning, normalization, feature extraction; 70/30 train/test split; we didn't find mention of time period or sample size	Yes
Khanna et al., 2022	Experimental comparison of four machine learning models	Long Short-Term Memory, Extreme Gradient Boosting, Support Vector Machine, Random Forest	Direction of stock/index movement (classification)	3 major stocks/indices; 11 years (January 2007–December 2017); 10 technical indicators; we didn't find mention in the abstract of data source or sample size	No

Study	Study Design	Machine Learning Models Used	Prediction Type	Dataset Characteristics	Full text retrieved
Rohit et al., 2025	Comparative analysis of four machine learning models	Random Forest, Logistic Regression, Support Vector Machine, Gradient Descent	Stock price prediction (regression/classification)	Apple Inc. (AAPL); 10 years; Kaggle; we didn't find mention in the abstract of sample size or preprocessing	No
Jonathan et al., 2025	Evaluation of multiple machine learning models	Linear Regression, Decision Trees, Random Forests, Adaptive Boosting, Extreme Gradient Boosting, K-Nearest Neighbours, Support Vector Regressor, Long Short-Term Memory	Stock price forecasting (regression)	New York Stock Exchange Composite, Swiss Market Index, others; we didn't find mention of time period, sample size, or data source; feature engineering (lag features, moving averages)	Yes

Study	Study Design	Machine Learning Models Used	Prediction Type	Dataset Characteristics	Full text retrieved
Nabipour et al., 2020	Comparative analysis of machine learning and deep learning models	Decision Tree, Random Forest, Adaptive Boosting, Extreme Gradient Boosting, Support Vector Classifier, Naïve Bayes, K-Nearest Neighbors, Logistic Regression, Artificial Neural Network, Recurrent Neural Network, Long Short-Term Memory	Trend prediction (classification/regression)	4 sectors (Tehran Stock Exchange); 10 years; technical indicators as continuous/binary data; we didn't find mention in the abstract of sample size or data source	No
Pushkala et al., 2024	Comparative study of three models	Long Short-Term Memory networks, Random Forest, Linear Regression	Stock price prediction (regression)	Historical price patterns; sentiment and economic indicators; we didn't find mention in the abstract of data source, time period, or sample size	No
Vijh et al., 2020	Application of two models	Artificial Neural Network, Random Forest	Next-day closing price prediction (regression)	5 companies; open/high/low/close data; feature engineering; we didn't find mention in the abstract of data source, time period, or sample size	No

Study	Study Design	Machine Learning Models Used	Prediction Type	Dataset Characteristics	Full text retrieved
Majumder et al., 2021	Application of five regression models	Linear Regression, Random Forest, Support Vector Regression, Vector Autoregression, Long Short-Term Memory	Stock market prediction (regression)	Historical stock price data; we didn't find mention in the abstract of data source, time period, sample size, or preprocessing	No
Senthilnathan and Thangaraj, 2025	Comparative study of several machine learning models	Linear Regression, Decision Trees, Random Forests, Support Vector Machines, Neural Networks (Feed Forward and Long Short-Term Memory)	Stock price prediction (regression)	Historical stock market data; we didn't find mention in the abstract of data source, time period, sample size, or preprocessing	No
Li, 2024	Evaluation of four machine learning algorithms	Decision Tree, Linear Regression, Random Forest, Support Vector Regression	Stock price forecasting (regression)	Apple Inc.; 3 years; we didn't find mention in the abstract of data source, sample size, or preprocessing	No

Summary of Model Use and Dataset Reporting:

- Most frequently evaluated models:
  - Random Forest: 10 studies
  - Long Short-Term Memory: 6 studies
  - Linear Regression and Extreme Gradient Boosting: 4 studies each
  - Support Vector Machine/Support Vector Classifier/Support Vector Regression: 4 studies
  - Other models (Decision Tree, Adaptive Boosting, Logistic Regression, K-Nearest Neighbors, Artificial Neural Network, Bagging, Extra Trees, Voting Classifier, Gradient Descent, Naïve Bayes, Recurrent Neural Network, Vector Autoregression, Feed Forward Neural Networks): 1–3 studies each
- Prediction task focus:

- 2 studies: classification only (direction of movement)
  - 6 studies: regression only (price or value prediction)
  - 2 studies: both classification and regression
- Dataset characteristics reporting:
  - Data source mentioned in 2 studies (Yahoo Finance Application Programming Interface and Kaggle)
  - Time period mentioned in 4 studies (ranging from 3 to 11 years)
  - We didn't find mention of sample size in any study
  - For the remaining studies, we didn't find mention of data source, time period, or sample size
- General observation:
  - Among the included studies, Random Forest and Long Short-Term Memory were the most commonly evaluated models, and most studies focused on regression tasks. Reporting of dataset characteristics was limited, with few studies providing data source or time period, and none mentioning sample size.

Effects

Comparative Model Performance

Study	Model Type	Accuracy Range	Best Performance Metrics	Key Features
Ampomah et al., 2020	Tree-based ensembles	We didn't find mention of accuracy range	Extra Trees: best on accuracy, precision, F1, area under the curve (values not reported)	45 predictors, feature extraction, normalization
Khanna et al., 2022	Long Short-Term Memory, Extreme Gradient Boosting, Support Vector Machine, Random Forest	78–85%	Random Forest: 84.89% accuracy, 89.33% F1	10 technical indicators, continuous data
Rohit et al., 2025	Random Forest, Logistic Regression, Support Vector Machine, Gradient Descent	78–86%	Random Forest: 86.24% accuracy; Logistic Regression: 85.71%; Support Vector Machine: 84.65%	Kaggle data, Apple Inc., focus on hyperparameter tuning
Jonathan et al., 2025	Long Short-Term Memory, Random Forest, Extreme Gradient Boosting, others	We didn't find mention of accuracy range	Long Short-Term Memory: root mean squared error 0.296 (lowest), Decision Tree: root mean squared error 0.350	Lag features, moving averages, time-series focus

Study	Model Type	Accuracy Range	Best Performance Metrics	Key Features
Nabipour et al., 2020	9 machine learning models, Recurrent Neural Network, Long Short-Term Memory	We didn't find mention of accuracy range	Recurrent Neural Network/Long Short-Term Memory: best for continuous and binary data (metrics not reported)	Technical indicators, binary/continuous input
Pushkala et al., 2024	Long Short-Term Memory, Random Forest, Linear Regression	We didn't find mention of accuracy range	Long Short-Term Memory: highest accuracy (value not reported)	Sentiment, economic indicators, feature engineering
Vijh et al., 2020	Artificial Neural Network, Random Forest	We didn't find mention of accuracy range	No single model identified as best; both efficient (low root mean squared error/mean absolute percentage error, values not reported)	Feature engineering, sectoral diversity
Majumder et al., 2021	Long Short-Term Memory, Random Forest, Support Vector Regression, Vector Autoregression, Linear Regression	We didn't find mention of accuracy range	Long Short-Term Memory: explained variance score 0.9726, root mean squared error 0.6220, mean squared error 0.3869, mean absolute error 0.4306, R-squared 0.9716	Multiple regression metrics, historical data
Senthilnathan and Thangaraj, 2025	Linear Regression, Decision Tree, Random Forest, Support Vector Machine, Neural Networks (Feed Forward, Long Short-Term Memory)	We didn't find mention of accuracy range	No single model identified as best; neural networks and Random Forest better for dynamic markets, Linear Regression for low-volume	Model selection by market context
Li, 2024	Decision Tree, Linear Regression, Random Forest, Support Vector Regression	We didn't find mention of accuracy range	Support Vector Linear Regression: best (mean squared error, value not reported)	Apple Inc., 3 years, model selection focus

Model Types Evaluated:

- Tree-based models (Random Forest, Decision Tree, Extreme Gradient Boosting, Extra Trees): 9 studies
- Neural network models (Long Short-Term Memory, Recurrent Neural Network, Artificial Neural Network, Feed Forward Neural Network): 8 studies
- Linear models (Logistic Regression, Linear Regression, Support Vector Regression, Vector Autoregression): 5 studies
- Other models (Support Vector Machine, Gradient Descent, etc.): 5 studies

Accuracy and Performance Reporting:

- Accuracy ranges were mentioned in 2 studies (78–85% and 78–86%).
- In 7 studies, a specific best-performing model was identified (including Extra Trees, Random Forest, Long Short-Term Memory, Support Vector Regression).
- In 2 studies, no single model was identified as best, with performance depending on context or multiple models performing similarly.
- Performance metrics with values were reported in 4 studies (accuracy, F1, root mean squared error, explained variance score, mean squared error, mean absolute error, R-squared).
- Performance metrics were mentioned but without values in 4 studies.
- We didn't find mention of performance metrics in 2 studies.

Key Features Used:

- All 10 studies described some form of feature engineering, selection, or normalization.
- Technical indicators were explicitly mentioned in 2 studies.
- Sentiment features were mentioned in 1 study and economic indicators in 1 study.
- Time-series features (lags, moving averages) were mentioned in 2 studies.
- Data source was specified in 2 studies (Kaggle, Apple Inc.).
- Model selection by market context was a focus in 2 studies.

Performance by Prediction Task

- Classification (directional movement):
  - In the included studies, tree-based ensembles (Random Forest, Extra Trees) and Random Forest classifiers were reported to achieve high accuracy, especially when technical indicators were used as features.
- Regression (price prediction):
  - In the included studies, Long Short-Term Memory models were reported to outperform others in capturing temporal dependencies, as indicated by lower root mean squared error and higher explained variance scores where reported. Support Vector Regression also demonstrated strong performance in some contexts.
- Mixed/Context-dependent:
  - Some studies (such as Senthilnathan and Thangaraj, 2025) emphasized that model performance was context-dependent, with no single model universally optimal.

Feature Impact Analysis

- Technical Indicators:
  - The computation and inclusion of technical indicators (such as moving averages, relative strength index) were associated with improved model performance in the studies that reported on this.

- Feature Engineering:
  - Lag features and moving averages were commonly used to enhance predictive power, particularly in time-series models, as described in several studies.
- Sentiment and Economic Data:
  - Incorporation of sentiment analysis and economic indicators was highlighted as a novel approach in one study, though its quantitative impact was not uniformly assessed across the included studies.

Implementation Considerations

Model Complexity and Resources

- Deep learning models (Long Short-Term Memory, Recurrent Neural Network):
  - In the included studies, these models were reported to require larger datasets and greater computational resources, and were described as offering improved performance in capturing complex temporal patterns.
- Tree-based ensembles (Random Forest, Extra Trees):
  - These models were described as offering a balance between performance and interpretability, with relatively lower computational demands compared to deep learning models.
- Simplicity versus overfitting:
  - Simpler models such as Linear Regression were reported as preferable in low-volume or less volatile markets to avoid overfitting.

Practical Applications

- Model selection:
  - The choice of model was reported to depend on the specific market context, data availability, and prediction task.
- Feature engineering:
  - Effective preprocessing and feature engineering were consistently described as critical for maximizing model performance.
- Generalizability:
  - The limited reporting of dataset characteristics and the use of single stocks or specific markets in several studies restricts the generalizability of findings.

References

Anup Majumder, Md. Mahbubur Rahman, A. Biswas, Md. Sabab Zulfiker, and Sarnali Basak. "Stock Market Prediction: A Time Series Analysis ." *Smart Systems: Innovations in Computing*, 2021.

DANIEL JONATHAN, SOWMYA BURUGUPALLI, MOUNIKA UGGINA, HESHMA KANCHERALA, GOWTHAMI SINDHU PRIYA CHILUKURI, IFEANYI MARTINS NWANEGBO, and NONSO FREDRICK CHIOBI. "STOCK PRICE PREDICTION USING MACHINE LEARNING." *International Journal of Engineering Processing and Safety Research*, 2025.

Ernest Kwame Ampomah, Zhiguang Qin, and G. Nyame. "Evaluation of Tree-Based Ensemble Machine Learning Models in Predicting Stock Price Direction of Movement." *Inf.*, 2020.

K.R Senthilnathan, and Rajasekaran Thangaraj. "Stock Market Price Prediction Using Machine Learning and LSTM Model." *Social Science Research Network*, 2025.

Mehar Vijh, Deeksha Chandola, V. Tikkiwal, and Arun Kumar. “Stock Closing Price Prediction Using Machine Learning Techniques.” *Procedia Computer Science*, 2020.

Mojtaba Nabipour, Pooyan Nayyeri, H. Jabani, S. S., and A. Mosavi. “Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis.” *IEEE Access*, 2020.

Munish Khanna, Mohak Kulshrestha, Law Kumar Singh, Shankar Thawkar, and Kapil Shrivastava. “Performance Evaluation of Machine Learning Algorithms for Stock Price and Stock Index Movement Prediction Using Trend Deterministic Data Prediction.” *International Journal of Applied Metaheuristic Computing*, 2022.

Pushkala B, Sandhya Rao, Satwik Bhat, and Sriya Madhapur. “Stock Market Price Prediction.” *International Conference on Recent Trends in Computing &Amp; Communication Technologies (ICRCCT'2K24)*, 2024.

Tianyao Li. “Forecasting Stock Price Trends with Machine Learning Techniques.” *Applied and Computational Engineering*, 2024.

Y. S. S. Rohit, T. Rushik, and Gayathri Ramasamy. “A Comparative Analysis of Machine Learning Models for Stock Price Prediction Performance Evaluation and Improvement Strategiesve Analysis of Machine Learning Models for Stock Price Prediction.” *Social Science Research Network*, 2025.