Project Tittle

Predictive Analysis of Stock Market Trends Using Machine Learning

Team Members:-

i. Sayantika Chowdhury(E23CSEU2165)

ii. Ishita Shrivastava(E23CSEU2190)

iii. Tanya Mishra(E23CSEU2161)

Literature survey related to the topic:-

1.Related survey done by Ishita:-

Link of the 1st research paper -

https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9165760

SUMMARY-

Domain: Focuses on predicting stock prices in volatile markets, considering various influencing factors.

Datasets: Uses historical stock prices and sentiment data from news articles, processed for polarity scores.

Methodology: Implements a Random Forest algorithm to preprocess data, perform sentiment analysis, and create predictive models, comparing results with logistic regression using multiple regression techniques.

Performance Metrics: Evaluated with variance score, MAE, MSE, and MSLE. Results show the Random Forest model outperforms logistic regression across these metrics.

Link of the 2nd research paper –

https://d1wqtxts1xzle7.cloudfront.net/60241969/IRJET-V6I581720190808-86852-1sjdu6a-libre.pdf?1565324109=&response-content-disposition=inline%3B+filename%3DIRJET_Prediction_of_Stock_Market_using_M.pdf&Expires=1729832480&Signature=L3g6XngUbVsEsAeyZkTZbnnAte6-BT~JRLI9yOHaJZwuGQiN4zFNA-LVyHbsKWG~l1jNjWLLHvJ-VDnGlp~li2iFRXgaZBrsXpqDpxPobbVcJbfislQTSJHoexFJjPJZzrt4b3nGw0OnOtJqvOYH-pd3wBI8FMkiUmyDHpqucZzP0hhGwEHxAfVa3PIvsvqbp~CdwIXszXeintTDyOETl8UEPNt3f~vQp4wj5hTdiRF8ATnFYwjgRLf4A52VSrCDlytEyT3FtsXtO~r3~qrauGWu~oKqzTjTaQra-K0gQg8cDblaQffbzJiOuLG2qUJdIQVx0i82KwZaTQNm~efN3g__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

SUMMARY-

Domain: Stock Market Prediction

Focuses on forecasting stock prices using machine learning to analyze complex financial patterns.

Datasets:

Uses historical price data and technical indicators, including moving averages and volatility measures, to inform models.

Methodology:

Employs machine learning models such as Artificial Neural Networks (ANNs) and Support Vector Machines (SVMs) to recognize trends. Data is preprocessed, and models are compared to determine effectiveness in prediction accuracy.

Performance Metrics:

Evaluates accuracy, Mean Squared Error (MSE), and Return on Investment (ROI) to assess model reliability and profitability.

Link of the 3rd research paper –

https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012098/pdf

SUMMARY-

Domain: Stock Market Prediction

Analyzes stock price forecasting using machine learning methods for financial decision-making.

Datasets:

Employs historical data, technical indicators, and fundamental metrics to capture market patterns.

Methodology:

Uses hybrid techniques, combining Genetic Algorithms and neural networks to improve predictive accuracy. The model preprocesses data and applies advanced algorithms to identify trends and insights.

Performance Metrics:

Assesses model performance with Mean Squared Error (MSE), Mean Absolute Error (MAE), and accuracy metrics to determine prediction reliability and profitability.

2. Related Survey Done By Sayantika:-

Link for 1st research paper-

https://iaeme.com/MasterAdmin/Journal_uploads/IJCA/VOLUME_1_ISSUE_1/IJCA_01_01_001.pdf

SUMMARY-

Domain: The study focuses on stock market prediction, an area characterized by its volatility and the complexity of accurately forecasting stock prices based on various influencing factors.

Datasets: The analysis utilizes historical stock price data from publicly listed companies and sentiment data extracted from news articles, which are processed to derive polarity scores indicating positive or negative impacts on stock prices.

Methodology: The methodology involves implementing a Random Forest algorithm, which includes preprocessing the data, conducting sentiment analysis, and generating predictive models. The performance of the Random Forest model is then compared to that of logistic regression using various regression techniques.

Performance Metrics: Key performance metrics for evaluation include variance score, mean absolute error (MAE), mean squared error (MSE), and mean squared log error (MSLE), with the Random Forest algorithm demonstrating better results across these metrics compared to logistic regression.

Link for 2nd research paper-

https://doi.org/10.1016/j.procs.2022.12.028

SUMMARY-

Domain: This research focuses on stock price prediction using machine learning to address high market volatility and uncertainty.

Datasets: It utilizes historical stock data from 12 major Indian companies collected over seven years, including stock prices and other key trading attributes.

Methodology: Five algorithms—K-Nearest Neighbors, Linear Regression, Support Vector Regression, Decision Tree Regression, and Long Short-Term Memory (LSTM)—were implemented and compared. Data preprocessing included normalization, and the dataset was split into training and testing sets.

Performance Metrics: Evaluation was based on Symmetric Mean Absolute Percentage Error (SMAPE), R-squared (R^2), and Root Mean Square Error (RMSE). LSTM outperformed other models across all metrics, showing higher accuracy and stability.

Link for 3rd research paper-

https://journalofbigdata.springeropen.com/articles/10.1186/s40537-020-00333-6

SUMMARY-

Domain: The study focuses on short-term stock market price trend prediction using deep learning, specifically targeting the volatile and complex nature of financial markets.

Datasets: A custom dataset was constructed from two years of Chinese stock market data using open-source APIs, including stock prices, trading data, financial indicators, and investor behavior.

Methodology: The approach integrates deep learning (LSTM) with feature engineering. Key steps include feature selection via recursive feature elimination (RFE), dimensionality reduction through PCA, and

transformation of raw data into labeled time series for LSTM model training.

Performance Metrics: The model's accuracy was evaluated using metrics such as binary accuracy, precision, recall, and training efficiency. The proposed system, leveraging LSTM and customized feature engineering, demonstrated high accuracy, surpassing traditional machine learning models like SVM and Random Forest.

3. Related survey done by Tanya-

Link for 1st research paper-

https://www.researchgate.net/publication/341482418_A_Survey_on_Stock_Market_Prediction_Using_Machine_Learning_Techniques

SUMMARY-

Domain: Stock Market Prediction

Dataset: The paper does not specify a particular dataset but focuses on general techniques applicable to stock market data, which is typically time-variant and nonlinear. Parameters used include stock prices (open, high, low, close), technical indicators, and fundamental factors.

Methodology: The study reviews several machine learning and time series models, such as:

Holt-Winters Model: Suited for time series with trend and seasonal components.

Artificial Neural Networks (ANN): Uses back-propagation in a multilayer perceptron setup.

Hidden Markov Model (HMM): Used for optimization, especially in trend analysis.

ARIMA Model: Applied for robust short-term forecasting.

Time Series Linear Model (TSLM): Integrates traditional and seasonal trends in a linear framework.

Recurrent Neural Networks (RNN): Leverages feedback loops for predicting stock prices based on recent history.

Performance Metrics: The models' performance is evaluated based on prediction accuracy, ability to handle non-linear data, and effectiveness in short-term versus long-term forecasting. Specific metrics include prediction error rates, accuracy in trend identification, and robustness to variations and overfitting.

This comparative study highlights the strengths and limitations of each model for stock market forecasting **and** proposes combining multiple methods to improve prediction accuracy.

Link for 2nd research paper-

https://scholarworks.lib.csusb.edu/cgi/viewcontent.cgi?article=1435&context=jitim

SUMMARY-

Domain:

This review explores stock market prediction using machine learning (ML) to improve investment decision-making by predicting stock prices, trends, or index values.

Dataset:

The studies use stock index data from global markets like S&P 500, NASDAQ, DAX, Hang Seng, and KOSPI, often spanning years or decades. Common features include daily closing prices, technical indicators, and economic variables. Some datasets also incorporate high-frequency trading data and financial news archives to capture market sentiment.

Methodology:

The paper categorizes ML methods into four main approaches:

Artificial Neural Networks (ANNs): Used primarily for numerical predictions. Some studies integrate deep learning with feature extraction techniques (e.g., PCA) to improve high-frequency trading accuracy.

Support Vector Machines (SVMs): Employed for classification tasks, such as predicting stock movement direction (rise or fall). Hybrid models often include feature selection methods to optimize prediction accuracy.

Genetic Algorithms (GAs) with Other Techniques: GAs are combined with ANNs or SVMs for feature optimization, helping focus on key predictive variables and improving long-term accuracy.

Hybrid and Other AI Techniques: Multiple ML methods (e.g., decision trees, random forests, fuzzy logic) are combined to improve robustness. This includes using reinforcement learning and rule-based expert systems for complex market predictions.

Performance Metrics:

Models are evaluated on prediction accuracy (alignment with actual prices), classification accuracy (correct market movement prediction), and error rates (MSE, MAE). Applied metrics also include risk-

adjusted returns to assess trading profitability in simulated environments. Hybrid models generally outperform single models but are often more complex.

The paper concludes that while hybrid models show promise, practical complexity remains a challenge. Future research should focus on generalizability across different markets and economic cycles, incorporating financial theories to enhance predictive value.

Link for 3rd research paper-

https://sist.sathyabama.ac.in/sist_naac/documents/1.3.4/1822-b.e-cse-batchno-237.pdf

SUMMARY-

Domain: Stock Market Prediction

Dataset: Historical stock data including attributes such as date, high, low, close prices, adjusted values, and additional features like latitude, longitude, hour, and month.

Methodology:

Data Preprocessing: Handling missing values and converting categorical attributes into numerical ones.

Feature Selection: Selection of relevant features (date, price, adjusted close, etc.) for model building.

Model Building: Various machine learning algorithms, including linear regression and ensemble classifiers like Random Forest, are used to predict stock prices.

Performance Metrics: The model's accuracy is evaluated based on its ability to predict market prices accurately, with a focus on minimizing error in predictions.

For dataset and model testing - https://github.com/Sayantika2327/stock-market-prediction-using-ml.git