

# **Stock Market Analysis and Prediction System**

DISSERTATION

Submitted in partial fulfillment of the requirements of the  
Degree : **M.Tech. in Artificial Intelligence and Machine Learning**

By

**NEELANSH GUPTA**  
**2022AC05003**

Under the supervision of

**ABHINAV NAG**  
**(VP of Engineering)**

BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE  
Pilani (Rajasthan) INDIA

(May, 2025)

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**  
**SECOND SEMESTER 2024-25**

**AIMLCZG628T DISSERTATION**

**Dissertation Title** : Stock Market Analysis and Prediction System

**Name of Supervisor** : Abhinav Nag

**Name of Student** : Neelansh Gupta

**ID No. of Student** : 2022AC05003

Courses Relevant for the Project & Corresponding Semester : 1. Machine Learning  
2. Advanced Deep Learning  
3. Deep Neural Network  
4. Statistical Methods

# Abstract

The rapid digitalization of financial markets, coupled with growing data volumes, has fostered tremendous interest in the application of advanced analytics and machine learning to stock market forecasting. This project presents the design and development of a comprehensive, modular stock market analysis and prediction system, focused on the Indian National Stock Exchange (NSE). The core aim is to empower investors and analysts with actionable Buy/Sell/Hold recommendations rooted in historical data and state-of-the-art machine learning techniques.

The system is structured as a robust pipeline that covers all essential phases of the analytical process. Beginning with data acquisition, it automatically downloads and organizes historical price and volume data for major NSE stocks. The data preprocessing component rigorously cleanses and normalizes these datasets, correcting missing or noisy entries to ensure analytical reliability. Feature engineering then transforms the raw data into a rich, informative set of indicators—such as Moving Averages, RSI, MACD, Bollinger Bands, volume-based and volatility measures—thus capturing a broad array of price dynamics, market momentum, and trend signals. Special attention is paid to integrating both statistical features and calendar-based effects specific to the Indian stock market environment.

The core model development phase leverages both classical and modern algorithms, including Random Forest, Gradient Boosting, XGBoost, LSTM, and hybrid CNN-LSTM architectures. A 3-class classification approach is employed, allowing the models to recommend Buy, Hold, or Sell signals based on predicted forward returns. Ensemble and deep learning methods are extensively compared and optimized to enhance predictive accuracy and robustness, while techniques like feature importance scoring guide the selection of the most informative inputs.

To ensure practical viability, the system integrates a custom backtesting and evaluation framework. Each model's out-of-sample performance is rigorously assessed using metrics such as total and annualized returns, Sharpe ratio, volatility, maximum drawdown, and win rate. The recommendation signals are benchmarked against baseline strategies (e.g., buy-and-hold) and subjected to detailed drawdown and risk analysis. Confidence-based filtering is incorporated so that the model defaults to "Hold" when prediction certainty falls below a defined threshold, thereby managing risk in uncertain market conditions.

Visualization modules play a key role in this system, enabling intuitive exploration of raw data, indicator trends, model predictions, and performance comparisons. All results and recommendations are presented through clear summary reports and dynamic plots, supporting transparency and informed decision-making.

The project's modular architecture facilitates easy adaptation to changing market conditions, expanded feature engineering, and the incorporation of additional modeling approaches. Comprehensive documentation ensures that the system can be maintained, extended, or deployed in varied research and real-world investment settings.

In conclusion, the developed system not only demonstrates the potential for machine learning and deep learning techniques in the context of Indian stock markets, but also addresses key challenges such as reliability, interpretability, and workflow automation. It represents a significant step towards data-driven investment analysis and can serve as a solid foundation for further research and advanced financial application development.

**Key Words:**

Stock Market Prediction, Machine Learning, Deep Learning, National Stock Exchange (NSE), Feature Engineering, Financial Forecasting, Buy/Sell/Hold Recommendations, Backtesting, Technical Analysis, Ensemble Learning, LSTM, CNN-LSTM, Indian Stock Market

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**  
**II SEMESTER 24-25**

AIMLCZG628T DISSERTATION

**Dissertation Outline (Abstract)**

**BITS ID No :** 2022AC05003

**Name of Student:** Neelansh Gupta

**Name of Supervisor** : Abhinav Nag

**Designation of Supervisor** : VP of Engineering

**Qualification and Experience** : B.Tech & MBA with 20+ years of experience

**Official E- mail ID of Supervisor** : abhinav.nag@credable.in

**Topic of Dissertation** : Stock Market Analysis and Prediction System



Neelansh Gupta  
(Signature of Student)

Date: May 25, 2025



Abhinav Nag  
(Signature of Supervisor)

Date: May 25, 2025

# Title : Stock Market Analysis and Prediction System

## 1. Broad Area of Work

This project is situated at the intersection of **Financial Data Science**, **Machine Learning**, and **Stock Market Prediction**. It aims to apply advanced computational methods to analyze historical market data and forecast future trends in the Indian stock market, specifically focusing on securities listed on the **National Stock Exchange (NSE)** of India.

The work integrates both traditional financial indicators and cutting-edge deep learning techniques to develop a robust, end-to-end predictive analytics system.

## 2. Objectives

The primary objectives of this project are:

- To develop a complete, modular pipeline for stock market analysis and prediction using historical NSE data.
- To design and implement machine learning and deep learning models (e.g., Random Forest, XGBoost, LSTM, CNN-LSTM) for forecasting stock price movements.
- To engineer meaningful features and technical indicators from raw market data to improve prediction accuracy.
- To evaluate and benchmark multiple predictive models using backtesting and robust performance metrics.
- To provide actionable Buy/Sell/Hold recommendations to support informed investment decisions.
- To develop user-friendly visualizations for data exploration and decision support.
- To automate end-to-end data collection, processing, model training, and recommendation generation to facilitate reproducibility.
- To integrate risk management techniques and confidence measures in the recommendation engine, ensuring reliability under uncertain markets.
- To support comparison of alternative analysis strategies through configurable modules (e.g., indicator choices, model swaps).
- To explore the effectiveness of feature importance and selection techniques in improving model generalization and reducing overfitting.

- To study the impact of various market-specific phenomena (such as calendar effects and sectoral trends) on model predictions.
- To maintain well-documented code and thorough technical documentation supporting ease of maintenance, reproducibility, and extensibility.
- To evaluate and interpret the results in the context of the Indian stock market, identifying both strengths and limitations of the deployed approaches.
- To summarize findings and limitations and suggest potential improvements and future research directions..

### 3. Scope of Work

The scope of this dissertation is to design, develop, and evaluate a comprehensive and modular Stock Market Analysis and Prediction System using machine learning techniques. The primary goal is to enable advanced, data-driven analysis and generate actionable investment recommendations for stocks listed on the National Stock Exchange (NSE) of India. The detailed scope includes:

- **Data Acquisition & Management:**
  - Develop automated modules to fetch and store historical stock market data directly from the NSE in various timeframes (e.g., daily, weekly).
  - Organize raw and processed data efficiently for reuse and scalability within the project structure.
- **Data Preprocessing:**
  - Implement robust data cleaning techniques to handle missing, inconsistent, or noisy data.
  - Normalize and transform data to formats suitable for machine learning, ensuring quality inputs for all downstream analyses.
- **Feature Engineering:**
  - Calculate a broad range of technical indicators such as Moving Averages, RSI, MACD, Bollinger Bands, ADX, and others.
  - Engineer custom features from price, volume, and volatility data to enhance model learning.
  - Incorporate calendar-based features (like day-of-week, month-end effects) and market-specific variables.
- **Model Development:**
  - Develop, train, and tune multiple machine learning and deep learning models, including Random Forest, XGBoost, LightGBM, SVM, LSTM, and CNN-LSTM.

- Implement model selection and hyperparameter optimization strategies to identify the most effective predictors for Indian stock market data.
- **Backtesting and Strategy Evaluation:**
  - Design and implement backtesting modules to evaluate predictive models and trading strategies using historical data.
  - Analyze key performance metrics such as return, volatility, drawdown, Sharpe ratio, and win rate to ensure statistical robustness.
- **Recommendation and Decision Support:**
  - Develop a robust recommendation system to provide clear Buy, Sell, or Hold signals based on predictive model outputs and confidence levels.
  - Integrate filtering mechanisms to manage risk and default to "Hold" under uncertain market conditions.
- **Visualization and Reporting:**
  - Provide rich, interactive visualizations for data exploration (e.g., price history, feature trends, indicator plots).
  - Visualize model performance, strategy results, and recommendations through dashboards and summary reports.
- **System Integration and Usability:**
  - Ensure the entire pipeline is modular, well-documented, and user-friendly, allowing for end-to-end execution and future extensibility.
  - Facilitate easy experimentation and comparison of different strategies, indicators, and models within a unified framework.
- **Documentation:**
  - Produce comprehensive documentation, including user manuals, technical reports, and presentation materials to communicate project outcomes to both technical and non-technical audiences.



## 4. Detailed Plan of Work (16 Weeks)

S.No.	Task/Phase	Duration	Deliverables
1	Project initiation & Literature Review	Week 1	Finalized project title, scope, defined objectives
2	Requirement Analysis & System Design	Week 2	Functional requirement documents, architecture diagram, technology stack listing
3	Data Collection setup	Week 3	Automated data collection scripts, initial data samples
4	Data storage & Preprocessing	Week 4	Structured raw and processed data directories, preprocessing scripts, data quality report
5	Exploratory Data Analysis	Week 5	EDA report, visualizations of trends and statistics
6	Feature Engineering - Technical Indicators	Week 6	Scripts for calculating technical indicators feature set, plots of selected indicators
7	Feature Engineering - Advanced Features	Week 7	Advanced feature scripts, correlation matrix, refined feature selection document
8	Baseline Model Development	Week 8	Baseline model code, performance metrics (accuracy, precision, recall), benchmark results
9	Advanced Machine Learning models	Week 9	Scripts for ensemble models, hyperparameter tuning results, comparison tables/charts
10	Deep Learning Model Design	Week 10	LSTM and CNN-LSTM model code, training/validation results, overfitting analysis
11	Model Evaluation and Selection	Week 11	Evaluation report, backtesting summary, selected model rationale
12	Strategy Backtesting and Validation	Week 12	Backtesting framework, performance metrics (returns, Sharpe, drawdown), Strategy analysis report

13	Recommendation System and Risk Management	Week 13	Recommendation module, Risk management filters, confidence thresholding documentation
14	Visualization and Dashboarding	Week 14	Data visualization modules, dashboard screenshots, interactive charts
15	Documentation, Testing and final Refinement	Week 15	User manual, technical documentation, test cases/results, feedback summary, refined codebase
16	Report Preparation and Presentation	Week 16	Final project report, presentation slides, summary of outcomes and future work

### *Week 1: Project Initiation and Literature Review*

- Finalize project topic and scope.
- Conduct an in-depth literature survey on stock prediction, machine learning techniques, technical analysis, and recent research in financial data science.
- Define high-level objectives and gather key references.

### *Week 2: Requirement Analysis and System Design*

- List functional and technical requirements.
- Define data needs (types, frequency, adequacy) and identify major NSE stocks for study.
- Design overall pipeline architecture—modular components for data, features, modeling, evaluation, and visualization.

### *Week 3: Data Collection Setup*

- Research sources and APIs for reliable NSE data acquisition.
- Build and test an automated data collection script.
- Start collecting historical data for selected stocks.

### *Week 4: Data Storage and Preprocessing*

- Design a structure for storing raw and processed datasets.
- Implement data preprocessing methods: handle missing values, remove outliers, normalize/scale data.
- Document preprocessing steps and initial observations.

### *Week 5: Exploratory Data Analysis (EDA)*

- Perform EDA to understand stock price trends, seasonalities, and anomalies.
- Visualize basic statistics and distributions.
- Summarize market-specific patterns and challenges.

### *Week 6: Feature Engineering - Technical Indicators*

- Implement calculation of key technical indicators: Moving Averages (SMA, EMA), RSI, MACD, Bollinger Bands, ADX, etc.
- Create feature sets using price, volume, volatility, and calendar effects.
- Visualize and validate features' relevance.

### *Week 7: Feature Engineering - Advanced & Custom Features*

- Develop complex features (e.g., lagged returns, rolling metrics, event-based features).
- Conduct correlation analysis to understand relationships between indicators.
- Refine feature selections for predictive modeling.

### *Week 8: Baseline Model Development*

- Build simple baseline models (e.g., logistic regression, SVM, decision trees).
- Set up training and test data splits.
- Evaluate initial model performance and establish reference metrics.

### *Week 9: Advanced Machine Learning Models*

- Implement ensemble learning models: Random Forest, XGBoost, LightGBM.
- Tune hyperparameters using appropriate techniques (e.g., grid search, cross-validation).
- Compare performance against baselines.

### *Week 10: Deep Learning Model Design*

- Develop and train LSTM and CNN-LSTM models for time-series forecasting.
- Address challenges like data formatting, sequence length, and overfitting.
- Summarize benefits/drawbacks of deep learning for stock data.

### *Week 11: Model Evaluation and Selection*

- Evaluate all models using accuracy, precision, recall, F1 score, and confusion matrix.
- Use backtesting to compare predictive power on out-of-sample data.
- Select the most robust model(s) for recommendation generation.

### *Week 12: Strategy Backtesting and Validation*

- Implement and execute a backtesting framework.
- Analyze metrics such as total/annual returns, Sharpe ratio, drawdown, and win rate.
- Refine strategies based on backtesting outcomes.

### *Week 13: Recommendation Engine & Risk Management*

- Integrate prediction outputs into a recommendation engine (Buy/Hold/Sell logic).

- Add confidence-based filtering for signal reliability.
- Apply risk management filters (e.g., volatility screens, drawdown limits).

#### *Week 14: Visualization and Dashboarding*

- Develop visualization modules for key outputs: price charts, indicator plots, model results, and backtest summaries.
- Build a basic dashboard for data exploration and decision support.

#### *Week 15: Documentation, Testing, and Final Refinement*

- Prepare comprehensive documentation (user manuals, technical notes, setup guides).
- Perform end-to-end testing of the complete pipeline.
- Collect feedback from peers and mentors, and incorporate final refinements.

#### *Week 16: Report Preparation and Presentation*

- Compile the final project report, describing methodology, results, conclusions, and future work.
- Prepare visual aids and slides for viva/presentation.
- Review all deliverables and submit the project.

## 5. Literature References

- **Géron, A. (2019).** Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow (2nd Ed.). O'Reilly Media.
- **Chan, E. (2013).** Algorithmic Trading: Winning Strategies and Their Rationale. Wiley.
- **Brownlee, J. (2018).** Machine Learning Mastery With Python.
- **Gudelek, M. U., Boluk, G., & Ozbayoglu, A. M. (2017).** "A Deep Learning Based Stock Trading Model with 2-D CNN Trend Detection," 25th SIU, IEEE.
- **National Stock Exchange of India.** <https://www.nseindia.com>
- **Patel, J., Shah, S., Thakkar, P., & Kotecha, K. (2015).** "Predicting stock and stock price index movement using Trend Deterministic Data Preparation and machine learning techniques," Expert Systems with Applications, 42(1), 259–268.
- **Fischer, T., & Krauss, C. (2018).** "Deep learning with long short-term memory networks for financial market predictions," European Journal of Operational Research, 270(2), 654–669.
- **Zhao, Z., et al. (2017).** "Time-series classification using deep forest," Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI), 3506–3512.
- **Hiransha, M., Gopalakrishnan, E. A., Menon, V. K., & Soman, K. P. (2018).** "NSE stock market prediction using deep-learning models," Procedia computer science, 132, 1351-1362.
- **Atsalakis, G. S., & Valavanis, K. P. (2009).** "Surveying stock market forecasting techniques – Part II: Soft computing methods," Expert Systems with Applications, 36(3), 5932–5941.
- **Chen, K., Zhou, Y., & Dai, F. (2015).** "A LSTM-based method for stock returns prediction: A case study of China stock market," Proceedings of the IEEE International Conference on Big Data, 2823–2824.
- **Tsantekidis, A., Passalis, N., Tefas, A., Kannianen, J., Gabbouj, M., & Iosifidis, A. (2017).** "Forecasting stock prices from the limit order book using convolutional neural networks," IEEE 19th Conference on Business Informatics (CBI), 7–12.

**Supervisor's Rating of the Technical Quality of this Dissertation Outline**

EXCELLENT / GOOD / FAIR/ POOR (Please specify): EXCELLENT

**Supervisor's suggestions and remarks about the outline (if applicable).**

Date : 25 May, 2025



(Signature of Supervisor)

Name of the supervisor : Abhinav Nag

Email Id of Supervisor : abhinav.nag@credable.in

Mob # of supervisor : 9871252551