PREDICTIVE MODELLING FOR STOCK TRADING

- Utsav Kailash Kothari
(210100637)

Introduction

Purpose of the presentation:

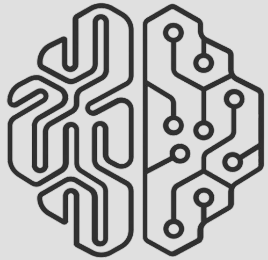
- Overview of the stock market's significance and the integration of machine learning (ML) with it.
- Explanation of ML's effectiveness in forecasting stock prices.
- Discuss on project aims, objectives, and the envisioned outcome.
- Discussion on selected ML Models, data-processing, model tuning, and more.
- Discuss completed phases and upcoming tasks.
- Discussion risk management and ethical considerations.

Stock Market - The Indian Stock Market

- The Indian stock market is a vital engine driving the India's rapid economic growth.
- I'm particularly interested in exploring the Indian market due to:
  - Global Recognition: Indian stock market is the fourth largest in the world.
  - Market Liquidity: The exchanges boast an extremely large trading volume of more than $100 billion daily.
  - Explosive growth: The number of demat accounts went to over 161 million in 2024, a significant surge in retail investors.

Why Machine Learning

- Data Driven Approach to Invest in Stocks - Technical Analysis(stock price trends, volume) and Fundamental Analysis(company financials, market conditions).
- Uncovers Hidden Patterns - ML uses advanced algorithms to detect complex patterns.
- Adapts to New Information - Integrating Sentiment analysis can help model to improve prediction over time and adapt to market dynamics
- Automation - Futuristic approach(Models can be automated to trade/invest).



THE UNDERLYING PROBLEM

- *Existing research primarily focuses on developed markets like the US.*
- *ML application to predict stock prices in India remains limited.*

Aims of the Project

Forecast Market Trends:

Develop a comprehensive predictive model utilising machine learning techniques and sentiment analysis to forecast stock market trends in the Indian Stock Market, specifically focusing on the NIFTY 50 and Sensex indices.

Leverage Sentiment Analysis:

To enhance investment strategies through the integration of a machine learning-based model capable of identifying and exploiting patterns in financial news sentiment, thereby predicting the impact on stock prices of the NIFTY 50 and Sensex indices.

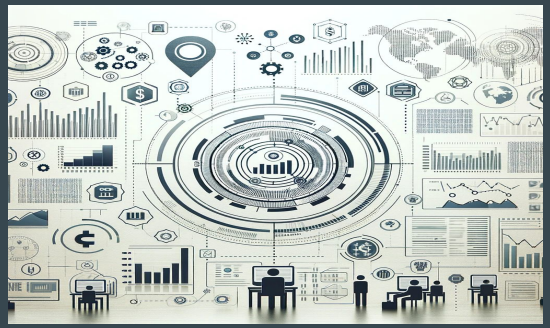
Manage Risk Proactively:

To implement a dynamic risk assessment framework within the predictive model, utilising advanced volatility prediction algorithms such as LSTM, ARIMA and GARCH for the Indian Stock Market.

Methodology

Data Collection:

- Nifty 50 and Sensex Indices represent the top 50 and top 30 companies listed on the NSE and BSE, respectively.
- Angel One API used to collect historical index prices and volume traded data.
- Financial News Articles from Moneycontrol are scraped for data collection.



Data Pre-Processing:

- Identify outliers in prices and sentiment data using IQR and z-scores. Remove and winsorize to reduce their impact on model training.
- Fill missing data points using mean/median imputation and forward fill where necessary.
- Scale price and sentiment data if necessary, to improve model performance and training efficiency.

Feature Selection:

- Integrate technical indicators like moving averages and MACD to capture market trends and volatility.
- Incorporate sentiment data(news headlines, conclusion and other parts) as a feature.
- Employ correlation analysis and feature importance scores to identify highly relevant features.

Methodology Continue - Models Selection

LSTM (Long Short - Term Memory)

- Captures long-term dependencies in time series data, ideal for finding historical patterns and trends in stock prices.

- Excellently handles non-linear relationships within the market data, providing accurate forecasts based on historical trends.

ARIMA (Autoregressive Integrated Moving Average)

- Effectively models linear trends in data, making it apt for indexes showing consistent directional movement.

- Parameters selection tailored for the Indian market, optimizing ARIMA's forecasting accuracy for stocks with linear growth trends.

GARCH (Generalized Autoregressive Conditional Heteroskedasticity)

- Explicitly models volatility in financial data, crucial for capturing market fluctuations and risk.

- Enhances investment decision-making by accurately forecasting volatility, unlike models assuming constant volatility.

Methodology Continue

Model Tuning (Hyperparameter Optimization):

- Fine-tuning Parameters: Optimize the internal settings of each model (LSTM, ARIMA, GARCH) to maximize their predictive accuracy for the Indian market.

- Grid Search or Randomized Search: Efficiently explore various parameter combinations to identify the optimal configuration for each model.

Model Evaluation:

- Evaluate model performance on unseen data segments using K-Fold Cross-validation, ensuring generalizability and avoiding overfitting.

- Use MSE, RMSE, and R-squared to assess model performance quantitatively.

Back-testing:

- Historical Data: Evaluate model prediction accuracy on historical data not used for training, ensuring real-world applicability.

- This allows for the adjustment and calibration of models based on performance insights, enhancing future prediction accuracy.

Project Planning (Gantt Chart):

- Illustrates the project timeline, clearly delineating phases from data collection through model development, tuning, evaluation, and final reporting.

- Highlights the structured approach to project management, ensuring timely completion of each phase and facilitating efficient progress monitoring.

Risks and Ethics

Technical Risks:

- Back-testing Generalizability:
  - ❏ Mitigation: Implement rolling window back-testing and real-time model monitoring to adapt to changing market conditions.
- GARCH Model Volatility Misalignment:
  - ❏ Mitigation: Employ dynamic parameter tuning to ensure GARCH model parameters reflect current market volatility.

General Risks:

- ❏ Overfitting: Addressed through rigorous cross-validation and separate test datasets.
- ❏ Technology Reliance: Mitigation strategies include staying updated and having contingency plans for potential API/platform issues.
- ❏ Accuracy & Reliability: Project emphasizes the inherent uncertainty of predictions to avoid misleading users.
- ❏ Time Management: Utilize project management tools and adhere to the defined timeline.

Ethical Considerations:

Data anonymization and responsible information handling will be practiced, ensuring data privacy and ethical use of predictive analytics.

What Has Been Achieved So Far

- **Project Proposal Submitted:** Completed and submitted the detailed project proposal, outlining the research scope, methodology, and planned execution steps, marking the start of the project's development phase.
- **Data Collection Achieved:** Successfully utilized Angel One's API for historical price data and employed web scraping to extract relevant financial news from Moneycontrol for NIFTY 50 and Sensex indices.
- **Data Pre-Processing Completed:** Implemented outlier detection and handling, missing value imputation, and normalization to ensure data integrity and readiness for analysis.
- **Feature Selection Finalized:** Identified key financial indicators and utilized sentiment analysis data, enhancing the model's predictive capability by incorporating market trends and investor sentiment.
- **Model Selection Established:** Selected and began the implementation of LSTM, ARIMA, and GARCH models, each chosen for their unique strengths in capturing the complexities of the stock market.

What is still to be done

Hyperparameter Tuning, Model Evaluation & Back-testing:

- ❏ Employ hyperparameter optimization techniques such as Grid Search and Random Search to enhance all three models' accuracy.
- ❏ Use k-Fold and Stratified k-Fold cross-validation to test model robustness and prevent overfitting.
- ❏ Back-test models against historical data of NIFTY 50 and Sensex indices to validate predictive capabilities.
- ❏ Analyze outcomes using performance metrics like MSE, RMSE, and R-squared to ensure model efficacy.

Project Report Writing:

- Develop draft report covering literature review, methodology, results, and discussion. Integrate data, model evaluation findings, and back-testing results to create a comprehensive narrative.

- Incorporate feedback from the supervisor to refine the report content.

Poster Creation & Submission:

Canva & Design: Use Canva to create an informative and visually appealing poster, the final design should effectively communicate the project's essence.

References

- Baker, M., & Wurgler, J. (2006). Investor Sentiment and the Cross-Section of Stock Returns. The Journal of Finance, 61(4), 1645–1680. doi:10.1111/j.1540-6261.2006.00885.x
- Bao, W., Yue, J., & Rao, Y. (2017). A deep learning framework for financial time series using stacked autoencoders and long-short-term memory. PLoS ONE, 12(7), e0180944.
- Beyaz, E. (2018). Effective stock price forecasting using machine learning techniques whilst accounting for the state of the market. International Journal of Engineering and Applied Sciences, 10(2), 115-122.
- Patel, M. (2020). Stock Price Prediction Using Machine Learning, BIRLA VISHVAKARMA MAHAVIDYALAYA.
- 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.
- Note - ChatGPT and Google Gemini used for photo generation and structure.


THANK YOU