



DEPARTMENT
OF
COMPUTER SCIENCE AND BUSINESS SYSTEMS

Project Review - 1

AI based smart stock selection using historical data

Team:

1. Name:Suriyakaant.S.R.
Reg. No.211722244107
- 2.Name:Shakthidharan .T
Reg. No.211722244095
- 3.Name:Mathiganth.P
Reg. No.211722244301

Supervisor Name :Mr.S.Rathna Sabapathy

Qualification:

Designation

Department of CSBS



OUTLINE

- Proposed Research Area
- Background and Motivation
- Objectives
- Literature survey
- Modules
- Expected Results
- References



PROPOSED RESEARCH AREA

- **Stock Prediction Optimization:** Applying AI to analyze historical stock price trends for improved future return predictions.
- **Risk-Return Balancing:** Exploring how models weigh expected returns versus risks, offering smarter portfolio suggestions.
- **Feature Engineering from Historical Data:** Using past trading volumes, price movements, and technical indicators as AI input features.
- **Model Comparison:** Evaluating different AI approaches—like regression, decision trees, or neural networks—for selection performance.
- **Automated Decision Support:** Creating a tool that helps investors by automatically recommending stocks based on learned patterns.



BACKGROUND AND MOTIVATION

- **Process Vast Data:** Use AI to analyze market data at a scale impossible for humans.
- **Remove Human Bias:** Make objective, data-driven decisions free from emotions like fear and greed.
- **Boost Speed & Efficiency:** Automate analysis to identify opportunities much faster than manual methods.
- **Find Hidden Patterns:** Identify subtle, predictive signals in historical data that humans would miss.
- **Democratize Investing:** Provide powerful, AI-driven analytical tools to everyday investors.



INTRODUCTION

- **Problem Statement:** Traditional stock-picking relies heavily on human judgment; AI can add data-driven precision.
- **Opportunity:** With large-scale historical data available, AI offers the chance to uncover patterns not visible to humans.
- **Goal:** Develop an AI framework that suggests top-performing stocks by learning from past behavior.
- **Scope:** Focus on a defined market (e.g., S&P 500, Indian stock exchanges) and a specific timeframe (e.g., past 5 years).
- **Impact:** Helps investors make informed decisions quickly, potentially enhancing profitability and reducing emotional bias.



LITERATURE SURVEY

| Paper Name & Year | Techniques & Features | Advantages & Applications | Limitations |
|---|--|--|---|
| Stock Price Prediction Using Transformer Models (2021) | Transformers on time-series data | Long-range dependency modeling | Data-hungry; needs regularization |
| Decision Support for Apple Futures Using Prediction Fusion (2021) | Ensemble fusion over historical data | Robustness via model fusion | Commodity-specific; limited transferability |
| Deep Reinforcement Learning for Trading Automation (2022) | DRL agent on historical bars | Optimizes trading objectives directly | High variance; simplified frictions |
| DiffSTOCK: Probabilistic Relational Stock Forecasting (2024) | DDPM + relational transformer | Captures cross-asset structure; uncertainty modeling | Complex; requires quality relation graphs |
| CNN + Bi-LSTM for Stock Index Prediction (2019) | CNN + Bi-LSTM hybrid | Better directional accuracy on indices | Index-level only; ignores costs |
| Temporal Fusion Transformer for Stock Prediction (2021) | TFT variant on historical features | Handles inputs & variable selection | Requires careful feature engineering |
| HATS: Hierarchical Graph Attention Network (2020) | Graph attention on inter-stock relations | Captures industry/market structure | Graph construction non-trivial |
| Spatiotemporal Hypergraph Convolution Network (2020) | Hypergraph convolution networks | Models higher-order stock relations | Increased complexity; scalability issues |



LITERATURE SURVEY

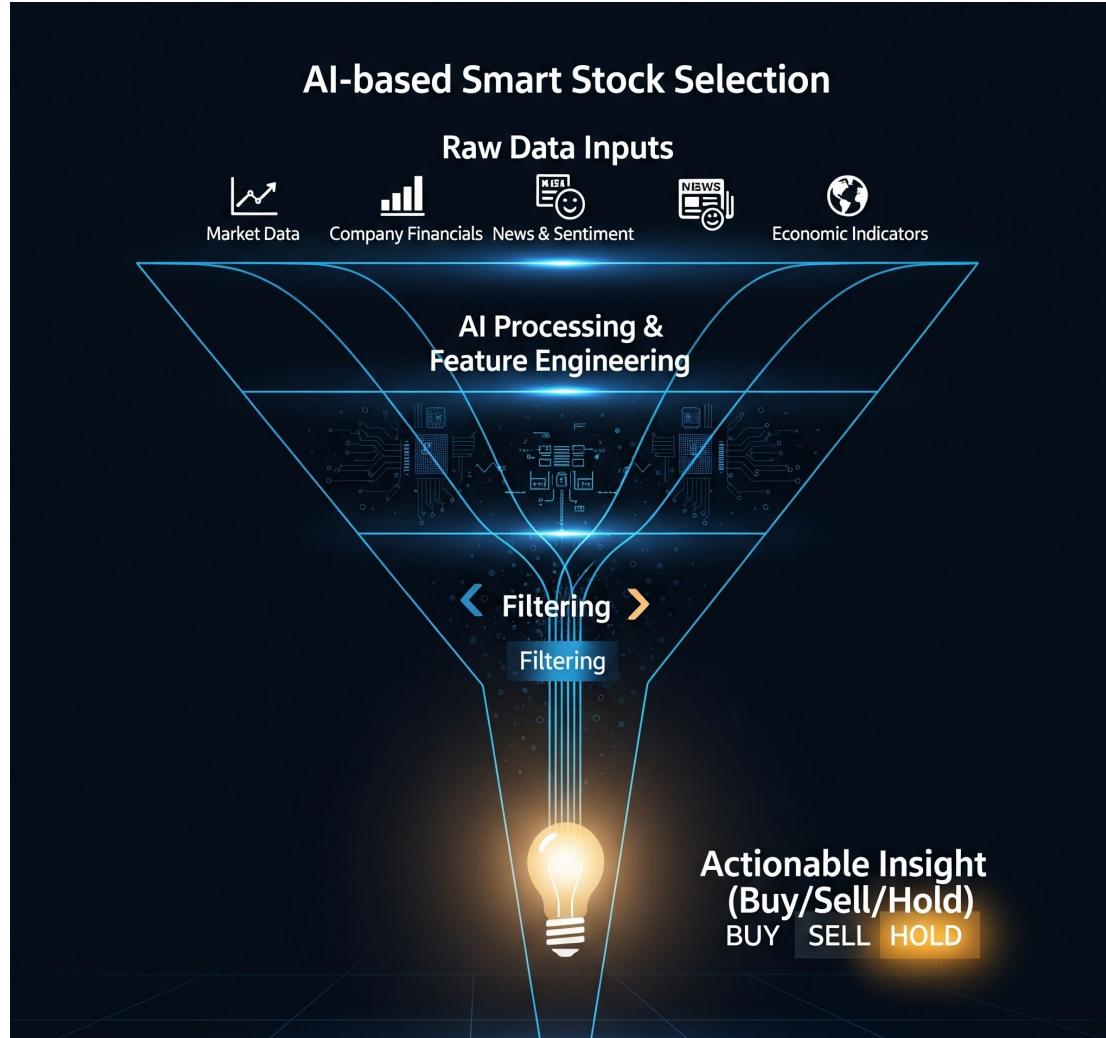
| Paper Name & Year | Techniques & Features | Advantages & Applications | Limitations |
|---|--|--|---|
| Stock Price Prediction Using Transformer Models (2021) | Transformers on time-series data | Long-range dependency modeling | Data-hungry; needs regularization |
| Decision Support for Apple Futures Using Prediction Fusion (2021) | Ensemble fusion over historical data | Robustness via model fusion | Commodity-specific; limited transferability |
| Deep Reinforcement Learning for Trading Automation (2022) | DRL agent on historical bars | Optimizes trading objectives directly | High variance; simplified frictions |
| DiffSTOCK: Probabilistic Relational Stock Forecasting (2024) | DDPM + relational transformer | Captures cross-asset structure; uncertainty modeling | Complex; requires quality relation graphs |
| CNN + Bi-LSTM for Stock Index Prediction (2019) | CNN + Bi-LSTM hybrid | Better directional accuracy on indices | Index-level only; ignores costs |
| Temporal Fusion Transformer for Stock Prediction (2021) | TFT variant on historical features | Handles inputs & variable selection | Requires careful feature engineering |
| HATS: Hierarchical Graph Attention Network (2020) | Graph attention on inter-stock relations | Captures industry/market structure | Graph construction non-trivial |
| Spatiotemporal Hypergraph Convolution Network (2020) | Hypergraph convolution networks | Models higher-order stock relations | Increased complexity; scalability issues |

PROPOSED METHODOLOGY

- **Data Collection:** Gather historical stock data (prices, volumes, market indicators) from reputable sources (e.g., Yahoo Finance, Bloomberg).
- **Preprocessing:** Clean the data, handle missing values, normalize scales, and engineer indicators like moving averages or RSI.
- **Model Selection:** Train multiple AI models—such as Random Forest, LSTM Neural Networks, and Gradient Boosting—to predict stock performance.
- **Feature Importance Analysis:** Identify which historical indicators (e.g., momentum, volatility) most influence predictions.



BLOCK DIAGRAM



EVALUATION METRICS

- **Accuracy of Predictions:** Percentage of times the model correctly identifies outperforming vs. underperforming stocks.
- **Sharpe Ratio of Portfolio:** Measures risk-adjusted return of the selected portfolio.
- **Cumulative Return:** Total return over a test period compared against benchmarks like market indices.
- **Precision/Recall for Top Stocks:** Especially useful if framing this as a classification (top-10 selection, etc.).
- **Statistical Significance (e.g., p-value):** Determine whether observed performance improvements are reliably different from random selection.



MODULES

- **Data Ingestion Module:** Downloads and stores historical data from chosen sources.
- **Preprocessing Module:** Cleans data, performs normalization, handles nulls, computes technical indicators.
- **Model Training Module:** Implements AI algorithms, handles cross-validation, hyperparameter tuning.
- **Prediction & Ranking Module:** Uses trained model to predict and rank stocks by expected return or other criteria.
- **Evaluation & Feedback Module:** Calculates metrics (returns, Sharpe ratio, accuracy), compares against benchmarks, and feeds insights back into model refinement.



REFERENCES

- **Fama, E.F., & French, K.R.** (1993). Common Risk Factors in the Returns on Stocks and Bonds. *Journal of Financial Economics*.
- **Zhang, L., & Zhou, T.** (Year). Stock Market Prediction with Artificial Neural Networks. [Journal/Conference].
- **Chen, K., Zhou, Y., & Dai, F.** (2015). A LSTM-based Method for Stock Returns Prediction: A Case Study of China Stock Market.
- **Gu, S., Kelly, B., & Xiu, D.** (2020). Empirical Asset Pricing via Machine Learning. *The Review of Financial Studies*.
- **Fama, E.F.** (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. *Journal of Finance*.

QUERIES?



THANK YOU...

