Stocks Unveiled: LSTM vs. Classical Models in Price Prediction

Introduction:

This project aims at assessing and comparing the performance of conventional machine learning models and deep learning models, especially LSTM, in predicting stock prices. The models employed are Linear Regression, Random Forest, Support Vector Machines (SVM), RNN-LSTM, CNN-LSTM, and two hybrid methods—one is a mix of CNN-LSTM with RNN-LSTM, and another is a fusion of the highest-performing conventional model with the hybrid deep learning framework. Preprocessed historical stock market data from Alpha Vantage was used with Python, Pandas, and Scikit-learn libraries to study and analyze. The performance measures employed were Accuracy, RMSE, and MSE.

Objectives:

- To implement and evaluate traditional machine learning models for stock price prediction.
- To develop and assess deep learning models, including LSTM, CNN-LSTM, and hybrid architectures.
- To compare the performance of traditional learning appropriate metrics.

Explanation of Architecture Diagram:

The architecture for this stock price prediction project is structured into three core layers: the **data layer** gathers historical stock data using the Alpha Vantage API; the **preprocessing and feature engineering layer** cleans and normalizes data, generating time-series features such as lag values and technical indicators; and the **modeling layer**, which includes traditional models (Linear Regression, Random Forest, KNN, etc.), deep learning models (LSTM, CNN-LSTM), and a final hybrid model that integrates CNN-LSTM, LSTM and a Traditional ML model to capture complex patterns. The **output layer** visualizes predictions, accuracy metrics, and investment insights.

Batch No. 6

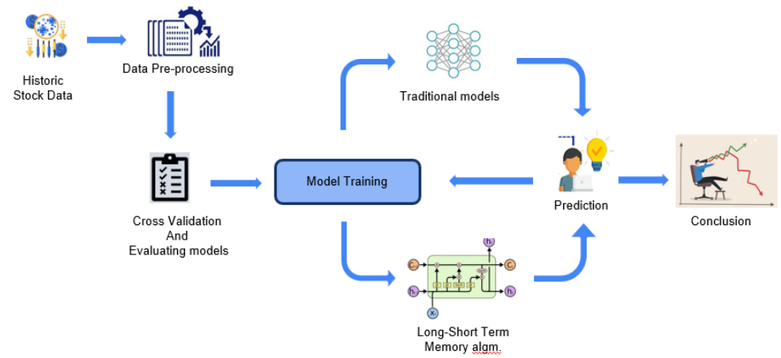
Team Member Details:

Rangineni Poojith (23245A6703)
Vadugu Karthik (22241A6759)
Lingampalli Koushik Rao (22241A6729)

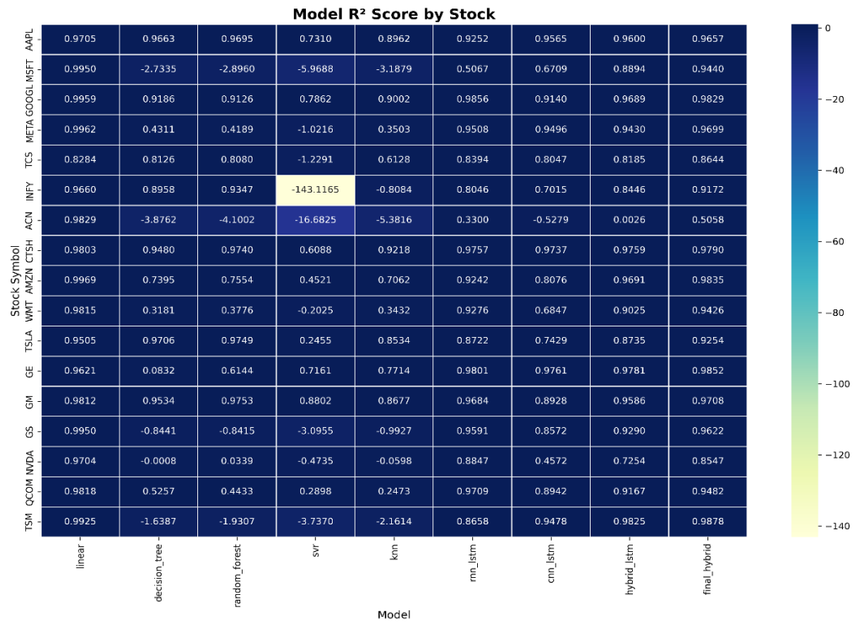
Name of the Guide:

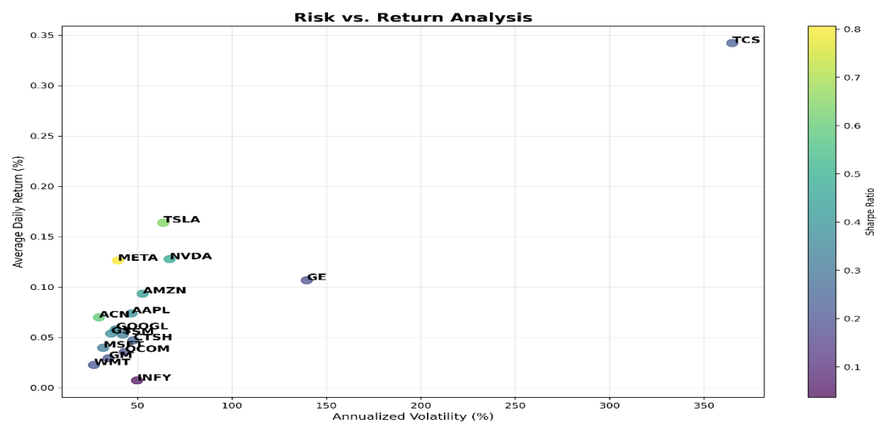
Mrs. P. Sindhuja,
Assistant Professor, Department of Data Science

Architecture Diagram:



Results:





Methodology:

- **Data Source:** Fetched daily stock prices for 18 companies via Alpha Vantage API.
- **Preprocessing:** Handled missing values, normalized data, and converted time series to supervised format.
- **Features:** Included lag values, moving averages, and technical indicators (e.g., RSI, MACD).
- **Models Used:**
  - *Traditional:* Linear Regression, Random Forest, SVR, Decision Tree, KNN
  - *Deep Learning:* RNN, LSTM, CNN-LSTM, LSTM-ANN
  - *Hybrid:* Final Hybrid (CNN + RNN + LSTM)
- **Training:** Used Python (Scikit-learn, TensorFlow); optimized using GridSearchCV and dropout.
- **Evaluation:** Assessed using R^2, MSE, and visual comparisons.
- **Visualization:** Plotted actual vs predicted prices, R^2 heatmaps, and risk-return graphs.

Conclusion:

This study compared traditional ML and deep learning models for stock price prediction across 18 companies in sectors like tech, retail, and automotive. The Final Hybrid model (CNN LSTM + RNN LSTM + Best Performed Traditional Model) outperformed others, achieving the highest R^2 scores, especially for volatile stocks. Traditional models like Linear Regression and Random Forest performed adequately for stable stocks but struggled with high-volatility stocks, often yielding poor or negative R^2 values. Risk-return analysis revealed TSLA and META offer high returns with high risk, while MSFT and AAPL present safer investment options. Sector-wise, predictions were generally robust, though companies like INFY and ACN showed variability due to domain-specific and macroeconomic influences.