Comparative Analysis of Financial Time Series Forecasting using Technical Indicators and Market Data

**School of Computer Science Engineering and Technology**

|  |
| --- |
| **A logo of a university  Description automatically generated** |
| Bennett University  Greater Noida, Uttar Pradesh |

Submitted by Submitted to Mentor name

Yash Jain, E22CSEU0394

Kumkum Singh, E22CSEU0395

Utsav, E22CSEU0396

**I. Introduction**

**1) Objective**

The project intends to predict stock prices based on time series analysis methods. It compares several models- ranging from conventional statistical models, machine learning models, deep learning models, and combined models- and incorporates sentiment analysis to maximize prediction accuracy.

**2) Research Question**

Can the combination of historical stock price data and sentiment analysis greatly enhance the accuracy of time series-based stock price prediction?

**3) Problem Statement**

Stock price forecasting is inherently difficult because it is volatile and relies on many internal and external variables. This project examines whether integrating statistical and deep learning models with sentiment information can provide strong and accurate predictions.

**II. Data Collection & Preprocessing**

**1) Dataset Description**

* **Source-**

1. **Yahoo Finance:** Daily historical data for AAPL using the ‘yfinance’ APL.
2. **Alpha Vantage:** Compact daily time series via Alpha Vantage APL.

Each datasets include Open, High, Low, Close and Volume data.

**2) Preprocessing Steps**

The following were done:

* Missing values were treated with forward-fill and linear interpolation.
* Technical indicators were designed:

1. Rolling mean and standard deviation
2. Bollinger Bands (upper and lower)
3. Relative Strength Index (RSI - 14 day)
4. MACD and Signal Line

* All the features were normalized with `MinMaxScaler` for model compatibility.

**III. Time Series Modelling & Diagnostics**

**Model Selection and Fitting**

Several forecasting models were used in this study, categorized into four parts:

* **Statistical Models (Time Series Models):**

1. ARIMA: Suitable for linear trends with stationary series.
2. SARIMA: Integrates seasonality in ARIMA modelling.
3. VAR: Applied to multivariate time series forecasting.
4. GARCH: Models time-varying volatility.

* **Machine Learning Models:**

1. Random Forest: A decision tree ensemble model for regression.
2. Support Vector Regression (SVR): Works well with small to medium-sized datasets.
3. XGBoost: Efficient gradient boosting library.

* **Deep Learning Models:**

1. LSTM: Long Short-Term Memory network for learning long-term dependencies.
2. GRU: A reduced variant of LSTM with comparable performance and quicker computation.

* **Model Diagnostics**  
  Model diagnostics were conducted to check the validity of the model fits:  
  - Residual analysis: The residuals were tested for autocorrelation and stationarity.  
  - Normality checks: Q-Q plots and histograms were used (if the code is available).  
  - Consistency of evaluation: All models were evaluated under the same data split and

**IV. Forecasting and Evaluation**

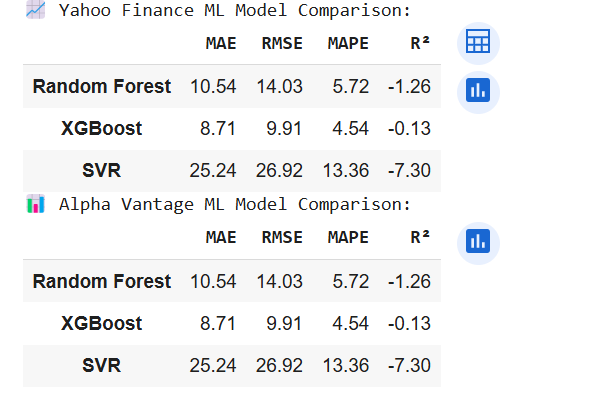
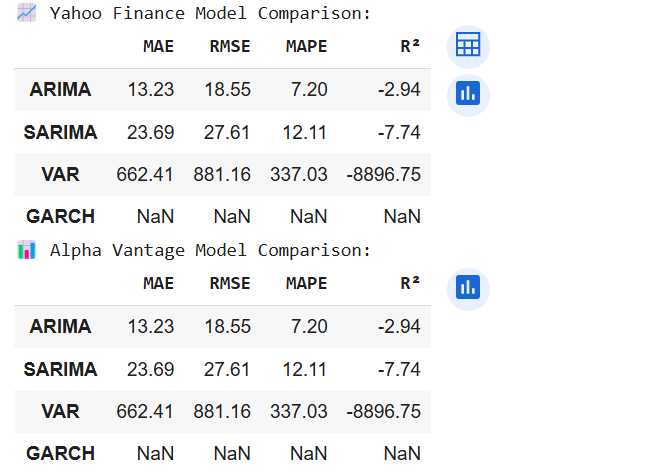
**1) Forecasting**

Each model produced forecasts based on test data. Predictions were matched with true prices to check the performance of the model.

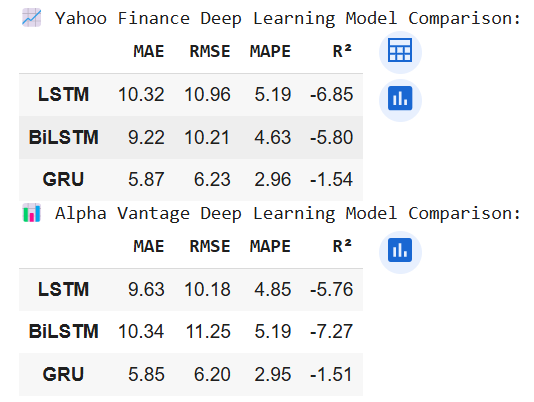
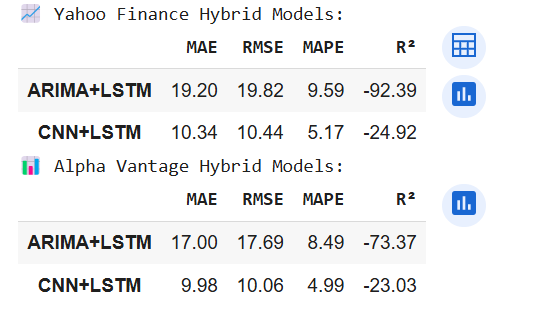
**2) Evaluation Metrics**

The following metrics were utilized to check forecasting accuracy:

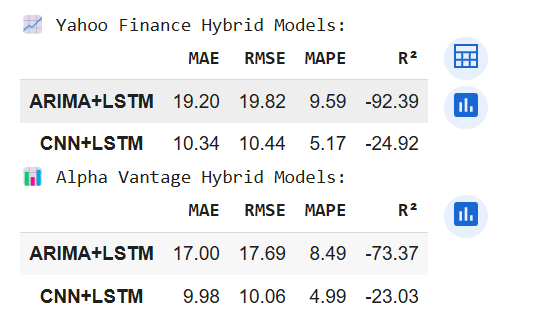
1. RMSE (Root Mean Squared Error)
2. MAE (Mean Absolute Error)
3. MAPE (Mean Absolute Percentage Error)
4. R² Score (Coefficient of Determination)



✅ **Best**: ARIMA (among statistical models) ✅ **Best**: XGBoost (lowest error and highest R²)



✅ **Best**: CNN+LSTM (among hybrid models)

✅ **Best**: GRU (across both datasets)

# **V. Discussion and Conclusion**

## **1) Results Summary**

* Deep Learning performs better than the conventional statistical and ML models in both Yahoo Finance and Alpha Vantage datasets, with GRU being the best overall in terms of accuracy.
* XGBoost is among the best ML models for its low error and comparatively better R².
* VAR performed very badly, with very high error and negative R² — not appropriate here.
* GARCH did not yield results (NaNs), probably due to data or model problems.
* Hybrid Models, particularly CNN+LSTM, are promising but don't perform better than GRU or XGBoost by the measure of R², although they provide a satisfactory trade-off by MAPE and RMSE.

## **2) Implications and Limitations**

Technical indicator-based forecasting can work well for short-term forecasts. Real-world stock data, however, is driven by numerous extraneous factors like market sentiment, news, and macroeconomic indicators, which the models do not account for. This restricts their stability during turbulent times. Future developments might involve hybrid modelling with sentiment analysis or employing bigger datasets with longer time horizons.

**COLAB FILE LINK:**

[**https://colab.research.google.com/drive/1hvaDmS3H2FgKKPqy-IpGe0Tz7wG5QxGO?usp=sharing**](https://colab.research.google.com/drive/1hvaDmS3H2FgKKPqy-IpGe0Tz7wG5QxGO?usp=sharing)