ROBERT GORDON UNIVERSITY ABERDEEN

SCHOOL OF COMPUTING SCIENCE AND DIGITAL MEDIA

MSc Business Analytics

Student Name:P.G.L.Ojithma	Matriculation Number:
Supervisor: Dr. Sameera Viswakula	Second Marker:
Project Title: Stock Price Predictability: Unravelling the Forecasting Capabilities of LSTM (Long Short Term Memory), XGBoost (Extreme Gradient Boosting), ARIMA (Autoregressive Integrated Moving Average), VAR (Vector Autoregression) and GARCH(Generalized Autoregressive Conditional Heteroskedasticity)	
	Start Date: January 2023
	Submission Date:21/08/2023

CONSENT

I agree ■
I do not agree □

That the University shall be entitled to use any results, materials or other outcomes arising from my project work for the purposes of non-commercial teaching and research, including collaboration.

DECLARATION

I confirm:

- **That the work contained in this document has been composed solely by myself and that I have not made use of any unauthorised assistance.**
- **That the work has not been accepted in any previous application for a degree.**
- **All sources of information have been specifically acknowledged and all verbatim extracts are distinguished by quotation marks.**

Student Signature: P.G.L.Ojithma	Date Signed:21/08/2023

1 Abstract

Purpose - Stock price forecasting has evolved significantly with advancements in computational models and the wealth of ample and easy data availability. This research seeks to investigate the predictive power of LSTM, XGBoost, VAR, and ARIMA models on stock prices of prominent entities like "Apple", "Eli Lilly and Company", "ExxonMobil", and the overarching S&P 500 index, utilizing web-scrapped data from Yahoo Finance from 2018 to 2023.

Design/methodology/approach- The study utilized a comprehensive data acquisition and preprocessing methodology, which involved web scraping and normalization using the min-max technique along with the Haar wavelet transformation that was applied for denoising. The research leveraged specific time steps to create input sequences for LSTM targeting short-term forecasting for the next day and the next week. Simultaneously, the XGBoost model underwent walk-forward validation, while hyperparameter tuning was accomplished for LSTM using Adam as the optimization technique. In parallel, the stationarity of the series was tested using the Dickey-Fuller Test, which is a prerequisite assumption for predictions via ARIMA and VAR. Volatility predictions for the close price returns were made using the GARCH model visualization which could be used to ascertain the risks associated with the assets.

Findings - XGBoost demonstrated superior accuracy across the entities studied, consistently delivering the lowest RMSE and MAE values. LSTM presented promising results, especially when it came to leveraging sequential data, outshining VAR and often paralleling ARIMA. Thus, it was indicated that the LSTM model should be subjected to more rigorous hyperparameter training methodologies and validation methodologies for better accuracies similar to XGBoost. However, VAR's larger deviations imply it might not be apt for precise stock price forecasting. ARIMA's performance was inconsistent, showcasing superiority in some cases and lagging in others. Thus, for stock market predictions, XGBoost and LSTM emerge as promising contenders, while ARIMA and VAR might be more situational.

Originality/value - This research amalgamates the power of machine learning, deep learning, and traditional statistical modeling to predict stock prices, bridging a gap in the current literature by incorporating more than the stock index analysis but the stock index with its selected constituent stocks analysis while opening up avenues for more comprehensive future research. By evaluating the forecasting power of the chosen models, this study offers invaluable insights for stakeholders, suggesting the potential of machine learning techniques, comparatively, in future frameworks of stock market predictions of the interested stock index and the constituent stocks.

Conclusion - The dynamic world of stock market predictions has witnessed a paradigm shift with the advent of robust computational models incorporating machine learning and deep learning

methods. This research underscored the unmatched precision of XGBoost in forecasting and the promising results of LSTM that could be improved by changing the conditions for validation and hyperparameter tuning according to the potential environment resources. While traditional models still hold relevance, contemporary techniques, as highlighted in the study, are carving the way forward, providing investors with more accurate tools and potential frameworks for stock market predictions.

2 Table of Contents