SECURE AND TRANSPARENT ACCOUNTING ANALYTICS IN REAL TIME DATA USING DEEP LEARNING

Advanced BiLSTM-Based Financial Prediction with Explainable AI and Data Security

Arockia Selvaraj A[1], Dr.P.D.R. Vijayakumar[2], Chaya Kumari L[3], Hema D[4], Nithya Shri D[5], Subasri S[6].
[2] Professor and Head, Department of Computer Science and Engineering, JCT College of Engineering and Technology, India.
[1] Assistant Professor, Department of Information Technology,
[3, 4, 5, 6] Department of Information Technology.
Info Institute of Engineering, Kovilpalayam, Coimbatore, India.

ABSTRACT

Financial data has increased rapidly in recent years, making it difficult for traditional accounting systems to manage large quantities of data quickly and efficiently. Traditional accounting systems cannot process data in real-time and therefore do not support predictive analytics or make it easier for organizations to make good decisions. This study proposes an explainable and secure real-time accounting analytics system that uses deep learning and big data technologies. This study uses a BiLSTM model to perform financial predictions and uses explainable AI techniques (SHAP and LIME) to increase transparency. The study incorporates security measures, such as encryption and anomaly detection, to secure financial data. The proposed system will facilitate the real-time processing of data, improve the accuracy of the predictions generated, and enable an organization to obtain a secure and transparent manner of analyzing financial information across Multiple industries.

Index Terms - *Real-time analytics, Deep Learning, Big Data, Explainable AI, Financial Analytics, Data Security*

INTRODUCTION

Due to a dramatic growth in the number of digital financial transactions, many accounting firms have created vast amounts of accounting data. Traditional accounting systems do not have the ability to process these large quantities of data quickly or in real-time. As a result, there are often delays in performing analysis and, as such, there is a limit to their ability to aid in making decisions. With technology advancing we can now. Analyze lots of financial data much faster. This is thanks to things like learning and big data analytics. They help us use data to make predictions and find patterns in data. Deep learning and big data analytics are really good at uncovering these patterns. They make it possible to look at amounts of financial data. Predictive analytics is one of the ways we can use these technologies to our advantage. It helps us make sense of data. Big data analytics and deep learning are key, to making this happen.

NEED OF THE STUDY

There are many limitations with existing financial analytic systems that limit their ability to effectively provide insight into the financial performance of companies and other entities. One major concern with these systems is that they can't process all of the financial data they receive in real time; therefore, data is processed and available for analysis only after the fact. A second issue with many of these predictive models is that the lack of explainability makes it difficult for users to understand how the outputs of the model were generated. There are also concerns surrounding the vulnerability of financial data due to its sensitivity; thus, the security of many analytic systems used to store and analyze that data often do not meet acceptable standards. Additionally, there is a need for greater anomaly detection, in order to identify irregular financial behavior. Therefore, the win-win solution to provide users with the ability to process financial information in real-time, have accurate prediction, have explainability and have strong security is to create a unified solution that meets those requirements. The proposed solution will accomplish those requirements.

RESEARCH METHODOLOGY

The approach proposed for Real Time Financial Data Processing and Analysis is a systematic and standardised approach. The first step in the system is to obtain financial data from multiple sources and create real-time data streams from them. The data is then pre-processed using Apache Spark to eliminate noise, and inconsistencies, and; finally the quality of the data is improved using normalisation techniques, such as Min-Max scaling and Robust scaling. The pre-processed financial data is transformed into time series data sequences using an algorithm known as the Sliding Window algorithm. The BiLSTM model has been selected for the prediction in this approach as it predicts data by capturing both past and future dependencies. Training the BiLSTM model will be improved by the use of the ADAMW optimiser during training, it applies Transfer learning and strong fine tuning for multiple datasets. to assist with increasing the efficiency of training the BiLSTM model. Interpretation of the model predictions and identifying the important features for the model, has been conducted using Explainable AI techniques, such as SHAP and LIME. To protect the financial data, security controls will be implemented including the use of AES encryption and Anomaly Detection technology. The performance of the system will be evaluated using the Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2 statistical measure.

PROPOSED SYSTEM ARCHITECTURE

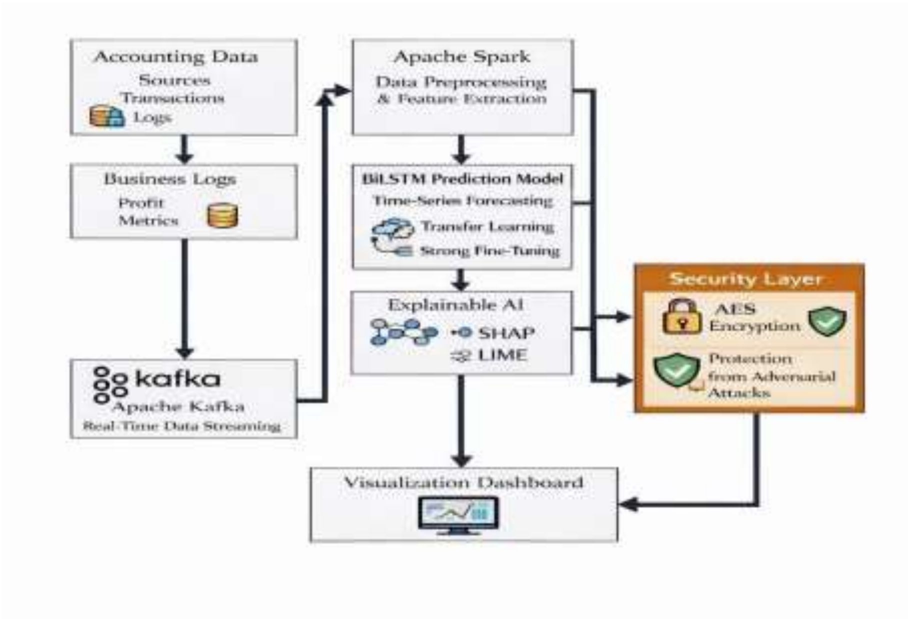

Fig.1. Proposed system architecture

The initial step of the system's implementation involves gathering accounting data, including transaction data, logs, and business metrics, from several different sources. This accounting data is then streamed in near real time through the use of Apache Kafka; by utilizing Kafka, the accounting data stream will offer a continuous flow of data. After the data is streamed, it is sent to an Apache Spark cluster for processing. During the Apache Spark data processing stage, the data will be pre-processed and features will be extracted in order to clean and provide the data for analysis. Once the data has been processed , it will be input into a Bidirectional Long Short-Term Memory (BiLSTM) model for time-series forecasting. Transfer learning and strong fine-tuning will be used to improve the time-series forecasting accuracy produced by the BiLSTM model. To provide interpretability to predictions made by the BiLSTM model, explainable AI approaches such as SHAP and LIME will help to provide interpretability into outputs produced by the BiLSTM model. A security layer protecting the financial data will also be created using AES encryption to help secure sensitive financial data and to provide protection against adversarial attacks. The results of the system will be presented through a dashboard providing visualizations of the results as well as supporting data to help support decision-making.

PROPOSED METHODS AND ALGORITHMS

The proposed system will use big data technologies, deep learning models, explainable AI techniques, and security mechanisms to do real-time financial analytics. The methodology focuses on processing large amounts of accounting data quickly while making sure that predictions are correct, data is safe, and data is clear.

Apple (manufacturing) and Walmart (e-commerce) are two of the many sources that provide financial data. The datasets have important information like open, high, low, close, and adjusted close prices, which are needed to look at financial trends and patterns. Apache Kafka streams the collected data all the time so that it can be processed in real time. This makes sure that data can be sent quickly and that the system can handle a lot of financial data without any problems. Using Apache Spark, the streamed data is then cleaned and preprocessed. To make the data more consistent and deal with outliers better, we use techniques like Min-Max normalization and Robust scaling.

Using a sliding window technique, the preprocessed data is turned into time-series sequences. This change helps to capture the time-based relationships in financial data, which are very important for making accurate predictions. The processed data is then sent to the BiLSTM (Bidirectional Long Short-Term Memory) model. This model can learn both forward and backward dependencies in time-series data, which makes it very good at predicting the future of finances. The ADAMW optimizer is used to train the model. This makes learning faster and more effective. Also, transfer learning and strong fine-tuning methods are used to make the model work better on a number of different datasets.

When the model is trained, it can make predictions about future financial values. Explainable Artificial Intelligence methods like SHAP and LIME are used to make it easier to understand. These methods help figure out which input features are most important and give clear explanations for each prediction, which makes it easier for users to understand. The system needs security. AES encryption keeps sensitive financial information safe from people who shouldn't have access to it. Also, anomaly detection methods are used to find unusual or harmful inputs, which makes sure the system is safe and reliable. We use standard metrics like Mean Squared Error (MSE), Mean Absolute Error (MAE), and R^2 score to see how well the model works. These metrics help you figure out how accurate and useful the prediction model is. Lastly, the results are shown on a visualization dashboard that has prediction graphs, 7-day forecasts, and analytical insights. This makes it easy for people like financial analysts, managers, and decision-makers to understand the results and make smart choices.

ALGORITHMS USED

The suggested system uses a mix of algorithms and methods to do accurate and efficient financial analysis:

1. BiLSTM is used to predict time series by finding connections between past and future financial data
2. The ADAMW Optimizer makes training better by speeding up convergence and lowering overfitting.
3. The Sliding Window Technique turns data into time-series sequences.
4. Min-Max Normalization puts data into a set range so that models work better.
5. Robust Scaler takes care of outliers in financial data.
6. SHAP and LIME help explain model predictions by interpreting them.
7. AES Encryption keeps data safe and private.

RESULT AND DISCUSSION

The proposed method uses dataset sets from different industries to evaluate the efficacy of their initial approach. Funding predicted by the projected system could correctly be classified as correct/failing during the evaluation phase. In addition, using a BiLSTM neural network to model historical time series improved the predictive ability of that model. Improving the quality of data produced through preprocessing allows for more accurate predictions by improving the perceived accuracy of predictions from the underlying model. Using Explainable AI (XAI), it was possible to clearly identify the most influential factors that contributed to the predictions, thus adding greater transparency to the output and allowing end-users/users of the model to better understand and use the output generated by the XAI processes. The development of security features would consist of incorporating measures such as data encryption and anomaly detection that would help protect the bank's financial institutions against unauthorized access and threats. It appears as if the entire process yielded an effective overall outcome in terms of accuracy, security, and transparency from a practical standpoint.

COMPARISON OF EXISTING & PROPOSED SYSTEM

Model / Dataset	Target Type	MSE	MAE	R^2 Score	Performance Level
Existing Method (NeuralACT)	Single Target	0.00048	0.01786	0.9243	Good
Proposed – APPLE Dataset	Single Target	0.00069	0.0183	0.9895	Excellent
Proposed – WALMART (High)	Multi Target	0.00134	0.02111	0.9473	Excellent
Proposed – WALMART (Low)	Multi Target	0.00142	0.02568	0.9443	Excellent

Table.1 Comparison of existing &Proposed system

Apple(Manufacturing)

Total datasets : 11170 Training : 8906 (80%)

Testing : 2227 (20%)

Target : Adj_close(Single Target) Future Prediction: Next 7 days

1.BiLSTM Training

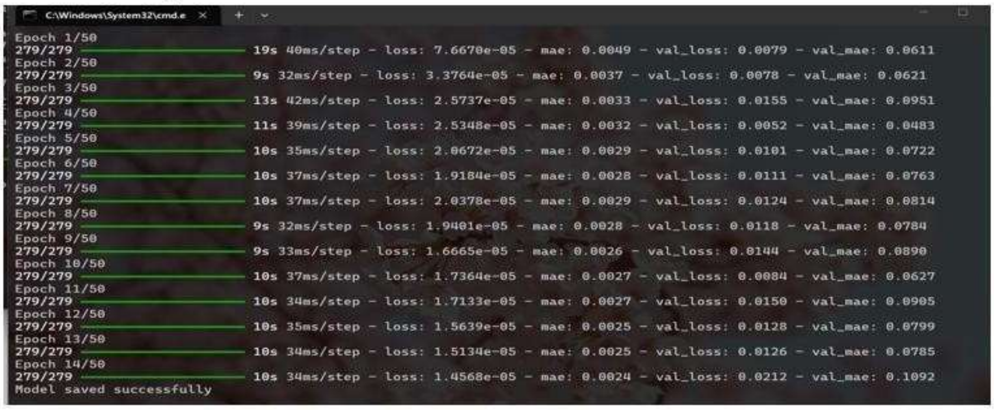

Fig.2. BiLSTM Model Training on Apple Dataset

The above-mentioned output highlights how the BiLSTM neural network is being trained using the dataset containing information about the manufacturing process at Apple. As one can see, the losses and MAE decrease consistently, which means that the training is going well and it finishes successfully with a saved model.

2.Testing Metrices

Fig.3. Performance Metrics for Apple Dataset (MSE, MAE, R² Score)

Adj_close Performance

MSE : 0.00069

→ Indicates very low prediction error

MAE : 0.0183

→ Shows minimal average deviation

R² Score: 0.9895

→ Indicates excellent model accuracy (≈ 98.95%)

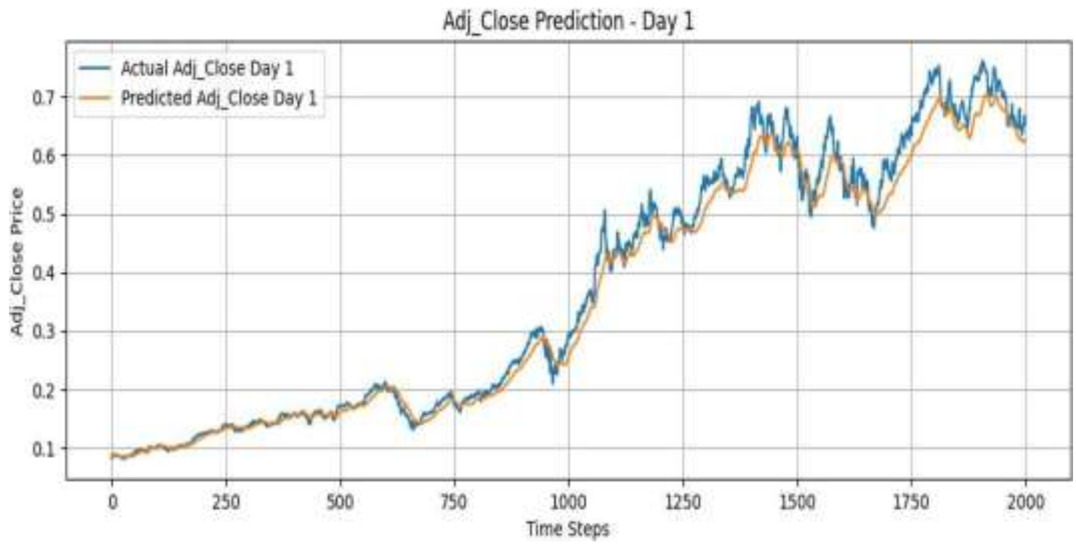

Fig.4. Actual vs Predicted Adj_Close Price (Day 1)

From these charts, one can observe how successful the use of the BiLSTM architecture was, in terms of errors like the MSE, MAE, and R^2, which indicate the accuracy of the prediction. There is a chart of the timeline, comparing the real values with those predicted by the model, and the two almost duplicate one another. Overall, it looks like the model is capturing the trends quite accurately.

3.Future Prediction for 7 days

```
Next 7 Days Adj Close Predictions

Day | Scaled   | Real Price
---------------------------------
1   | 0.864014 | 0.86
2   | 0.867442 | 0.87
3   | 0.870083 | 0.87
4   | 0.872016 | 0.87
5   | 0.873411 | 0.87
6   | 0.874170 | 0.87
7   | 0.874256 | 0.87
```

Fig.5. 7-Day Future Prediction of Adj_Close Prices

This shows the model's predicted adjusted closing prices for the next 7 days based on learned patterns. The values are gradually increasing, indicating a slight upward trend in the predicted prices.

4. Explainability AI

SHAP

```
--- Day 1 ---
Predicted Adj_Close : 0.86
Profit/Loss         : +0.01
Trend               : Increase

SHAP Contributions:
Open   : -0.058145 (26.77%) -> Decrease
High   : -0.037186 (17.12%) -> Decrease
Low    : -0.065324 (30.07%) -> Decrease
Close  : -0.056541 (26.03%) -> Decrease
Volume : +0.000036 (0.02%) -> Increase

Top Contributors:
- Low (Decrease)
- Open (Decrease)

Explanation:
Most features contributed negatively, pushing the price downward:
• Open negatively contributed (-0.058145)
• High negatively contributed (-0.037186)
• Low negatively contributed (-0.065324)
• Close negatively contributed (-0.056541)

Some features had minor positive impact:
• Volume positively contributed (0.000036)

However, despite negative contributions, the model predicts an INCREASE due to model bias and time-series learning.
```

Fig.6. SHAP-Based Explainability Analysis for Apple Dataset

Effect on feature describes the way each input affects the prediction. It shows whether an input increases the prediction or decreases it.

LIME

```
--- Day 1 ---
Predicted Adj_Close : 0.86
Profit/Loss         : +0.01
Trend               : Increase

LIME Contributions:
Open   : +0.049722 (41.23%) -> Increase
High   : +0.000000 (0.00%) -> Decrease
Low    : +0.049083 (40.70%) -> Increase
Close  : +0.021793 (18.07%) -> Increase
Volume : +0.000000 (0.00%) -> Decrease

Explanation:
Most features contributed positively, pushing the price upward:
• Open positively contributed (0.049722)
• Low positively contributed (0.049083)
• Close positively contributed (0.021793)

Hence, the model predicts an INCREASE in Adj_Close.
```

Fig.7. LIME-Based Explainability Analysis for Apple Dataset

This shows how LIME explains the model's prediction by showing how each feature affects the result. Most features contribute positively, indicating an increase in the predicted price.

WALMART(E-Commerce)

Total datasets: 13264 Training:10581(80%) Testing:2646 (20%)

Target: High and Low (MultipleTargets) Future Prediction: Next 7 days

1.BiLSTM Training

Fig.8. BiLSTM Training with Transfer Learning and strong fine tuning on Walmart Dataset

This shows the Walmart e-commerce dataset used with Transfer Learning and strong fine-tuning on multiple datasets.It indicates the model is effectively learning patterns from combined data for better prediction accuracy.

2.Testing Metrices

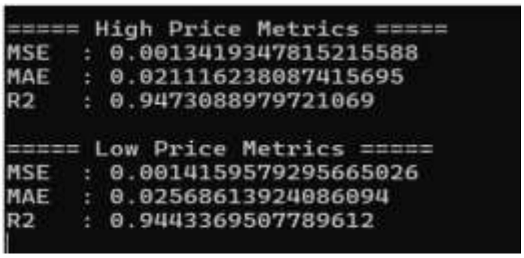

Fig.9. Performance Metrics for Walmart Dataset (High & Low Prices)

1. **High price performance**

MSE: 0.00134

⟶ Indicates Low prediction error MAE: 0.02111

⟶ Minimal average deviation R^2 Score: 0.9473

⟶ indicates excellent model accuracy (≈ 94.73%)

2. Low Price Performance

MSE: 0.00142

⟶ Slightly higher error than high price MAE: 0.02568

⟶ Minimal Average deviation R^2 Score: 0.9443

⟶ indicates excellent model accuracy (≈ 94.43%)

3.Future Prediction for 7 days

```
Next 7 Days Predictions (High & Low)

Day | Scaled High | Scaled Low | Real High | Real Low
-------------------------------------------------------------
1   | 0.262540 | 0.262463 | 0.26 | 0.26
2   | 0.249363 | 0.248864 | 0.25 | 0.25
3   | 0.235565 | 0.234595 | 0.24 | 0.23
4   | 0.223172 | 0.221787 | 0.22 | 0.22
5   | 0.212965 | 0.211269 | 0.21 | 0.21
6   | 0.204648 | 0.203410 | 0.20 | 0.20
7   | 0.198880 | 0.197805 | 0.20 | 0.20
```

Fig.10. 7-Day Future Prediction of High and Low Prices

This shows the predicted high and low prices for the next 7 days based on the trained model. It means the prices may slowly decrease over the coming days.

4.SHAP

```
==============================
Walmart High & Low Prediction
==============================
Current High: 0.27
Current Low : 0.27

=========== 7-DAY SHAP INTERPRETATION (HIGH & LOW) ===========

--- Day 1 ---
Predicted High : 0.17
High Change    : -0.10
Predicted Low  : 0.16
Low Change     : -0.10
Trend          : Decrease

SHAP Contributions:
Open   : +0.001938 (23.33%) -> Increase
High   : +0.002380 (28.64%) -> Increase
Low    : +0.002266 (27.27%) -> Increase
Volume : +0.001725 (20.76%) -> Increase

Top Contributors:
- High (Increase)
- Low (Increase)

Explanation:
Most features contributed positively, pushing values upward.
--------------------------------------------------
```

5.LIME

```
=========== 7-DAY LIME INTERPRETATION (HIGH & LOW) ===========

--- Day 1 ---
Predicted High : 0.17
High Change    : -0.10
Predicted Low  : 0.16
Low Change     : -0.10
Trend          : Decrease

LIME Contributions:
Open   : +0.006502 (26.99%) -> Increase
High   : +0.006797 (28.22%) -> Increase
Low    : +0.010787 (44.79%) -> Increase
Volume : +0.000000 (0.00%) -> Decrease

Explanation:
Most features contributed positively, pushing values upward.
--------------------------------------------------
```

Fig.10&11 Explainability for High and Low Price Prediction(SHAP&LIME)

This shows model explainability using SHAP and LIME techniques for high and low price predictions. SHAP highlights the contribution of each feature to the model's output.

LIME provides a local explanation for individual predictions. Most features like open, high, and low contribute positively to the results. Overall, these methods improve transparency and trust in the model's predictions.

6. Protection from Adversarial Attacks

Fig.12&13 Protection Against Adversarial Attacks

7. AES Encryption

Fig.14&15. AES Encryption Process for Securing Financial Data

This emphasizes the following security aspects: one is that of defending against any potential threat or attack from an adversary by removing harmful files, while the other is the implementation of AES for securing files using a password.

Comparison of R^2 scores

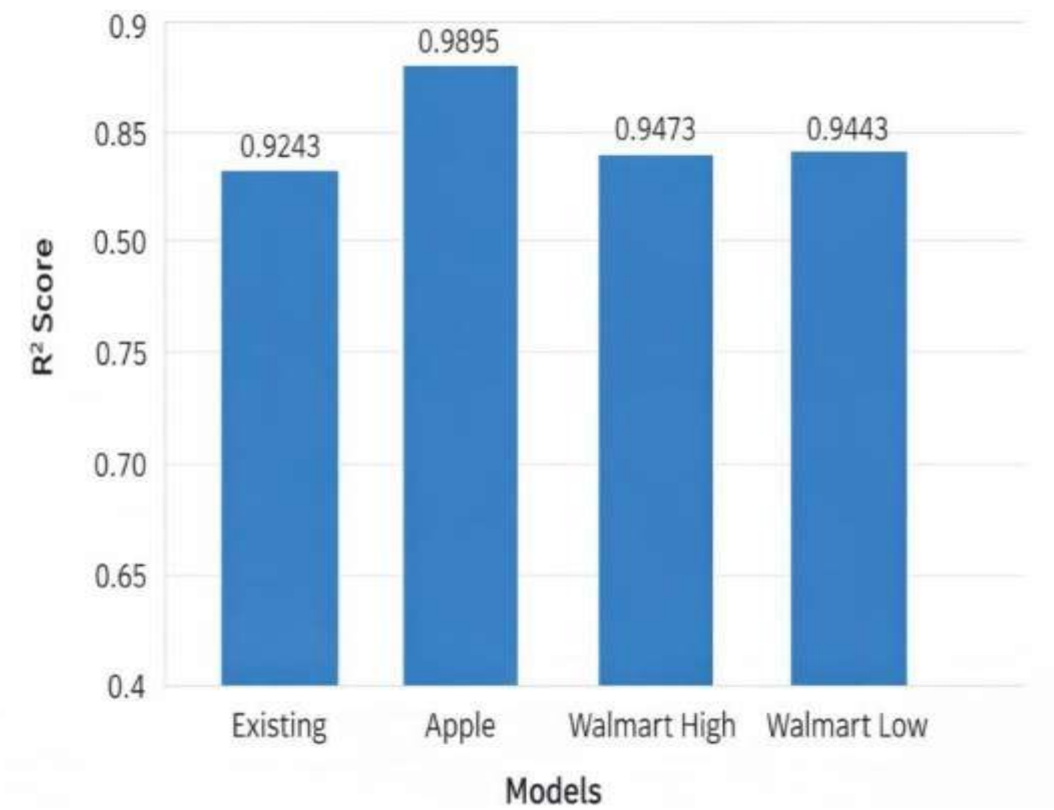

Fig.16 Comparison of R^2 scores Across Datasets

CONCLUSION

This paper introduces an innovative and trustworthy real-time financial analytics framework designed using the latest developments in deep learning, as well as big data. The framework effectively processes large quantities of financial information, predicts future behaviour, and offers users both visibility into their analysis and assurance regarding the security of their data. Utilising the BiLSTM algorithm will improve the accuracy of predictions made by the framework; on the other hand, implementing an explainable AI drives the interpretation of results to the user's level of understanding. The system also has security measures to keep records safe. This research provides a way to predict what people will do with their money in the future. It helps people make decisions because it is clear and secure. The financial records are protected so people can trust the system. This research is, about behaviours and it helps with decision-making by being transparent and secure.

FUTURE ENHANCEMENTS

The suggested system can be additionally improved through various enhancements that would make it more scalable, efficient, and user-friendly. Cloud hosting can be used to ensure scalability and real-time availability of the system regardless of location. Data base can be included into the system to facilitate storage and retrieval of large quantities of data. Moreover, integration of external databases containing information about market trends and other economic indicators will help improve financial forecasts generated by the system. The system can be enhanced through automated reports and machine learning algorithms which will allow for updating of the model using new data. Finally, the system can benefit from including an API that will integrate it with other financial platforms and business applications.

ACKNOWLEDGMENT

We would like to thank our project advisors for the help and support they have provided during this project; we would also like to thank our institution for providing us with the resources to complete this project.

REFERENCE

[1] S. Hochreiter and J. Schmidhuber, "Long Short-Term Memory," Neural Computation, 1997.

[2] A. Graves, "Bidirectional LSTM Networks," 2005.

[3] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," 2015.

[4] I. Loshchilov and F. Hutter, "Decoupled Weight Decay Regularization (AdamW)," 2019.

[5] S. M. Lundberg and S. I. Lee, "A Unified Approach to Interpreting Model Predictions (SHAP)," 2017.

[6] M. T. Ribeiro, S. Singh, and C. Guestrin, "Why Should I Trust You? Explaining the Predictions of Any Classifier (LIME)," 2016.

[7] National Institute of Standards and Technology (NIST), "Advanced Encryption Standard (AES)," 2001.

[8] Ribeiro, M. T., Singh, S., & Guestrin, C., "Why Should I Trust You? Explaining the Predictions of Any Classifier," *KDD*, 2016.

[9] Vaswani, A. et al., "Attention Is All You Need," *NeurIPS*, 2017.

[10] Kingma, D. P., & Ba, J., "Adam: A Method for Stochastic Optimization," *ICLR*, 2015.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.

[12] T. Chen and C. Guestrin, "XGBoost: A Scalable Tree Boosting System," *ACM SIGKDD*, 2016.

[13] J. Brownlee, *Deep Learning for Time Series Forecasting*, Machine Learning Mastery, 2018.

[14] A. Géron, *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow*, O'Reilly, 2019.

[15] M. Zaharia et al., "Apache Spark: A Unified Engine for Big Data Processing," *Communications of the ACM*, 2016.

[15] N. Kreps et al., "Kafka: A Distributed Messaging System for Log Processing," *LinkedIn Engineering*, 2011.

[16] C. Molnar, *Interpretable Machine Learning*, 2020.

[17] R. Guidotti et al., "A Survey of Methods for Explaining Black Box Models," *ACM Computing Surveys*, 2018.

[18] D. Silver et al., "Mastering the Game of Go with Deep Neural Networks," *Nature*, 2016.