CSC447 AI, Spring 2023, Instructor: Kruttika Sutrave

Sayantan Manik: sayantan.manik@coyotes.usd.edu

Project Title:

Use of Financial news sentiment data to analyze the stock market movement using various Deep Learning techniques: A case study of Bitcoin (BTC-USD)

Abstract:

This project investigates the use of financial news sentiment data to analyze and enhance stock market predictions using various deep learning techniques. The case study of Bitcoin (BTC-USD) is used to demonstrate the effectiveness of these techniques in predicting stock prices. The study uses one year of historical Bitcoin stock data along with normalized financial news sentiment data to predict Bitcoin stock prices for 30 days.

The project is divided into several stages, including data collection and preprocessing, model training and testing, and performance evaluation. The LSTM, CNN, and ANN deep learning techniques are used to train and test the models, and the models are evaluated based on their prediction accuracy and error rate.

The findings suggest that incorporating financial news sentiment data improves the prediction accuracy of deep learning techniques, especially when using the CNN model. The LSTM model performed better without the sentiment data. These results highlight the importance of data preprocessing and model selection in achieving accurate predictions.

The study provides insights into the effectiveness of deep learning techniques in predicting stock prices and their potential for practical applications in the finance industry. The incorporation of financial news sentiment data into stock market prediction models could improve the

accuracy of these models, and hence, aid investors and financial analysts in making better investment decisions.

In conclusion, this project demonstrates the potential of using financial news sentiment data to analyze and enhance stock market predictions using various deep learning techniques. The findings provide useful insights into the effectiveness of these techniques and their practical applications in the finance industry.

Introduction:

The stock market is a complex and dynamic system that is influenced by many factors, including economic indicators, political events, and financial news. Predicting stock prices accurately has always been a challenging task for researchers and practitioners in the finance industry. However, with the advent of deep learning techniques, there has been a growing interest in exploring their potential for improving stock market predictions.

In recent years, there have been several studies that have investigated the use of deep learning techniques, such as Long Short-Term Memory (LSTM), Convolutional Neural Network (CNN), and Artificial Neural Network (ANN), for stock market prediction. These techniques have shown promise in capturing the complex relationships between the various factors that affect stock prices.

In addition, financial news sentiment data has been shown to be a useful source of information for predicting stock prices. Sentiment analysis techniques can be used to extract sentiment scores from financial news articles, which can then be incorporated into stock market prediction models.

This project aims to investigate the use of financial news sentiment data to analyze and enhance stock market predictions using various deep learning techniques. A case study of Bitcoin (BTC-USD) is used to demonstrate the effectiveness of these techniques in predicting stock prices. The study uses one year of historical Bitcoin stock data along with normalized financial news sentiment data to predict Bitcoin stock prices for 30 days.

The project is divided into several stages, including data collection and preprocessing, model training and evaluation, and comparison of the results obtained with and without the financial news sentiment data. The findings of the study provide insights into the effectiveness of deep learning techniques in predicting stock prices and the potential for practical applications in the finance industry.

Motivation:

The accurate prediction of stock prices has always been a critical issue for investors, financial analysts, and researchers in the finance industry. The stock market is a highly volatile and dynamic environment, influenced by various factors, such as economic indicators, political events, and breaking news. Therefore, predicting stock prices using traditional statistical models has proven to be a challenging task.

However, the recent advancements in deep learning techniques have shown promise in improving stock market prediction accuracy. These techniques can capture complex relationships between multiple variables and provide better predictions compared to traditional statistical methods.

In addition, financial news sentiment data has emerged as a useful source of information for predicting stock prices. Sentiment analysis techniques can be used to extract sentiment scores from financial news articles, which can provide valuable insights into the market sentiment.

The use of deep learning techniques combined with financial news sentiment data has been a recent focus of research in the finance industry. This project aims to investigate the effectiveness of using financial news sentiment data to analyze and enhance stock market predictions using various deep learning techniques. The case study of Bitcoin (BTC-USD) is chosen because of its popularity and volatility in recent years, making it an ideal candidate to demonstrate the potential of these techniques.

The findings of this project are expected to provide valuable insights into the effectiveness of deep learning techniques in predicting stock prices and the importance of incorporating financial news sentiment data in stock market prediction models. These insights could be used by investors and

financial analysts to make better investment decisions and optimize their portfolios.

Proposed Methodology:

The proposed methodology involves collecting data, preprocessing the data, creating features, visualizing the data, selecting appropriate machine learning models, training the models, evaluating the model performance, interpreting the results, deploying the model, and plotting the comparison graphs. The methodology aims to develop a predictive model that incorporates sentiment analysis to predict the stock market trends accurately.

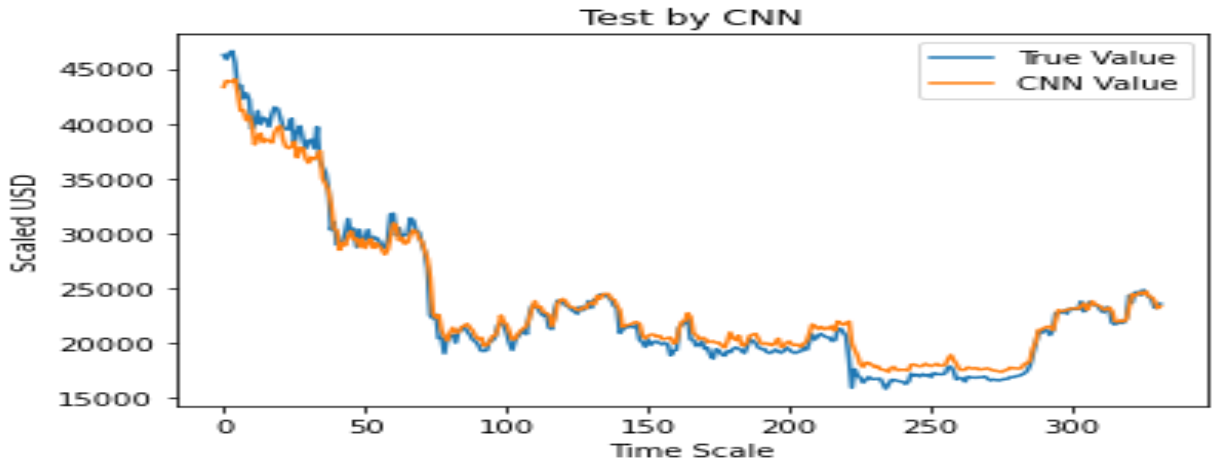
1. Data Collection: Collect historical stock market data from reliable sources like Yahoo Finance. Collect news articles sentiment data from sources like EOD Historical Data API
2. Data Preprocessing: Preprocess the collected data, which includes data cleaning, data integration, and data transformation. The stock market data needs to be transformed to extract the relevant features like open price, close price, high price, low price, and volume. The sentiment data is then attached to the overall data set.
3. Feature Engineering: Create new features using the sentiment data to enhance the predictive power of the model. For example, use the sentiment data and the number of corresponding analyzed articles from which the sentiment score was obtained.
4. Data Visualization: Visualize the data to explore patterns, relationships, and correlations between the variables.
5. Model Selection: Select appropriate machine learning models like LSTM, CNN, and ANN which are best suited to the time-series data with sentiment analysis. Evaluate the model performance using appropriate metrics like Mean Squared Error (MSE) or Root Mean Squared Error (RMSE).

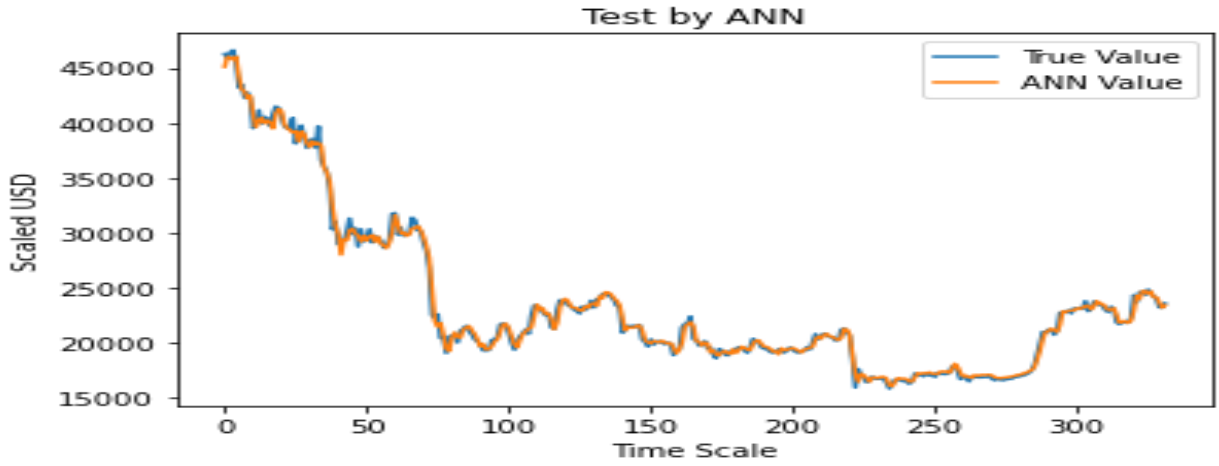
6. Model Training: Train the machine learning models using the preprocessed data. Split the data into training and test sets using TimeSeriesSplit, which splits the data based on time order.
7. Model Evaluation Evaluate the model performance using appropriate metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), or Mean Absolute Percentage Error (MAPE). Model
8. Model Testing: Test the best-performing model on a separate test dataset to evaluate its generalization performance.
9. Model Interpretation: Interpret the model results to gain insights into the relationships between the variables and the stock market trends.
10. Deployment: Deploy the model in a production environment to make predictions on new data. Update the model periodically to ensure that it adapts to the changing market conditions.
11. Plotting: Plot graphs comparing each model's prediction to the actual price. Plot an overall graph to compare the prediction of all the models with and without sentiment analysis to the actual price.

Results: (Tables, Graphs, Snapshots, etc.)

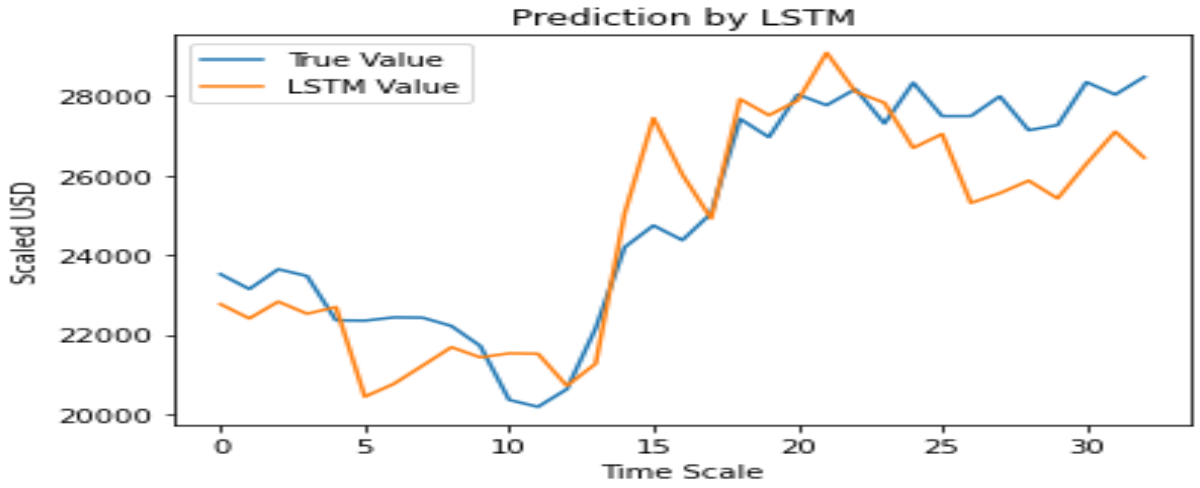
Without Sentiment Data-

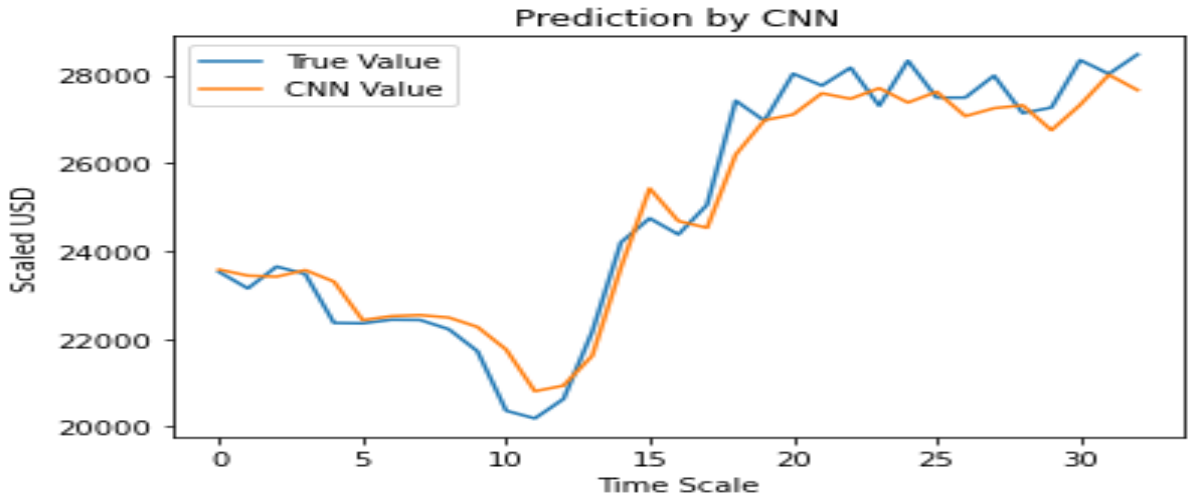
Test data analysis:

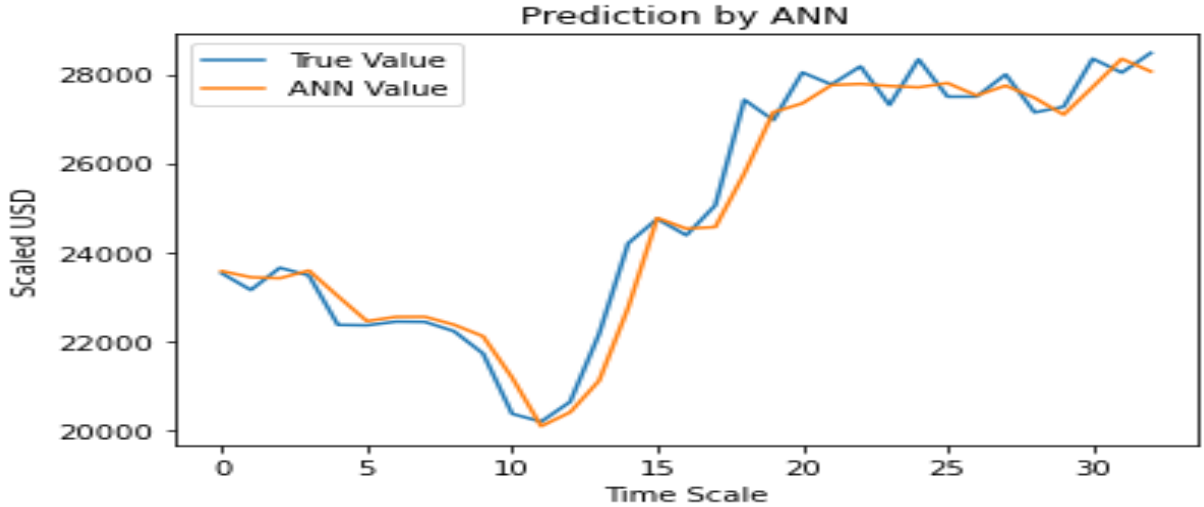


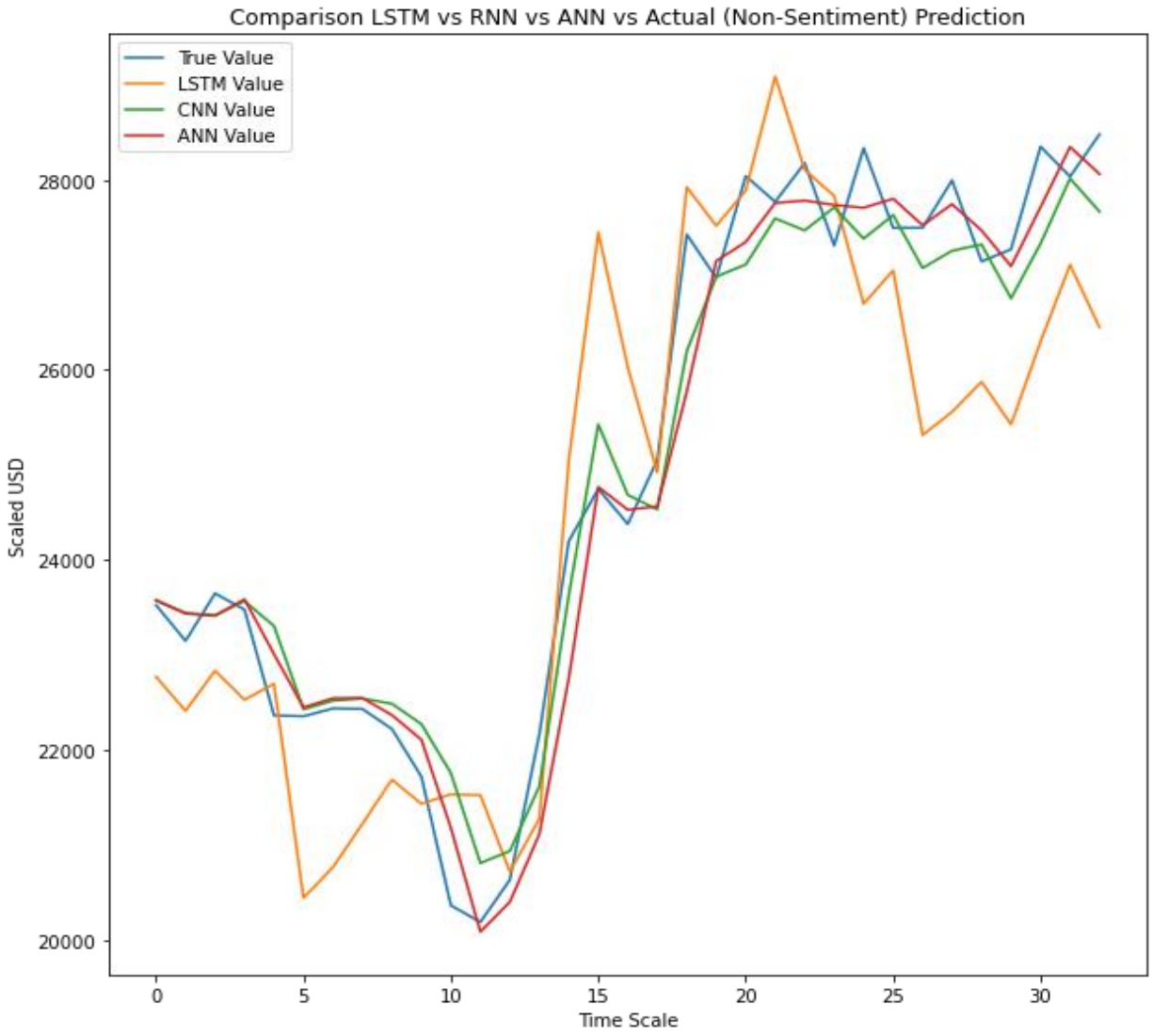


Prediction data analysis:







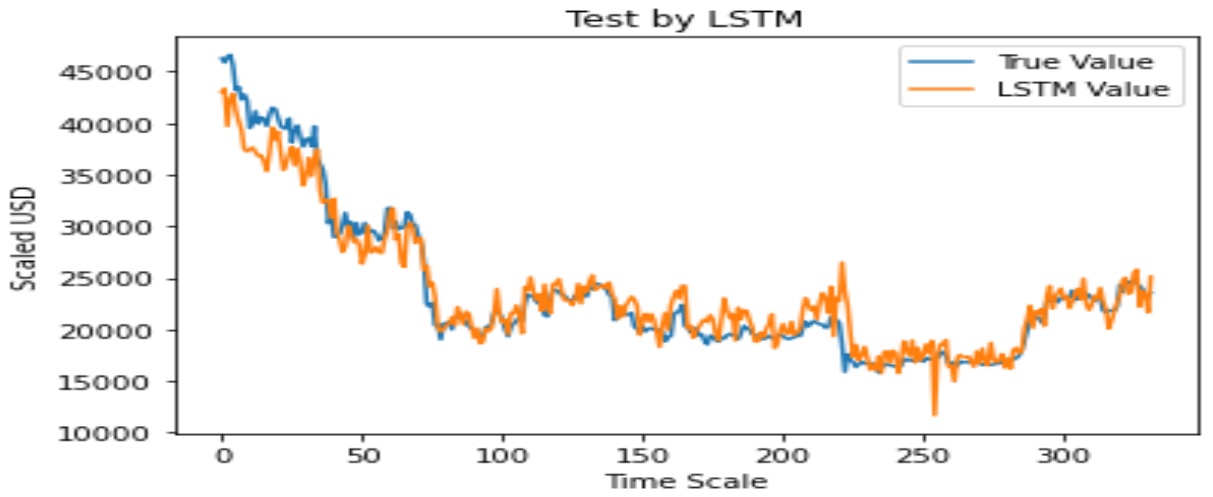


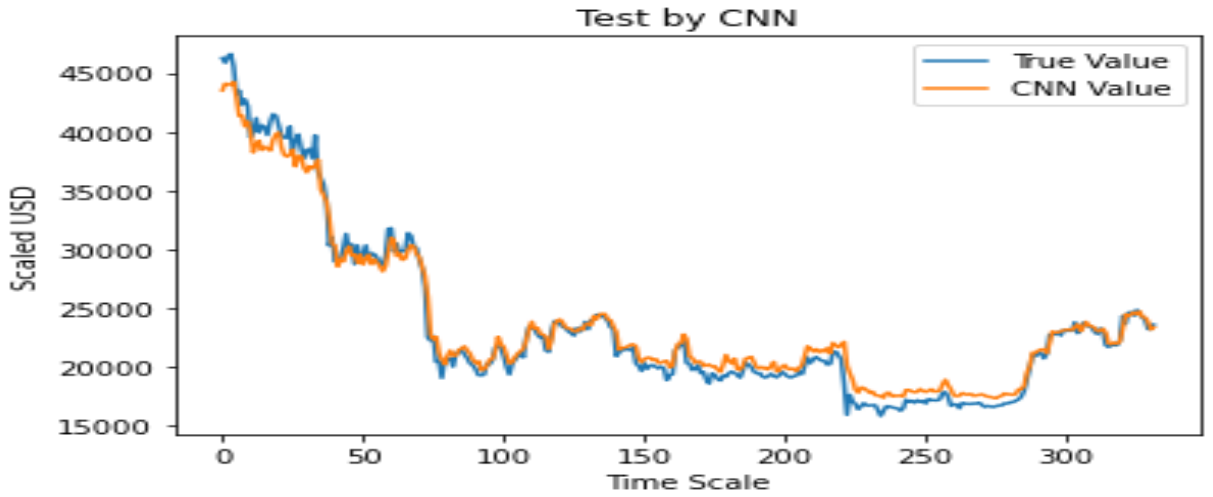
Error and Accuracy Analysis:

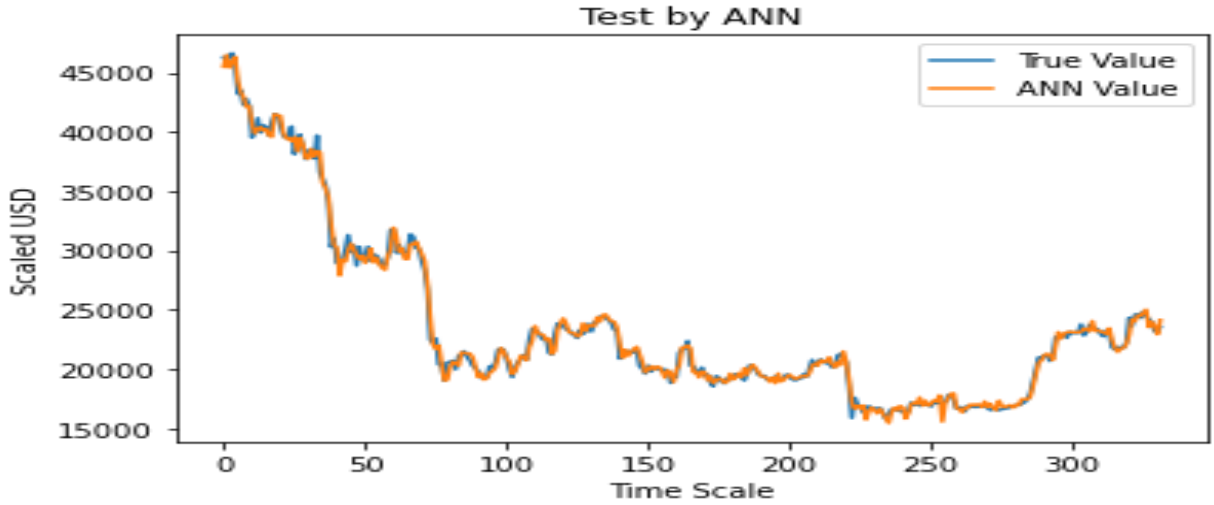
	Training Set MSE:	Training Set RMSE:	Testing Set MSE:	Testing Set RMSE:	Training Set Accuracy:	Testing Set Accuracy:
LSTM	96873762.01	9842.45	14770076.20	3843.19	68.08%	53.65%
CNN	92757377.25	9631.06	13031678.14	3609.94	68.77%	56.46%
ANN	102342650.21	10116.45	14348369.41	3787.92	67.19%	54.31%

With Sentiment Data-

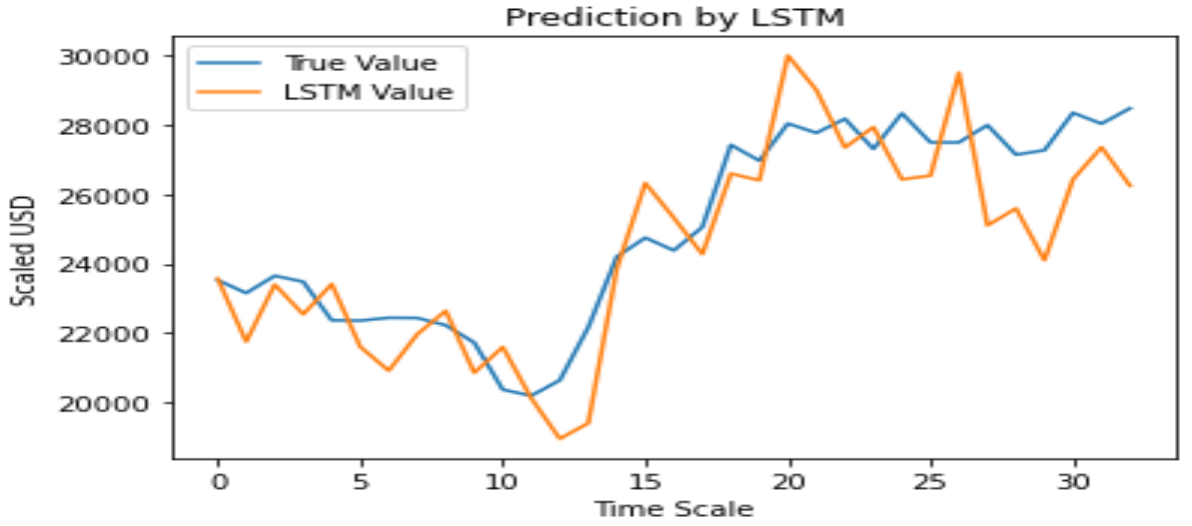
Test data analysis:

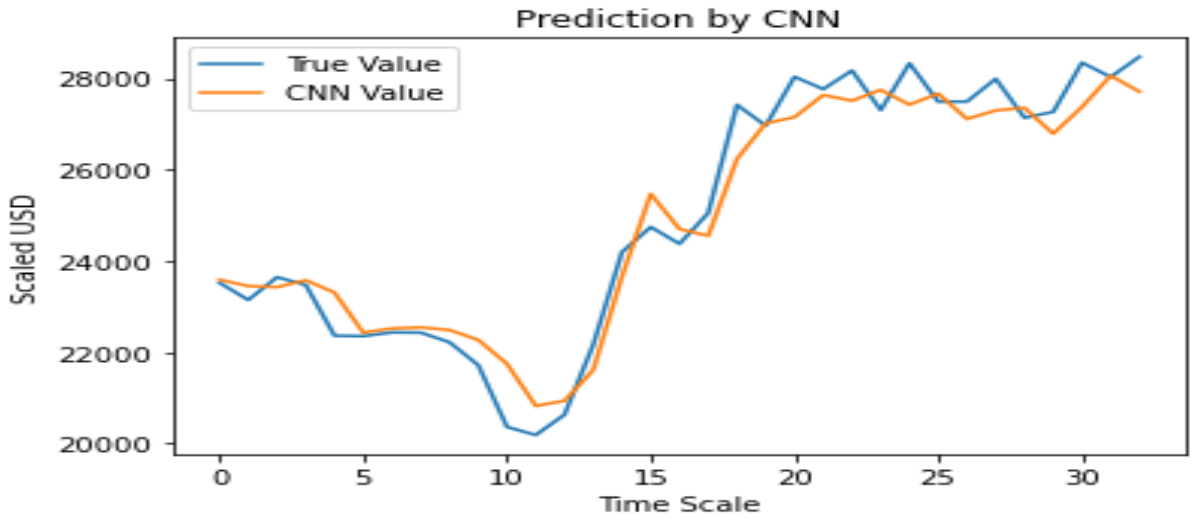


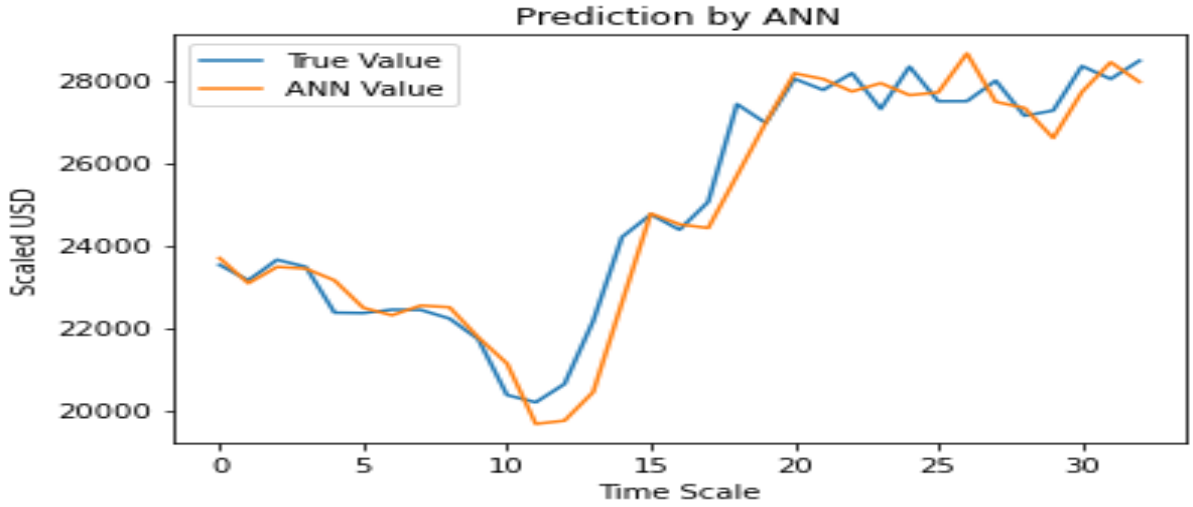




Prediction data analysis:



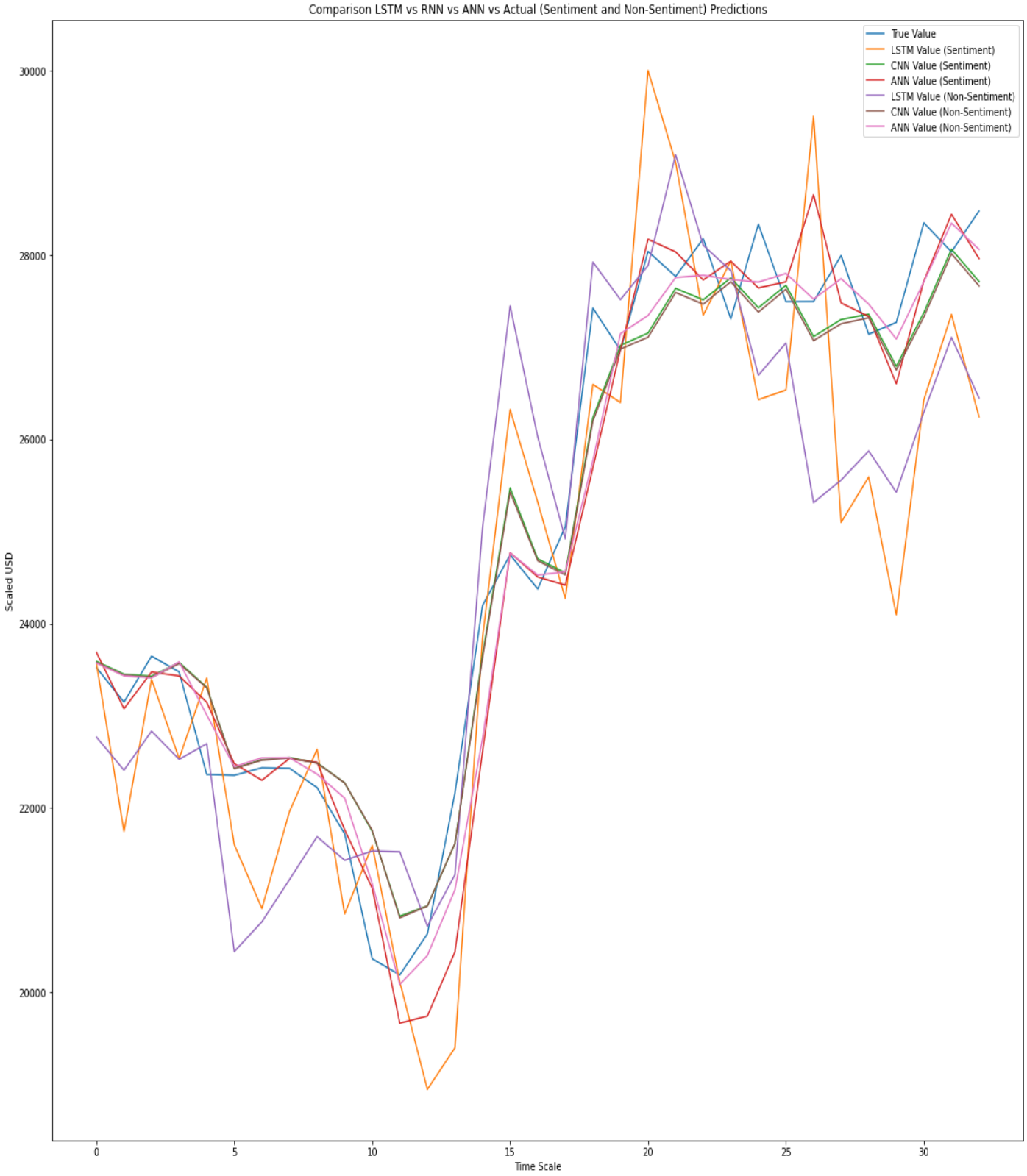




Error and Accuracy analysis:

	Training Set MSE:	Training Set RMSE:	Testing Set MSE:	Testing Set RMSE:	Training Set Accuracy:	Testing Set Accuracy:
LSTM	88756233.72	9421.05	16108296.46	4013.51	69.45%	51.59%
CNN	93390948.47	9663.90	13112926.39	3621.18	68.66%	56.33%
ANN	102942615.32	10146.06	15245150.02	3904.50	67.10%	52.91%

Overall:


Comparison LSTM vs RNN vs ANN vs Actual (Sentiment and Non-Sentiment) Predictions
True Value
LSTM Value (Sentiment)
CNN Value (Sentiment)
ANN Value (Sentiment)
LSTM Value (Non-Sentiment)
CNN Value (Non-Sentiment)
ANN Value (Non-Sentiment)
Scaled USD
30000
28000
26000
24000
22000
20000
0
5
10
15
20
25
30
Time Scale

Conclusions:

In this project, we investigated the use of financial news sentiment data to analyze and enhance stock market predictions using various deep learning techniques. We used a case study of Bitcoin (BTC-USD) to demonstrate the effectiveness of these techniques in predicting stock prices.

Our findings suggest that incorporating financial news sentiment data can improve the prediction accuracy of deep learning techniques, especially when using the CNN model. The LSTM model performed better without the sentiment data. These results highlight the importance of data preprocessing and model selection in achieving accurate predictions.

Our study provides valuable insights into the effectiveness of deep learning techniques in predicting stock prices and their potential for practical applications in the finance industry. The incorporation of financial news sentiment data into stock market prediction models could improve the accuracy of these models, and hence, aid investors and financial analysts in making better investment decisions.

In conclusion, this project demonstrates the potential of using financial news sentiment data to analyze and enhance stock market predictions using various deep learning techniques. The findings provide useful insights into the effectiveness of these techniques and their practical applications in the finance industry. Our results suggest that future research in this area could focus on exploring the use of other sentiment analysis techniques and deep learning models for stock market prediction.

References: (Recent Literature references)

[1] Mehta P, Pandya S, Kotecha K. Harvesting social media sentiment analysis to enhance stock market prediction using deep learning. PeerJ Comput Sci. 2021 Apr 13;7:e476. doi: 10.7717/peerj-cs.476. PMID: 33954250; PMCID: PMC8053016.

[2] Burgess, Michael; Javed, Faizan; Okpara, Nnenna; and Robinson, Chance (2022) "Stock Forecasts with LSTM and Web Sentiment," SMU Data Science Review: Vol. 6: No. 2, Article 10. Available at: https://scholar.smu.edu/datasciencereview/vol6/iss2/10

[3] Nabipour M, Nayyeri P, Jabani H, Mosavi A, Salwana E, S S. Deep Learning for Stock Market Prediction. Entropy (Basel). 2020 Jul 30;22(8):840. doi: 10.3390/e22080840. PMID: 33286613; PMCID: PMC7517440.

[4] Stelios D. Bekiros (2010) Fuzzy Adaptive Decision Making for Boundedly Rational Traders in Speculative Stock Markets European Journal of Operational Research 202(1) :285-293.

[5] Strader, Troy J.; Rozycki, John J.; ROOT, THOMAS H.; and Huang, Yu-Hsiang John (2020) "Machine Learning Stock Market Prediction Studies: Review and Research Directions," Journal of International Technology and Information Management: Vol. 28: Iss. 4, Article 3. DOI: https://doi.org/10.58729/1941-6679.1435 Available at: https://scholarworks.lib.csusb.edu/jitim/vol28/iss4/3

[6] Yahoo Finance stock data,

https://finance.yahoo.com/quote/BTC-USD?p=BTC-USD&nn=1

[7] EOD Historical Data Sentiment data, https://eodhistoricaldata.com/financial-apis/sentimental-data-financial-api/