**MINI PROJECT PAPER COMPARISION**

BATCH A8

| Name of the project | Algorithms and Datasets used | Performance | Disadvantages and future enhancements | Metrics and accuracy |
|---|---|---|---|---|
| Predicting Stock Market Trends Using Machine Learning and Deep Learning Algorithms Via Continuous and Binary Data; a Comparative Analysis – Mojtaba Nabipour, Pooyan Nayyeri, Hamed Jabani, Shahab S, Amir Mosavi | This study compares nine machine learning models (Decision Tree, Random Forest, Adaptive Boosting (Adaboost), extreme Gradient Boosting (XGBoost), Support Vector Classifier (SVC), Naïve Bayes, K-Nearest Neighbors (KNN), Logistic Regression and Artificial Neural Network (ANN)) and two powerful deep learning methods (Recurrent Neural Network (RNN) and Long short-term memory (LSTM). Dataset: Four stock market groups, namely diversified financials, | Nine machine learning models and two deep learning methods (were employed as predictors. We supposed two approaches for input values to models, continuous data and binary data, and we employed three classification metrics for evaluations. Our experimental works showed that there was a significant improvement in the performance of models when they use binary data instead of a continuous one. Indeed, deep learning algorithms were our superior models in both approaches. | Overall, it is obvious that all the prediction models perform well when they are trained with continuous values (up to 67%), but the models' performance is remarkably improved when they are trained with binary data (up to 83%). | For Continuous data: RNN – F1 Score ~ 0.85 Accuracy ~ 0.84 LSTM – F1 Score ~ 0.85 Accuracy ~ 0.85 SVC – F1 Score ~ 0.74 Accuracy ~ 0.71 Logistic Regression – F1 Score ~ 0.74 Accuracy ~ 0.71<br><br>For Binary data: RNN – F1 Score ~ 0.89 Accuracy ~ 0.89 LSTM – F1 Score ~ 0.89 Accuracy ~ 0.88 SVC – F1 Score ~ 0.86 Accuracy ~ 0.86 Logistic Regression – F1 Score ~ 0.86 Accuracy ~ 0.86 |

| | | | | |
|---|---|---|---|---|
| | petroleum, non-metallic minerals and basic metals, from the Tehran stock exchange. | | | |
| Indian Stock Market Prediction using Deep Learning – Ayan Maiti, Pushparaj Shetty D | Long Short-Term Memory (LSTM) model and the Generative Adversarial Network (GAN) model. Dataset: Yahoo Finance is being actively traded on India's National Stock Exchange (NSE). | LSTM and GAN models to predict the stock prices on the Indian market and compare their performance. LSTM outperformed GAN significantly with an error of 0.074 compared to 0.32 obtained by GAN. | The model is not able to predict the direction of the price movement. The increase in the number of look-back days did not have a significant improvement in results on the LSTM model. | The results of the model update cycles on the predictive performance show that M=1200 and N=300 produce the best results of RMSRE=0.073884, obtained at look-back of 90 days. |
| Stock Market Prediction Web Service Using Deep Learning by LSTM – Mohammad Mahabubul Hasan, Pritom Roy, Sabbir Sarkar and Mohammad Monirujjaman Khan | Deep learning Long Short-Term Memory (LSTM) model Dataset: DSE Bangladesh | The web-based application demonstrates machine learning outperforms statistical and regression techniques in forecasting share prices. | To add different parameters to our model for better prediction results and add new features to our web service to make this system much more robust and sustainable. | Test results have given 70% accurate prediction stats. |
| Deep learning for | Long Short-Term | Experimental | For future work, | This approach |

| | | | | |
|---|---|---|---|---|
| stock prediction using numerical and textual Information<br>-Ryo Akita, Akira Yoshihara, Takashi Matsubara, Kuniaki Uehara | Memory (LSTM) model.<br>Representation of textual information by employing the technique of Paragraph Vector.<br>Representation of numerical information using regression. Used MeCab as the morphological Analyser<br>Dataset.<br>We used the morning edition of the Nikkei newspaper published from 2001 to 2008 for our experiments, with the<br>news from year 2001 to 2006 as the training data, 2007<br>as validation data, and 2008 as test data. The target 10 companies were chosen from Nikkei 225 and the same industries | results showed that distributed representations of textual information are better than the numerical-data-only methods and Bag-of-Words based methods, LSTM was capable of capturing time series influence of input data than other models, and considering the companies in the same industry<br>was effective for stock price prediction. | we would like to incorporate more<br>technical indices such as the moving average (MA) and the moving average convergence divergence (MACD) for better<br>profit-making capabilities. | was able to make the more profits from all industries compared to considering a separated single company.<br>Hence, it is effective for predicting<br>stock prices to consider multiple companies together. |
| Stock Market Prediction Using Machine Learning –<br>Ishita Parmar, Navanshu | Regression-Based Model<br>Long Short-Term Memory (LSTM) Network-Based Model | A. Regression-Based Model:<br>Regression based Model is used for predicting | In the future, the accuracy of the stock market prediction system can be further improved | Regression:<br>The R-square confidence test resulted in a confidence score of 0.86625 |

| | | | | |
|---|---|---|---|---|
| Agarwal, Sheirsh Saxena | Dataset: Yahoo Finance (Five variables: open, close, low, high and volume) | continuous values through some given independent values. Regression uses a given linear function for predicting continuous values<br>B. Long Short-Term Memory (LSTM) Network-Based Model: LSTM is the advanced version of Recurrent-Neural Networks (RNN) where the information belonging to the previous state persists. LSTM regulates error by giving aid to the RNNs by retaining information for older stages making the prediction more accurate. Thus proving itself as much more reliable compared to other methods | by utilizing a much bigger dataset than the one being utilized currently. Furthermore, other emerging models of Machine Learning could also be studied to check for the accuracy rate resulting from them. | LSTM: The model resulted in a Train Score of 0.00106 MSE (0.03 RMSE) and a Test Score of 0.00875 MSE (0.09 RMSE). |
| Stock price prediction using machine learning techniques - Sumeet Sarode, | LSTM (Long Short-Term Memory) is used for predicting Dataset: Real-time news | The system combines price prediction based on historical and real-time data along with news | The future quant funds will obviate risks that are seized by unforeseen news events and | |

| | | | | |
|---|---|---|---|---|
| Harsha G. Tolani, Prateek Kak, Lifna C S | that is collected from various websites providing financial news. | analysis. It takes the latest trading information and analysis indicators as its input. For news analysis, only the relevant and live news is collected from a large set of business news. The filtered news is analysed to predict sentiment around companies. The results of both analyses are integrated together to get a response that gives a recommendatio n for future increases. | make it more pliable and sturdy. | |
| Machine Learning Approaches in Stock Price Prediction: A Systematic Review - Payal Soni, Yogya Tewari and Prof. Deepa Krishnan | Partial Least Squares Classifier (PLS Classifier)<br><br>LSTM coupled with sentiment analysis<br><br>ARIMA<br><br>Dataset: Yahoo and NSE-India | Explores the different techniques that are used in the prediction of share prices from traditional machine learning and deep learning methods to neural networks and graph-based approaches. The best performing model out of the ones created in is the GCN based graph, modelled from | The model considered 12 technical indicators to identify patterns in the stock market.However, the accuracy level lies between 50-70%, thus to increase the level of accuracy, a higher number of technical indicators can be used. Focus on combining the sentiment analysis of | For Partial Least Squares Classifier (PLS Classifier) - Average Error Value = 0.81225<br><br>LSTM coupled with sentiment analysis - Matthews Correlation Coefficient (MCC) = 0.04092 Accuracy = 52.27%<br><br>ARIMA - For 3 time steps |

| | | | | |
|---|---|---|---|---|
| | | news co-mentions. A reason for this could be that the graph is causation based instead of being correlation-based. | stocks related information and the numeric value associated with the historical value of stocks in predicting stock prices can be done. | ahead: RMSE: ~15% MAPE: 20-25% MAE: ≤15% For 9 time steps ahead: RMSE: 15-20% MAPE: 15-20% MAE: 10-15% |
| A Prediction Approach for Stock Market Volatility Based on Time Series Data. – Sheikh Mohammad Idrees, M. Afshar Alam, Parul Agarwal | ARIMA Model. ( Autoregressive integrated moving average) The publicly available time series data of Indian stock market has been used for this study. | Auto-regressive processes have a certain degree of unpredictability or randomness built in, that occasionally makes it capable to predict future trends pretty well. ARIMA approach is good enough for handling time series data, and as such can be very constructive in various real world problems like that of health sector, education, finance and other practical domains for prediction. | The predicted time series has been compared with the actual time series, which shows roughly a deviation of 5% mean percentage error for both Nifty and Sensex on average. | For Nifty: p-value = 0.9099<br><br>For Sensex: p-value = 0.8682. |
| Stock Closing Price Prediction using Machine Learning Techniques - | Linear Regression, Random Walk Theory (RWT), Moving Average | The comparative analysis based on RMSE, MAPE and MBE values clearly indicate | The historical dataset available on company's website consists of only few | For ANN, RMSE ~ 1.10 MAPE ~ 1.07% MBE ~ -0.0522 |

| | | | | |
|---|---|---|---|---|
| Mehar Vijh, Deeksha Chandola, Vinay Anand Tikkiwal, Arun Kumar | Convergence / Divergence (MACD) and also using some linear models like Autoregressive Moving Average (ARMA), Autoregressive Integrated Moving Average (ARIMA), for predicting stock prices. The dataset includes 10 year data from 4/5/2009 to 4/5/2019 of Nike, Goldman Sachs, Johnson and Johnson, Pfizer and JP Morgan Chase and Co. | that ANN gives better prediction of stock prices as compared to RF. Results show that the best values obtained by ANN model gives RMSE (0.42), MAPE (0.77) and MBE (0.013). | features like high, low, open, close, adjacent close value of stock prices, volume of shares traded etc., which are not sufficient enough. For future work, deep learning models could be developed which consider financial news articles along with financial parameters such as a closing price, traded volume, profit and loss statements etc., for possibly better results. | For RN,<br>RMSE ~ 1.29<br>MAPE ~ 1.14%<br>MBE ~ -0.0521 |
| STOCK PRICE PREDICTION USING LSTM, RNN AND CNN-SLIDING WINDOW MODEL -Sreelekshmy Selvin, Vinayakumar R, Gopalakrishnan E.A, Vijay Krishna Menon, Soman K.P | We have used three different deep learning architectures, RNN, LSTM and CNN for this work. The data set consists of minute wise stock price for 1721 NSE listed companies for the period of July 2014 to June | For comparison we have used ARIMA, which is a linear model used for forecasting. From error percentage obtained it is clear that deep learning models are outperforming ARIMA. | The changes occurring in the stock market may not always be in a regular pattern or may not always follow the same cycle. CNN architecture is capable of identifying the changes in trends. For the proposed methodology CNN is identified | ERROR PERCENTAGE<br>STOCK RNN LSTM CNN<br>Infosys 3.90 4.18 2.36<br>TCS 7.65 7.82 8.96<br>Cipla 3.83 3.94 3.63 |

| | | | | |
|---|---|---|---|---|
| | 2015. We trained the model using the data of Infosys and was able to predict stock price of Infosys ,TCS and Cipla. | | as the best model compared to the other two models. | |