PREDICTION OF ACCURACY FOR FUTURE PRICE OF STOCK PREDICTION USING LSTM MODEL COMPARED WITH SVM MODEL

R.Rohith Kumar[1], Sashi Rekha[2*]

Article History: Received: 12.12.2022 **Revised:** 29.01.2023 **Accepted:** 15.03.2023

Abstract

Aim: The objective of the work is to predict the Stock Price Prediction Using LSTM (Long Short Term Memory) Model Compared with SVM (Support Vector Machine). To achieve accuracy a novel SVClassifier is used, **Method and Materials:** Accuracy and loss are performed with a DATA dataset from the keras library. The total sample size is 20. The two groups Convolutional Long Short Term Memory (N=10) and Support Vector Machine algorithms (N=10),
Results: The result proved that LSTM with better accuracy of 97% than SVM accuracy of 85%. The independent sample T-Test value ($p < 0.05$) with confidence level of 95% $p=0.082$ ($p<0.05$) it is statistically insignificant with a pretest power of 80%. The two algorithms LSTM and SVM are statistically satisfied with
Conclusion: prediction of stock price significantly seems to be better in LSTM

Keywords: Stock market, Machine learning, Stock Price Prediction , Novel Support Vector Machine, Accuracy

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu, 602105
[2*]Project guide, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Saveetha University, Chennai, Tamil Nadu,602105

1. Introduction

The Indian financial exchange is an incredibly perplexing framework wherein enormous volumes of information are produced each second and change quickly because of different assorted elements(Ayan and Eken 2021). Organisations partition their possession as stocks. A stock is a speculation that represents proprietorship in an organisation. Buying a company‟s stock means buying possession in that organisation. Consequently the financial backers who put resources into a company‟s stock become a piece of the company‟s generally speaking benefit or misfortune (Jia et al. 2019). Henceforth by putting resources into the financial exchange one can proficiently increment one‟s total assets. Securities exchange is a stage for purchasing, selling and exchanging organisation stocks (Sunny, Maswood, and Alharbi 2020). The securities exchange is non-straight, broken and changes quickly as it is influenced by numerous assorted factors like reports in news, monetary exercises, political occasions, and different other numerical components (Yan 2021)

In the last five years, Google scholar identified almost 17,300 research articles on Stock Marking using machine learning Changed Manuscript Received on October 15, 2019 Shashank Singh, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu Maaz Ahmad, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu Aditya Bhattacharya, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu M. Azhagiri, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, Tamil Nadu Buying the right organisation stock and selling it when its worth is higher than its unique worth outcomes in making benefits(Roy, Ghosh, and Senapati 2021). At the point when the pattern of the market is effectively anticipated, benefits are made relative to the venture. The test is to effectively anticipate the stock costs precisely to limit misfortune and expand benefit(Wei 2019). AI (ML) and Deep Learning (DL) can assume a vital part in the field of money. By utilising the different ML and DL calculations, we can proficiently gather and picture the colossal volumes of financial exchange information and control it successfully to exactly conjecture the securities exchange patterns (Sunil 2021). The well known customarily utilised Machine Learning calculations for anticipating securities exchange patterns are the different relapses like the Linear Regression and The Polynomial relapse and Classifications like the Random Forest grouping, SVM order novel support vector machine, KNN calculation. However, the relapse calculations alone can't factor the different components influencing the securities exchange and don't give exact outcomes (Bathla 2020). All the previously mentioned ML calculations have been utilised in different investigations and demonstrated to offer poor to direct outcomes and couldn't give great exactness. A superior way to deal with anticipated securities exchange drifts all the more precisely and effectively is to make a Hybrid model that joins these famous ML grouping calculations and get a consolidated exact expectation. In the area of DL, RNN networks have been customarily put to use to foresee the financial exchange patterns, however the RNN networks are not extremely effective because of their concern of Vanishing Gradient. To place it in basic words, RNN holds data for just short time frames, for example in the event that we need the data after a little term of time, it could be accessible, however after various emphasess, the first data gets lost (Ali 2021). This limit is overwhelmed by utilising Long Short Term Memory organisations.Our team has extensive knowledge and research experience that has translated into high quality publications(K. Mohan et al. 2022; Vivek et al. 2022; Sathish et al. 2022; Kotteeswaran et al. 2022; Yaashikaa, Keerthana Devi, and Senthil Kumar 2022; Yaashikaa, Senthil Kumar, and Karishma 2022; Saravanan et al. 2022; Jayabal et al. 2022; Krishnan et al. 2022; Jayakodi et al. 2022; H. Mohan et al. 2022)

LSTM networks have ended up being exceptionally productive and adaptable in foreseeing the securities exchange patterns with extremely high exactness. Long Short Term Memory is a piece of the Recurrent Neural Network structure however varies from the traditional RNN in the manner it adjusts its past state information after each emphasis ie. The first information isn't lost after a couple of emphasis. The expectation of the financial exchange patterns includes not just figuring the verifiable and continuous information yet additionally calculating the news that fundamentally influences the securities exchange(Liu, Wang, and Zheng 2019). The financial exchange is impressively influenced by political occasions and other news in the media. Feeling Analysis is the most common way of deciding the opinion behind a piece of text and to decide the writer‟s demeanour and feeling behind composing the text to comprehend the feeling behind the news and foresee what it will mean for the stock value, a Sentiment Analyzer should be utilised to decide the opinion behind the news and measure its extremity (Zeng and Liu 2018). This work proposes a half and half model consolidating the different ML arrangement calculations viz SVM, KNN and Random Forest calculations utilising the Majority Voting calculation to foresee financial exchange patterns. The work likewise proposes a backpropagation enhanced Long Short Term Memory model for the financial exchange

expectation (Chen and Ge 2021). The two models factor a given company‟s verifiable stock value information just as the news influencing the organisation stocks information to anticipate the company‟s future stock patterns precisely and effectively. The two models utilise a Sentiment Analyzer to do the feeling examination on the organisation news (Ko and Chang 2021).

The research gap is identified as the SVM algorithm has less accuracy in stock price prediction . Determination stock price detection is shown at a very low percentage while analysing and manual input is not possible to add to the dataset. The classification of stock price detection in the existing model is less accurate due to selection of minimal functions. The aim of research work is to improve accuracy of predicting the stock price , and to reduce loss of data while training and testing dataset. The novel support vector machine is used to achieve accuracy.

2. Materials and Methods

The study setting of the proposed work is done in Saveetha School of Engineering. Two groups were identified for the study setting where group one LSTM and group two SVM. Using G power 10 samples sizes and totally 20 sample sizes have been carried out for our study, 95% confidence and pretest power 80%

The dataset named ‘DATA’ is downloaded from the public domain keras library. In our experiments here we used the data.csv dataset. Detailed descriptions of the features/attributes in the dataset can be found. The dataset consists of 5 lakhs instances. Dataset has two columns: url and label. The dataset was splitted into two parts namely the training part and testing part . 70% of the data was used for training and the remaining 30% was used for testing. The algorithm was implemented by evaluating the train and test. Input dataset collected from the link (tesla.csv))

Lstm Algorithm

Rolling LSTM Inputs: Time series Outputs: RMSE of the forecasted data
Split data into:
70\% training and 30\% testing data
1. size ← length(series) * 0.70
2. train ← series[0...size]
3. test ← series[size...length(size)] # Set the random seed to a fixed value
4. set random.seed(7)
Fit an LSTM model to training data Procedure fit_lstm(train, epoch, neurons) 5.
X ← train
6. y ← train - X
7. model = Sequential()
8. model.add(LSTM(neurons), stateful=True))
9. model.compile(loss=’mean_squared_error’, optimizer=’adam’)
10.for each i in range(epoch) do
11. model.fit(X, y, epochs=1, shuffle=False)
12. model.reset_states()
13.end for return model
Make a one-step forecast Procedure forecast_lstm(model, X)
14. yhat ← model.predict(X) return that
15. epoch ← 1
16. neurons ← 4
17. predictions ← empty # Fit the lstm model
18. lstm_model = fit_lstm(train,epoch,neurons) # Forecast the training dataset
19. lstm_model.predict(train) # Walk-forward validation on the test data
20. for each i in range(length(test)) do
21. # make one-step forecast
22. X ← test[i]
23. yhat ← forecast_lstm(lstm_model, X)
24. # record forecast
25. predictions.append(yhat)
26. expected ← test[i]
27. end for
28. MSE ← mean_squared_error(expected, predictions)
29. Return (RMSE ← sqrt(MSE))

Support Vector Machine (Svm) Algorithm

Support vector machine is a deep learning algorithm which is used for regression problems and classification problems as support vector classification (SVC) and support vector regression (SVR) (Zouina and Outtaj 2017). It takes too long to process so use the small dataset. SVM is a deep learning technique that works well for identifying speech recognizing and image recognizing.Accuracy of Phishing website detection using SVM algorithm(Accuracy = 97.32) is shown in Table 3 and Pseudo code shown in Table 7.

Equation (1) shows the accuracy measurement.

$$Accuracy = ((TP+TN)/(TP+TN+FP+FN)) \quad ----------(1)$$

Where,

TP - is the true positive classified by the model.
FP - is the no.of false positive classified by the model.
TN - is the no.of true negative classified by the model.
FN - is the no.of false negative classified by the model.

The software tool to evaluate SVM and LR algorithms is colab in Python programming language. The hardware configuration includes an intel i5 processor with a RAM size of 8GB. The

system used was a 64-bit windows 10 operating system

Statistical Analysis

For statistical implementation, the software to be used here is IBM SPSS V26.0. Statistical package for social sciences is used for calculating the statistical calculations such as mean, standard deviation, and also to plot the graphs etc.,. The independent variables are Url, Label and the dependent variable is 'accuracy'. In SPSS, the dataset is prepared using 10 as sample size for each group and accuracy is given as the testing variable.and the dependent variable is tesla stock samples. An independent T-Test analysis was performed

3. Results

For LSTM and support vector machine (SVM) both algorithms with their accuracy rate. For both proposed and existing algorithms 10 iterations were taken for each iteration the predicted accuracy was noted for analysing accuracy. The results of statistical packages of social sciences (IBM-SPSS v21) used for data analysis. With value obtained from the iterations Independent Sample T-test was performed . Significance values and group statistics values of proposed and existing algorithms are shown in Table 3. Whereas t-test equality is calculated. Confidence interval of the difference as lower and upper values range as shown in Table 4.
The bar graph is plotted by selected mean accuracy on Y-axis and the Group on X-axis. From the graph, it is clear that LSTM has significantly higher accuracy than SVM. The error bars are shown in the graph and the error rate is less for linear regression compared to LSTM

4. Discussion

In this study, the LSTM algorithm has better significant stock price prediction accuracy than SVM algorithm ($p<0.001$, Independent sample t-test). The improved accuracy and reduced loss for LSTM (Accuracy = 96.59%, Loss = 3.41%) than SVM (accuracy = 85.98%, Loss = 14.2%). Over the years, many models and algorithms have been employed to predict stock market data. In the field of Machine Learning (ML), the most popular algorithm used is the Support Vector Machine (SVM), the SVM has been widely used to predict stock market trends since it is noise-tolerant and gives a decent accuracy. Another ML algorithm used is the K-Nearest Neighbour (KNN), KNN has been used to predict stock market data to some extent but not with great accuracy. In the field of Deep Learning (DL), the most popular algorithms that have been used are the Artificial Neural Networks (ANN), ANNs have been conventionally used to predict stock market trends but haven‟t proved to be very efficient due to their limitation of data-overfitting.
The other very popular DL algorithm used is the Recurrent Neural Network (RNN), RNN has been utilised to predict stock market data and gives descent to impressive accuracy depending on the number of epochs. The RNN hasn‟t proved to be very accurate in predicting complex stock market trends due to their Vanishing Gradient problem(Sunny, Maswood, and Alharbi 2020). Other approaches attempted at predicting stock market data are based on the Sentiment Analysis of the news related to a company(Sunny, Maswood, and Alharbi 2020; Wei 2019). By determining the sentiment behind a company news, we can determine how a particular event or news will affect the stock price of a company. The Sentiment Analysis of a company news is implemented by doing the sentiment analysis of the Twitter news related to a company‟s stock. Several comparative studies have been done to compare the various individual ML and DL algorithms(Liu, Wang, and Zheng 2019).
Support Vector Machine (SVM) is a classification-cum-regression Supervised Machine Learning (ML) algorithm that can be employed for both classification and prediction problems. The algorithm takes every individual data entry as a point in a multi-dimensional space and the coordinates of the particle are its value. SVM algorithms employ mathematical functions called kernel functions for classification and prediction. The function of the kernel is to take input data and transform it into suitable classified form.The Majority Voting algorithm combines various predictive or regressive models to give better results and higher accuracy(Zeng and Liu 2018). The first step is to create multiple classification or predictive models for different subsets of the same data. Run the various models and get a prediction output from each. Store the result of each model in a matrix or a multidimensional array. The result or prediction of every model is counted as a vote. Then the matrix is traversed and the instances of each prediction are counted. The final output is the prediction that received more than half of the total votes(Sunil 2021).
LSTMs are a part of the Recurrent Neural Networks but help overcome the Vanishing Gradient problem encountered in the conventional RNN. Similar to RNN, we have time steps in LSTM but there is an extra feature called "MEMORY" in LSTM for every time step. The LSTM network consists of different memory blocks called cells. Two states are given as input from the previous cell to the next cell: the cell state and the hidden state (Bathla 2020). Three gates are used for managing information viz. Forget Gate, Input Gate and Output Gate. A forget gate removes

unwanted information from the cell state. The input gate appends new information to the cell state. It hence adds new information using the sigmoid function as well as creates a vector of all the possible values suitable for addition to the current cell state.

5. Conclusion

In this research work, The results indicate that our proposed LSTM based model by using a support vector machine to detect previously unseen stock price with improved accuracy of 97%. In future, the performance can be improved by adopting some more advanced classification models as well as suitable optimised features can also be selected by using some optimization techniques.

Declaration

Conflict Of Interest

The authors do not have any conflict of interest associated with this manuscript

Author Contribution

Author RRK is involved in data collection, data analysis, manuscript, writing. Author SKM involved in conceptualization, data validation, and critical review of manuscript

Acknowledgment

The authors of this work would like to express their gratitude to the Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Saveetha University for providing the opportunities and facilities needed to carry out research studies.

Funding

We thank the following organisation for providing financial support that enabled us to complete the study.

1. Insysness Technology Pvt.Ltd.Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

6. References

Ali, F. 2021. "Peer Review #2 of 'LSTM-Based Sentiment Analysis for Stock Price Forecast (v0.1).'" https://doi.org/10.7287/peerj-cs.408v0.1/reviews/2.

Ayan, Ebubekir, and Süleyman Eken. 2021. "Detection of Price Bubbles in Istanbul Housing Market Using LSTM Autoencoders: A District-Based Approach." *Soft Computing*, March, 1–17.

Bathla, Gourav. 2020. "Stock Price Prediction Using LSTM and SVR." *2020 Sixth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. https://doi.org/10.1109/pdgc50313.2020.9315800.

Chen, Shun, and Lei Ge. 2021. "Exploring the Attention Mechanism in LSTM-Based Hong Kong Stock Price Movement Prediction." *Machine Learning and AI in Finance*. https://doi.org/10.4324/9781003145714-7.

Jayabal, Ravikumar, Sekar Subramani, Damodharan Dillikannan, Yuvarajan Devarajan, Lakshmanan Thangavelu, Mukilarasan Nedunchezhiyan, Gopal Kaliyaperumal, and Melvin Victor De Poures. 2022. "Multi-Objective Optimization of Performance and Emission Characteristics of a CRDI Diesel Engine Fueled with Sapota Methyl Ester/diesel Blends." *Energy*. https://doi.org/10.1016/j.energy.2022.123709.

Jayakodi, Santhoshkumar, Rajeshkumar Shanmugam, Bader O. Almutairi, Mikhlid H. Almutairi, Shahid Mahboob, M. R. Kavipriya, Ramesh Gandusekar, Marcello Nicoletti, and Marimuthu Govindarajan. 2022. "Azadirachta Indica-Wrapped Copper Oxide Nanoparticles as a Novel Functional Material in Cardiomyocyte Cells: An Ecotoxicity Assessment on the Embryonic Development of Danio Rerio." *Environmental Research* 212 (Pt A): 113153.

Jia, Mingzhu, Jian Huang, Lihua Pang, and Qian Zhao. 2019. "Analysis and Research on Stock Price of LSTM and Bidirectional LSTM Neural Network." *Proceedings of the 3rd International Conference on Computer Engineering, Information Science & Application Technology (ICCIA 2019)*. https://doi.org/10.2991/iccia-19.2019.72.

Ko, Ching-Ru, and Hsien-Tsung Chang. 2021. "LSTM-Based Sentiment Analysis for Stock Price Forecast." *PeerJ Computer Science*. https://doi.org/10.7717/peerj-cs.408.

Kotteeswaran, C., Indrajit Patra, Regonda Nagaraju, D. Sungeetha, Bapayya Naidu Kommula, Yousef Methkal Abd Algani, S. Murugavalli, and B. Kiran Bala. 2022. "Autonomous Detection of Malevolent Nodes Using Secure Heterogeneous Cluster Protocol." *Computers and Electrical Engineering*. https://doi.org/10.1016/j.compeleceng.2022.107902.

Krishnan, Anbarasu, Duraisami Dhamodharan, Thanigaivel Sundaram, Vickram Sundaram, and Hun-Soo Byun. 2022. "Computational Discovery of Novel Human LMTK3 Inhibitors by High Throughput Virtual Screening Using NCI Database." *Korean Journal of Chemical Engineering*. https://doi.org/10.1007/s11814-022-1120-5.

Liu, Yiwei, Zhiping Wang, and Baoyou Zheng. 2019. "Application of Regularized GRU-LSTM Model in Stock Price Prediction." *2019 IEEE 5th International Conference on Computer and Communications (ICCC)*. https://doi.org/10.1109/iccc47050.2019.9064035.

Mohan, Harshavardhan, Sethumathavan Vadivel, Se-Won Lee, Jeong-Muk Lim, Nanh Lovanh, Yool-Jin Park, Taeho Shin, Kamala-Kannan Seralathan, and Byung-Taek Oh. 2022. "Improved Visible-Light-Driven Photocatalytic Removal of Bisphenol A Using V2O5/WO3 Decorated over Zeolite: Degradation Mechanism and Toxicity." *Environmental Research*. https://doi.org/10.1016/j.envres.2022.113136.

Mohan, Kannan, Abirami Ramu Ganesan, P. N. Ezhilarasi, Kiran Kumar Kondamareddy, Durairaj Karthick Rajan, Palanivel Sathishkumar, Jayakumar Rajarajeswaran, and Lorenza Conterno. 2022. "Green and Eco-Friendly Approaches for the Extraction of Chitin and Chitosan: A Review." *Carbohydrate Polymers* 287 (July): 119349.

Roy, Ranjan Kumar, Koyel Ghosh, and Apurbalal Senapati. 2021. "Stock Price Prediction: LSTM Based Model." *Proceedings of Intelligent Computing and Technologies Conference*. https://doi.org/10.21467/proceedings.115.19.

Saravanan, A., P. Senthil Kumar, B. Ramesh, and S. Srinivasan. 2022. "Removal of Toxic Heavy Metals Using Genetically Engineered Microbes: Molecular Tools, Risk Assessment and Management Strategies." *Chemosphere* 298 (July): 134341.

Sathish, T., R. Saravanan, V. Vijayan, and S. Dinesh Kumar. 2022. "Investigations on Influences of MWCNT Composite Membranes in Oil Refineries Waste Water Treatment with Taguchi Route." *Chemosphere* 298 (July): 134265.

Sunil, Anjali. 2021. "Stock Price Prediction Using LSTM Model and Dash." *International Journal for Research in Applied Science and Engineering Technology*. https://doi.org/10.22214/ijraset.2021.32760.

Sunny, Md Arif Istiake, Mirza Mohd Shahriar Maswood, and Abdullah G. Alharbi. 2020. "Deep Learning-Based Stock Price Prediction Using LSTM and Bi-Directional LSTM Model." *2020 2nd Novel Intelligent and Leading Emerging Sciences Conference (NILES)*. https://doi.org/10.1109/niles50944.2020.9257950.

Vivek, J., T. Maridurai, K. Anton Savio Lewise, R. Pandiyarajan, and K. Chandrasekaran. 2022. "Recast Layer Thickness and Residual Stress Analysis for EDD AA8011/h-BN/B4C Composites Using Cryogenically Treated SiC and CFRP Powder-Added Kerosene." *Arabian Journal for Science and Engineering*. https://doi.org/10.1007/s13369-022-06636-5.

Wei, Dou. 2019. "Prediction of Stock Price Based on LSTM Neural Network." *2019 International Conference on Artificial Intelligence and Advanced Manufacturing (AIAM)*. https://doi.org/10.1109/aiam48774.2019.00113.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Algal Biofuels: Technological Perspective on Cultivation, Fuel Extraction and Engineering Genetic Pathway for Enhancing Productivity." *Fuel*. https://doi.org/10.1016/j.fuel.2022.123814.

Yaashikaa, P. R., P. Senthil Kumar, and S. Karishma. 2022. "Review on Biopolymers and Composites – Evolving Material as Adsorbents in Removal of Environmental Pollutants." *Environmental Research*. https://doi.org/10.1016/j.envres.2022.113114.

Yan, Yangtian. 2021. "Prediction of Stock Price Based on LSTM Model." *Proceedings of the 6th International Conference on Financial Innovation and Economic Development (ICFIED 2021)*. https://doi.org/10.2991/aebmr.k.210319.037.

Zeng, Yawen, and Xing Liu. 2018. "A-Stock Price Fluctuation Forecast Model Based on LSTM." *2018 14th International Conference on Semantics, Knowledge and Grids (SKG)*. https://doi.org/10.1109/skg.2018.00044.

TABLES AND fIGURES

Table 1.accuracy of stock price prediction using LSTM algorithm

Test size	**Accuracy**
Test 1	96.59
Test 2	96.13

Test 3	96.10
Test 4	96.07
Test 5	95.98
Test 6	95.93
Test 7	93.91
Test 8	97.87
Test 9	95.81
Test 10	95.79

Table 2. Accuracy of stock price detection using SVM algorithm

Test Size	**Accuracy**
Test 1	85.98
Test 2	85.87
Test 3	85.76
Test 4	86.65
Test 5	85.61
Test 6	87.98
Test 7	84.87
Test 8	84.76
Test 9	84.54
Test 10	84.32

Table 3. Group statistics results (Mean of LSTM 95.7580 is more Compared with SVM 85.6140 and Std.Error Mean for LSTM is .31511 and SVM is .44473)

Groups	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
LSTM	10	95.7580	0.99648	0.31511
SVM	10	85.6140	1.40635	0.44473

Table 4. T-test with Independent Samples The result is calculated with a 95% confidence interval and a significance threshold of 0.05 (the support vector machine algorithm looks to perform significantly better than the artificial neural network with a value of (p = 0.08).

	Equal Variance	**Levene's Test for Equality of Variance**	**T-test for Equality of Means**

		F	Sig	t	df	Sig(2-tailed)	Mean Difference	Std.Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Accuracy	Equal variances Assumed	3.386	0.08	18.61	18	<.001	10.144	0.545	8.9988	11.289
	Equal variances Not Assumed			18.61	16.218	< .001	10.144	0.545	8.9898	11.298

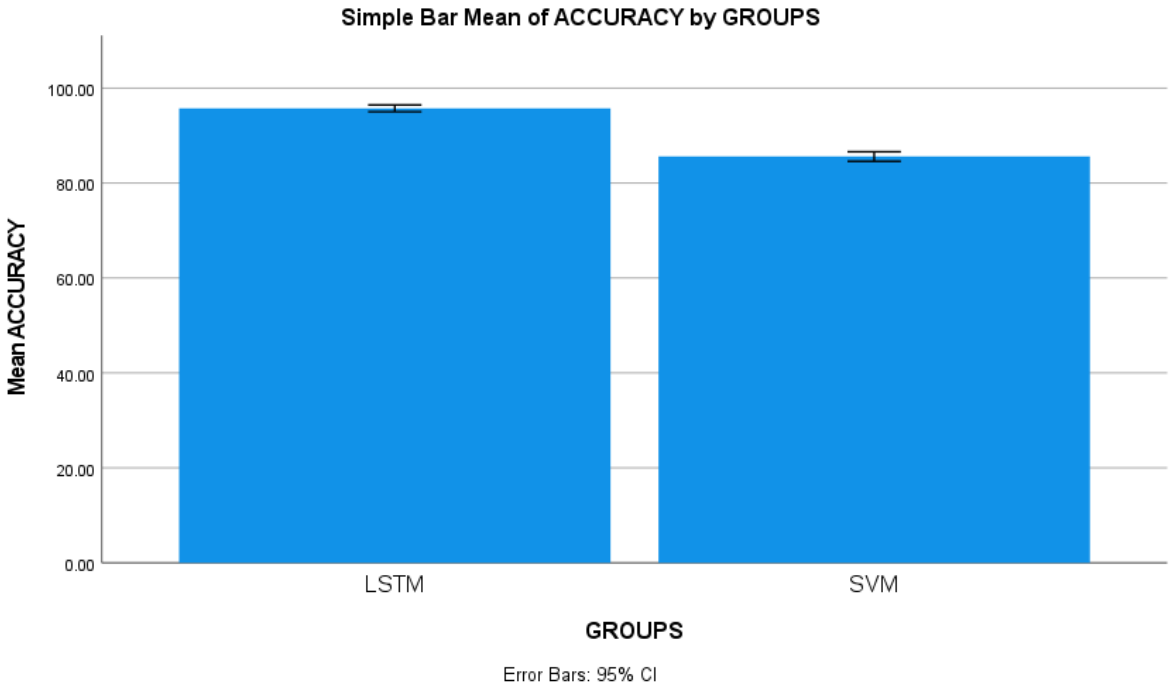

Fig. 1. Clustered Bar mean of accuracy , mean of loss by LSTM & SVM classifier in terms of mean accuracy. The mean accuracy of LSTM is better than SVM and standard deviation of LSTM is slightly better than SVM. X-axis : LSTM vs SVM algorithm Y-axis :