# IMPROVED ACCURACY IN STOCK PRICE PREDICTION SYSTEM USING A NOVEL DECISION TREE ALGORITHM COMPARED TO LINEAR DISCRIMINANT ANALYSIS (LDA) ALGORITHM

**Juhaina[1], Terrance Frederick Fernandez[2*]**

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

**Abstract**

**Aim:**The Main target of this work is comparative study of Novel Decision Tree Algorithm and Linear Discriminant Analysis(LDA) Algorithm for optimizing Stock price prediction to improve the Accuracy of Stock Exchange.

**Materials and Methods:** Novel Decision Tree Algorithm (N=10) and Linear Discriminant Analysis Algorithm (N=10) are simulated by varying the Novel Decision Tree parameter and Linear Discriminant Analysis parameter to optimize the pH. Sample size is calculated using Gpower 80% for two groups and there are 20 samples used in this work.

**Results:**The examination of Accuracy rate is finished by independent Sample size utilizing SPSS Software. The Linear Discriminant Analysis Algorithm produces 16.91% Accuracy though Decision Tree algorithm produces 87.29% Accuracy. The Statistical significance difference between Decision Tree and LDA was found to be 0.063($p<0.05$).

**Conclusion:** The Outcome shows that the performance of Decision Tree is better than the performance of Linear Discrimination in the terms of Accuracy.

**Keywords:** Machine Learning, Stock price prediction, Novel Decision Tree Algorithm, Linear Discriminant Analysis Algorithm, Accuracy, Analysis.

[1]Research Scholar, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University,Chennai,Tamilnadu,India,Pincode: 602105.

[2*]Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical And Technical Sciences, Saveetha University,Chennai,Tamilnadu,India,Pincode: 602105.

## 1. Introduction

Estimating stock market values is both an interesting and demanding field of study. Forecasting stock market prices is always a difficulty for many company experts and researchers. Due to the Volatile nature and non-linear structure of the financial stock market, accurately predicting stock market returns is a difficult endeavor(Saud and Shakya 2020). The Machine Learning model had a significant impact in the following research work on Stock Price Prediction, which generates forecasts based on the present stock market's values by training on their prior values(Vijh et al. 2020). While proponents of the efficient market hypothesis feel it is impossible to effectively predict stock prices, formal propositions show that accurate modeling and selection of appropriate variables can lead to models that can reliably predict stock prices and stock price movement patterns.(Mehtab, Sen, and Dutta 2021). Predicting the stock market with 100% accuracy is extremely difficult due to extrinsic factors such as social, political, psychological, and economic influences(Hogenboom, Brojba-Micu, and Frasincar 2021). The primary function of stock price prediction is to serve as a trading platform where various investors can sell and buy shares in accordance with the stock's availability. Predictive analytics or predictive modeling are other terms for machine learning. Machine Learning is based on the premise of using increasing data to discover the best model for new data among past data.(Çelik 2018). These predictions will aid in future stock prediction. It has always been a hotspot for buyers and investors looking to understand the fluctuating regularity of the stock market price and predict its trend(Agrawal et al. 2022).

In the last 5 years, more than 65 papers have been published on IEEE xplore and google scholar on Stock price predictions which can be greatly invested in stocks for big companies(Pang et al. 2020).A near gets anticipating the speculation can lead towards benefit. In this article investigation of Novel Decision Tree Algorithm (Tharwat et al. 2017; Navada et al. 2011) and Linear Discriminant Analysis Algorithm (Tharwat et al. 2017) in superior execution effectiveness has been made utilizing an exploratory methodology. A comparison gets the job done of estimating whether or not an investment will yield a return. This particular article presents the comparative analysis between the accuracy control of Novel Decision Tree Algorithm using conventional controllers like stock prediction controllers and Stock Price Controller (SPC). A novel method for Decision Tree using Linear Discriminant Analysis has been proposed in this Article for the Accuracy improvement(Nti, Adekoya, and Weyori 2021). Our team has extensive knowledge and research experience that has translated into high quality publications(Pandiyan et al. 2022; Yaashikaa, Devi, and Kumar 2022; Venu et al. 2022; Kumar et al. 2022; Nagaraju et al. 2022; Karpagam et al. 2022; Baraneedharan et al. 2022; Whangchai et al. 2022; Nagarajan et al. 2022; Deena et al. 2022)

In a past report the Accuracy improvement of the Naive Bayes(A.Jabbar Alkubaisi, Kamaruddin, and Husni 2018) calculation with stock value forecast was not as expected carried out to further develop the log misfortune pace of the stock value prediction. To beat this issue a clever long transient memory calculation is carried out to further develop log misfortune pace of financial exchange expectation (Ampomah et al. 2021). The aim of this study is to get a better accuracy rate for Stock Price Prediction using the Decision Tree Algorithm.

## 2. Materials and Methods

The following Research work is performed in the Department of Computer Science and Engineering, Saveetha School of Engineering, SIMATS, Chennai. The work is carried out on 1236 records taken from a Stock dataset. The Accuracy in predicting the stock price was performed by evaluating two groups. A total of 10 iterations was performed on each group to achieve better accuracy. The dataset was downloaded from Kaggle website.The dataset contains 1236 rows and 8 columns. Some of the important attributes taken for experimental setup are Total trade, Turnover, Loss etc.

The example size has been determined utilizing the GPower programming by looking at both of the regulators in Supervised Learning. The example size was determined as 600 in each gathering utilizing G power. Test size was determined utilizing clinical examination, with alpha and beta qualities 0.05 and 0.5, separately. 95% certainty stretches and pretest power 80% and enrolment proportion 1 were utilized. Two calculations ( Novel Decision Tree Algorithm and Linear Discriminant Algorithm) are executed utilizing Technical Analysis programming. In this work, no human and creature tests were utilized so there is no requirement for any moral endorsement expected for this examination work.

Two methods that are widely used in general are namely Fundamental Analysis and Technical Analysis

Fundamental Analysis: It is vital to have competitive analysis strength and economic conditions in which they are interested in order to identify the accurate product value, dependable and accurate information on the financial report of the company. Fundamental analysis is important for long-term predictions, and the advantages are due to their systematic approval and ability to predict changes.

Technical Analysis: Technical analysis is based on the premise that investors constantly change quantities in reaction to various forces and factors, resulting in stock price trends and movements. Trend indicators, the lowest and highest daily values, daily ups and downs, stock volume, indices, and other technical variables of quantitative characteristics can all be employed for analysis. It is feasible to derive rules from data, and investors can use these rules to make future judgments. As an input to the system, technical analysis data is preferred above fundamental analysis data.

**Decision Tree Algorithm**

Decision Tree is a supervised learning technique that may be used to solve both classification and regression problems, however it is most commonly employed to solve classification issues. Internal nodes represent dataset attributes, branches represent decision rules, and each leaf node provides the conclusion in this tree-structured classifier. It's a graphical representation for obtaining all feasible solutions to a problem/decision depending on certain parameters. In statistics, data mining, and machine learning, a decision tree is one of the predictive modeling approaches. In decision analysis, a decision tree can be used to visually and explicitly represent decisions and decision making.

Implementation of Decision Tree Algorithm:

Input: Training and Testing data
Output:Accuracy Score
Step 1. From sklearn import Decision Tree
Step 2. Import dataset
Step 3. Preprocess the data
Step 4. x,y(Define the features needed)
Step 5. Split the dataset into Training and Testing sets
Step 6. Fit the Decision Tree algorithm on the training and testing dataset
Step 7. Train the Decision Tree algorithm
Step 8. Predict the result
Step 9. Calculate the accuracy score.

**Linear Discriminant Analysis Algorithm**

LDA (Linear Discriminant Analysis) is a technique for reducing dimensionality. It's utilized in Machine Learning and pattern categorization applications as a pre-processing phase. The purpose of LDA is to project features from a higher-dimensional space onto a lower-dimensional space, avoiding the dimensionality curse while simultaneously saving resources and reducing dimensional costs.

Implementation of LDA:

Input: Training and Testing data
Output:Accuracy Score
Step 1. From sklearn import LDA
Step 2. Import dataset
Step 3. Preprocess the data
Step 4. x,y(Define the features needed)
Step 5. Split the dataset into Training and Testing sets
Step 6. Fit the LDA algorithm on the training and testing dataset
Step 7. Train the LDA algorithm
Step 8. Predict the result
Step 9. Calculate the accuracy score.

The hardware configuration was an Intel core i5 processor with a RAM size of 8GB. The system type used was a 64bit OS, X64 based processor with SDD of 256GB. The operating system used was windows 10 and the tool used for implementation was google colab© with python programming language.

**Statistical Analysis**

For statistical analysis of Decision Tree Algorithm and LDA Algorithm based approaches, SPSS software is employed. The dependent variable is efficiency, while the independent variable is Decision Tree accuracy. The accuracy of the Decision is calculated using separate T test analysis for both techniques.

## 3. Results

Table 1 shows the simulation result of the proposed Linear Discriminant Analysis Algorithm and the existing system Novel Decision Tree Algorithm were run at different times in the google colab with a sample size of 10. From table 1, it was observed that the mean accuracy of the Decision Tree Algorithm was 87.29% and the Linear Discriminant Analysis Algorithm was 16.91%.

Table 2 represents the T-test comparison of both Decision Tree algorithm and LDA algorithm. The Mean, Standard Deviation and Standard Error Mean were calculated by taking an independent variable T test among the study groups. The Decision Tree algorithm produces a significant difference than the LDA algorithm with a value of 0.063 .

Table 3 represents the Mean of Decision Tree algorithm which is better compared with LDA algorithm with a standard deviation of Decision Tree is 1.21979 and LDA algorithm is 0.61742 respectively. From the results, the Decision Tree algorithm (87.29%) gives better accuracy than the Linear Discriminant Analysis algorithm (16.91%). Figure 1 gives the comparison chart of the Decision Tree of Linear Discriminant Analysis algorithms in terms of mean and accuracy. The mean accuracy of the Decision Tree algorithm is better than Linear Discriminant Analysis. Figure 2 shows the error mean of Decision Tree algorithm (0.38) and Linear Discriminant Analysis algorithm (0.20).

## 4. Discussions

Decision Tree and Linear Discriminant Analysis algorithms are implemented and compared for stock market prediction to improve the accuracy by Stock price prediction. From obtained results it is concluded that the Decision Tree algorithm provides better accuracy results compared to the Linear Discriminant Analysis algorithm.

In this paper,(Nair, Dharini, and Mohandas 2010) an automated decision tree-adaptive neuro-fuzzy hybrid automated stock market trend prediction framework is proposed. The proposed framework utilizes specialized examination (generally utilized by stock merchants) to include extraction and decision tree for highlight determination.The proposed framework is tried on four significant worldwide financial exchanges. The outcomes show that the proposed mixture framework creates a lot higher accuracy when contrasted with independent decision tree based framework and ANFIS based framework without include determination and dimensionality decrease.

In this paper, they(Gurrib, Kamalov, and Smail 2021) propose another technique for foreseeing the course of bitcoin value utilizing linear discriminant analysis (LDA) . Solidly, they train a LDA-based classifier that utilizes the current bitcoin value data and Twitter feature news to gauge the following day course of bitcoin cost. Specifically, the proposed approach produces estimated accuracy of 0.828 and AUC of 0.840 on the test information. The proposed arrangement of this paper(Gurrib, Kamalov, and Smail 2021; Karim, Alam, and Hossain 2021) works in two strategies - Linear Regression and Decision Tree Regression. Two models like Linear Regression and Decision Tree Regression are applied for various sizes of a dataset for uncovering the stock value figure expectation precision.The outcome uncovers that Linear Regression is given better precision for both little and large datasets. Then again, Decision Tree Regression communicates the helpless forecast cost in view of the size of the dataset.

From the above conversation, a couple of articles guarantee that they give preferred execution over the proposed Decision tree and Linear discriminant analysis algorithm for further developing accuracy of stock price prediction. Likewise, the current value forecast requires no extra expense and subsequently got serious consideration as of late. In this way, we can construe that the proposed Decision tree and Linear discriminant algorithm can be utilized to work on the exactness of value forecast by managing the stock addition.

Stock market prediction has limited price prediction ability based on future price significant profit which makes better price prediction in future. Deep Learning and Machine Learning algorithms can address future stock prediction.

## 5. Conclusion

The work involves the Novel Decision Tree algorithm that is aimed to find the Stock Price Prediction. It is proved with better accuracy of 87.29% when compared to Linear Discriminant Analysis accuracy is 16.91% for predicting Stock price.

## DECLARATIONS

**Conflict of Interests**

No conflict of Interest in this manuscript.

**Authors Contributions**

Author JU was involved in data collection, data analysis and manuscript writing. Author TFF was involved in the conceptualization, data validation and critical review of manuscript.

**Acknowledgements**

The authors would like to express their gratitude towards Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (Formerly known as Saveetha University) for providing the necessary infrastructure to carry out this work successfully.

**Funding**

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

1. Railsdata software Pvt.Ltd
2. Saveetha University
3. Saveetha Institute of Medical And Technical Sciences
4. Saveetha School of Engineering

## 6. References

Agrawal, Manish, Piyush Kumar Shukla, Rajit Nair, Anand Nayyar, and Mehedi Masud. 2022. "Stock Prediction Based on Technical Indicators Using Deep Learning Model."

Computers, Materials & Continua 70 (1): 287–304.

A.Jabbar Alkubaisi, Ghaith Abdulsattar, Siti Sakira Kamaruddin, and Husniza Husni. 2018. "Stock Market Classification Model Using Sentiment Analysis on Twitter Based on Hybrid Naive Bayes Classifiers." Computer and Information Science 11 (1): 52.

Ampomah, Ernest Kwame, Gabriel Nyame, Zhiguang Qin, Prince Clement Addo, Enoch Opanin Gyamfi, and Micheal Gyan. 2021. "Stock Market Prediction with Gaussian Naïve Bayes Machine Learning Algorithm." Informatica. An International Journal of Computing and Informatics 45 (2). https://doi.org/10.31449/inf.v45i2.3407.

Baraneedharan, P., Sethumathavan Vadivel, C. A. Anil, S. Beer Mohamed, and Saravanan Rajendran. 2022. "Advances in Preparation, Mechanism and Applications of Various Carbon Materials in Environmental Applications: A Review." Chemosphere. https://doi.org/10.1016/j.chemosphere.2022.134596.

Çelik, Özer. 2018. "A Research on Machine Learning Methods and Its Applications." Journal of Educational Technology and Online Learning, September. https://doi.org/10.31681/jetol.457046.

Deena, Santhana Raj, A. S. Vickram, S. Manikandan, R. Subbaiya, N. Karmegam, Balasubramani Ravindran, Soon Woong Chang, and Mukesh Kumar Awasthi. 2022. "Enhanced Biogas Production from Food Waste and Activated Sludge Using Advanced Techniques – A Review." Bioresource Technology. https://doi.org/10.1016/j.biortech.2022.127234.

Gurrib, Ikhlaas, Firuz Kamalov, and Linda Smail. 2021. "Bitcoin Price Forecasting: Linear Discriminant Analysis with Sentiment Evaluation." In ArabWIC 2021: The 7th Annual International Conference on Arab Women in Computing in Conjunction with the 2nd Forum of Women in Research, Sharjah, UAE. New York, NY, USA: ACM. https://doi.org/10.1145/3485557.3485561.

Hogenboom, Alexander, Alex Brojba-Micu, and Flavius Frasincar. 2021. "The Impact of Word Sense Disambiguation on Stock Price Prediction." Expert Systems with Applications 184 (115568): 115568.

Karim, Rezaul, Md Khorshed Alam, and Md Rezaul Hossain. 2021. "Stock Market Analysis Using Linear Regression and Decision Tree Regression." In 2021 1st International Conference on Emerging Smart Technologies and Applications (eSmarTA). IEEE. https://doi.org/10.1109/esmarta52612.2021.9515762.

Karpagam, M., R. Beaulah Jeyavathana, Sathiya Kumar Chinnappan, K. V. Kanimozhi, and M. Sambath. 2022. "A Novel Face Recognition Model for Fighting against Human Trafficking in Surveillance Videos and Rescuing Victims." Soft Computing. https://doi.org/10.1007/s00500-022-06931-1.

Kumar, P. Ganesh, P. Ganesh Kumar, Rajendran Prabakaran, D. Sakthivadivel, P. Somasundaram, V. S. Vigneswaran, and Sung Chul Kim. 2022. "Ultrasonication Time Optimization for Multi-Walled Carbon Nanotube Based Therminol-55 Nanofluid: An Experimental Investigation." Journal of Thermal Analysis and Calorimetry. https://doi.org/10.1007/s10973-022-11298-4.

Mehtab, Sidra, Jaydip Sen, and Abhishek Dutta. 2021. "Stock Price Prediction Using Machine Learning and LSTM-Based Deep Learning Models." In Communications in Computer and Information Science, 88–106. Communications in Computer and Information Science. Singapore: Springer Singapore.

Nagarajan, Karthik, Arul Rajagopalan, S. Angalaeswari, L. Natrayan, and Wubishet Degife Mammo. 2022. "Combined Economic Emission Dispatch of Microgrid with the Incorporation of Renewable Energy Sources Using Improved Mayfly Optimization Algorithm." Computational Intelligence and Neuroscience 2022 (April): 6461690.

Nagaraju, V., B. R. Tapas Bapu, P. Bhuvaneswari, R. Anita, P. G. Kuppusamy, and S. Usha. 2022. "Role of Silicon Carbide Nanoparticle on Electromagnetic Interference Shielding Behavior of Carbon Fibre Epoxy Nanocomposites in 3-18GHz Frequency Bands." Silicon. https://doi.org/10.1007/s12633-022-01825-1.

Nair, Binoy B., N. Mohana Dharini, and V. P. Mohandas. 2010. "A Stock Market Trend Prediction System Using a Hybrid Decision Tree-Neuro-Fuzzy System." In 2010 International Conference on Advances in Recent Technologies in Communication and Computing. IEEE. https://doi.org/10.1109/artcom.2010.75.

Navada, Arundhati, Aamir Nizam Ansari, Siddharth Patil, and Balwant A. Sonkamble. 2011. "Overview of Use of Decision Tree Algorithms in Machine Learning." In 2011 IEEE Control and System Graduate Research Colloquium. IEEE. https://doi.org/10.1109/icsgrc.2011.5991826.

Nti, Isaac Kofi, Adebayo Felix Adekoya, and

Benjamin Asubam Weyori. 2021. "A Novel Multi-Source Information-Fusion Predictive Framework Based on Deep Neural Networks for Accuracy Enhancement in Stock Market Prediction." Journal of Big Data 8 (1). https://doi.org/10.1186/s40537-020-00400-y.

Pandiyan, P., R. Sitharthan, S. Saravanan, Natarajan Prabaharan, M. Ramji Tiwari, T. Chinnadurai, T. Yuvaraj, and K. R. Devabalaji. 2022. "A Comprehensive Review of the Prospects for Rural Electrification Using Stand-Alone and Hybrid Energy Technologies." Sustainable Energy Technologies and Assessments. https://doi.org/10.1016/j.seta.2022.102155.

Pang, Xiongwen, Yanqiang Zhou, Pan Wang, Weiwei Lin, and Victor Chang. 2020. "An Innovative Neural Network Approach for Stock Market Prediction." The Journal of Supercomputing 76 (3): 2098–2118.

Saud, Arjun Singh, and Subarna Shakya. 2020. "Analysis of Look Back Period for Stock Price Prediction with RNN Variants: A Case Study on Banking Sector of NEPSE." Procedia Computer Science 167: 788–98.

Tharwat, Alaa, Tarek Gaber, Abdelhameed Ibrahim, and Aboul Ella Hassanien. 2017. "Linear Discriminant Analysis: A Detailed Tutorial." AI Communications. The European Journal on Artificial Intelligence 30 (2): 169–90.

Venu, Harish, Ibham Veza, Lokesh Selvam, Prabhu Appavu, V. Dhana Raju, Lingesan Subramani, and Jayashri N. Nair. 2022. "Analysis of Particle Size Diameter (PSD), Mass Fraction Burnt (MFB) and Particulate Number (PN) Emissions in a Diesel Engine Powered by Diesel/biodiesel/n-Amyl Alcohol Blends." Energy. https://doi.org/10.1016/j.energy.2022.123806.

Vijh, Mehar, Deeksha Chandola, Vinay Anand Tikkiwal, and Arun Kumar. 2020. "Stock Closing Price Prediction Using Machine Learning Techniques." Procedia Computer Science 167: 599–606.

Whangchai, Niwooti, Daovieng Yaibouathong, Pattranan Junluthin, Deepanraj Balakrishnan, Yuwalee Unpaprom, Rameshprabu Ramaraj, and Tipsukhon Pimpimol. 2022. "Effect of Biogas Sludge Meal Supplement in Feed on Growth Performance Molting Period and Production Cost of Giant Freshwater Prawn Culture." Chemosphere 301 (August): 134638.

Yaashikaa, P. R., M. Keerthana Devi, and P. Senthil Kumar. 2022. "Advances in the Application of Immobilized Enzyme for the Remediation of Hazardous Pollutant: A Review." Chemosphere 299 (July): 134390.

**Tables and Figures**

Table 1: Accuracy of the samples using Novel Decision tree algorithm and Linear Discriminant Analysis algorithm

| SAMPLES | NOVEL DECISION TREE ALGORITHM (ACCURACY) | LINEAR DISCRIMINANT ANALYSIS ALGORITHM (ACCURACY) |
|---|---|---|
| 1 | 87.23 | 16.00 |
| 2 | 88.00 | 16.23 |
| 3 | 86.32 | 17.00 |
| 4 | 85.34 | 16.85 |

| | | |
|---|---|---|
| 5 | 87.00 | 16.45 |
| 6 | 89.00 | 17.00 |
| 7 | 89.00 | 18.00 |
| 8 | 88.00 | 17.93 |
| 9 | 87.09 | 16.78 |
| 10 | 86.00 | 16.87 |

Table 2: Comparison of mean of Accuracy using Novel Decision tree and Linear Discriminant algorithms.

| **Matrics** | **Algorithm** | **No of samples** | **Mean** | **Std. Deviation** | **Std.Error Mean** |
|---|---|---|---|---|---|
| Accuracy | Decision tree algorithm | 10 | 87.2980 | 1.21979 | 0.38573 |
| | LDA | 10 | 16.9110 | 0.61742 | 0.20464 |

**Group statistics**

Group statistics comparison of accuracy for predecting Stock price using Decision tree Algorithm and Linear Discriminant algorithm is done. Novel Decision tree algorithm has higher mean compared to Linear Discriminant algorithm.

Decision tree algorithm = 87.2980

Linear Discriminant analysis Algorithm= 16.9110

Table 3: Independent sample T-test is performed for the two groups for significance and standard error determination p>0.05 for the test basis.

| | **Levene's Test for Equality of Variance** | | **t_test for Equality of Means** | | | | | **95%Confidence interval of interest** | |
|---|---|---|---|---|---|---|---|---|---|
| | F | sig. | t | df | sig.(2-) | Mean Differenc | std.Error Differenc | Lower | Upper |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | | | | tailed) | e | e | | |
| Accuracy | Equal Variance assumed | 3.932 | 0.063 | 161.197 | 18 | 0.000 | 70.38700 | 0.43665 | 69.46963 | 71.30437 |
| | Equal variance not assumed | | | 161.197 | 13.694 | 0.000 | 70.38700 | 0.43665 | 69.44851 | 71.32549 |

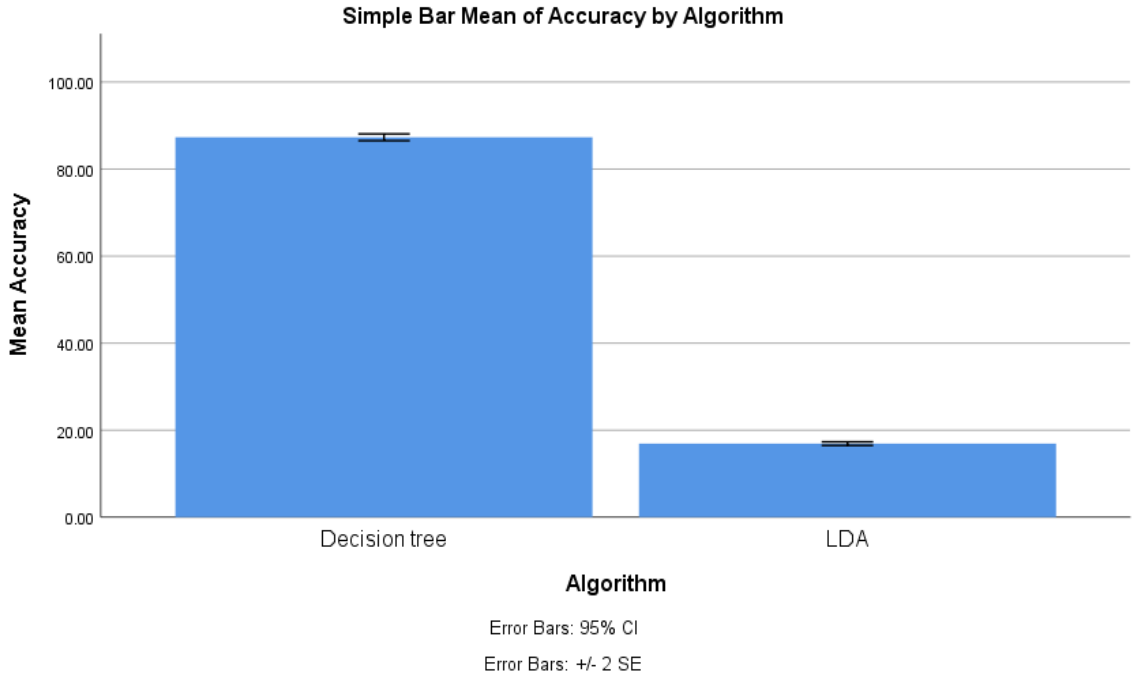


Fig-1: Sample mean of accuracy by using Decision tree algorithm and Linear Discriminant Analysis algorithm.