IMPROVED ACCURACY IN STOCK PRICE PREDICTION SYSTEM USING A NOVEL NAIVE BAYES ALGORITHM COMPARED TO K-NEAREST NEIGHBOR ALGORITHM (KNN)

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

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

Abstract

Aim: This following research compares the Novel Naive Bayes Algorithm and the K-Nearest Neighbor Algorithm for stock price prediction optimization in order to enhance the Accuracy of real-time stock exchange.
Materials and Methods: To optimize the pH, the Novel Naive Bayes Algorithm (N=10) and K-Nearest Neighbor (N=10) are simulated by adjusting the Novel Naive Bayes parameters and K-Nearest Neighbor parameters**.** Gpower 80 percent is utilized to compute sample size for two groups, and 20 samples are investigated in this research.
Results: Utilizing SPSS Software, an independent sample size is used to evaluate the accuracy rate. Although K-Nearest Neighbor generates 46.50 percent accuracy, Naive Bayes produces 82.19 percent accuracy. The difference in statistical significance between Naive Bayes and KNN was discovered to be 0.016 ($p<0.05$).
Conclusion: In terms of accuracy, the Naive Bayes algorithm outperforms the K-Nearest Neighbor algorithm.

Keywords: Machine learning, Stock price prediction, Novel Naive Bayes Algorithm, K-Nearest Neighbor 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

Forecasting stock market prices is always a difficulty for many company experts and researchers. Estimating stock market values is both an interesting and demanding field of study.(Ding and Qin 2020) Predicting the stock market with 100% accuracy is extremely difficult due to extrinsic factors such as social, political, psychological, and economic influences. 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(Hsieh, Hsiao, and Yeh 2011). 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(Huang, Nakamori, and Wang 2005). Predictive analytics or predictive modeling are other terms for machine learning. Predicting stock exchange trends is seen as a critical topic that requires special attention, as successfully predicting stock prices with supervised learning may result in generating money by making the right decisions.(M et al. 2018) These stock price predictions are most often used in business applications. These predictions will aid in future stock prediction. (Raza 2017)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(Islam, Salam, and Hasan 2015)(W. Chen et al. 2021).

More than 68 articles on stock price predictions have been published on IEEE xplore and Google scholar in the past 5 years, which can be used to invest heavily in stocks for large corporations(Mwaanga and Njebele 2017). w(Taunk et al. 2019) algorithms in terms of high-performance efficiency. This particular article presents the comparative analysis between the accuracy control of Novel Naive Bayes (NNB)(S. Chen et al. 2020) and K-Nearest Neighbor Algorithm (KNN) using conventional controllers like stock prediction controllers and Stock Price Controller (SPC). A novel method for Naive Bayes using K-Nearest Neighbor has been proposed in this Article for the efficiency and Accuracy improvement.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)

From this literature survey, it can be concluded that this Naive Bayes Algorithm for stock price prediction helps to predict the Stock price to improve Accuracy which brings the effective results as the previous results were given less accuracy.The accuracy improvement of the Logistic Regression technique with stock price prediction was not adequately performed in a prior study to improve the stock price prediction's log loss rate(Sriram 2020). The aim of this study is to achieve a higher accuracy rate for Stock price prediction using Naive Bayes Algorithm.

2. Material 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.

By comparing both of the controllers in Supervised Learning, the sample size was estimated using the GPower software. Using G power, a sample size of 600 was computed for each group. Clinical analysis was used to calculate sample size, with alpha and beta values of 0.05 and 0.5, respectively. We employed 95 percent confidence intervals, 80 percent pretest power, and a one-to-one enrolment ratio. Technical Analysis software is used to implement two methods (Naive Bayes Algorithm and K-Nearest Neighbor Algorithm). And since no human or animal samples were utilized in this study, no ethical approval was required for this research.

Fundamental analysis and Technical analysis are two methodologies that are extensively employed in general.
Fundamental Analysis: It is essential 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 values 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.

Naive Bayes Classifier Algorithm (Generative Learning Model):

It is a classification technique based on Bayes' Theorem with an assumption of independence among predictors. In simple terms, a Naive Bayes classifier assumes that the presence of a particular feature in a class is unrelated to the presence of any other feature. Even if these features depend on each other or upon the existence of the other features, all of these properties independently contribute to the probability. The Naive Bayes model is easy to build and particularly useful for very large data sets. Along with simplicity, Naive Bayes is known to outperform even highly sophisticated classification methods.

Steps to Implement Naive Bayes Model:

- Step 1: Separate by Class.
- Step 2: Summarize Dataset.
- Step 3: Summarize Data By Class.
- Step 4: Gaussian Probability Density Function.
- Step 5: Class Probabilities

K-Nearest Neighbor Algorithm:

The k-nearest-neighbors algorithm is a classification algorithm, and it is supervised: it takes a bunch of labeled points and uses them to learn how to label other points. To label a new point, it looks at the labeled points closest to that new point (those are its nearest neighbors), and has those neighbors vote, so whichever label the most of the neighbors have is the label for the new point (the "k" is the number of neighbors it checks).

Steps to implement K-Nearest Neighbor Algorithm :

- Step 1: Data Preprocessing step
- Step 2: Fitting the K-NN algorithm to the Training set
- Step 3: Predicting the test result
- Step 4: Test accuracy of the result(Creation of Confusion matrix)
- Step 5: Visualizing the test set result.

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:

Apart from the trail investigation, the work was statistically evaluated using the Statistical Package for Social Sciences (SPSS). The mean, standard deviation, and standard error mean were calculated during the analysis. To compare the parameters in both groups, an independent variable T test was utilized. The dependent variable is efficiency, and the independent variable is Naive Bayes efficiency. The accuracy of the Naive Bayes is calculated using separate T test analysis for both approaches.

3. Results

Table 1 shows the simulation results of proposed Novel Naive Bayes Algorithm and the existing system K-Nearest Neighbor 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 Novel Naive Bayes Algorithm was 82.19% and the K-Nearest Neighbor algorithm was 46.50%.

Table 2 represents the T-test comparison of both Novel Naive Bayes Algorithm and K-Nearest Neighbor algorithm. The Mean, Standard Deviation and Standard Error Mean were calculated by taking an independent variable T test among the study groups. The Naive Bayes algorithm produces a significant difference than the KNN algorithm with a value of 0.016 .

Table 3 represents the Mean of Novel Naive Bayes Algorithm which is better compared with the K-Nearest Neighbor algorithm with a standard deviation of 1.0628 and 0.6630 respectively. From the results, the Novel Naive Bayes Algorithm (82.19%) gives better accuracy than the K-Nearest Neighbor algorithm (46.50%). Figure 1 gives the comparison chart of the Novel Naive Bayes Algorithm of K-Nearest Neighbor Algorithms in terms of mean and accuracy. The mean accuracy of the Naive Bayes Algorithm is better than KNN Algorithm. Figure 2 shows the error mean of Naive Bayes Algorithm (0.33) and KNN Algorithm (0.20).

4. Discussions

Naive Bayes and KNN 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 Naive Bayes algorithm provides better accuracy results compared to the KNN algorithm.

In the recent survey, the proposed (J. W. Lee 2002) Showed that a technique for applying reinforcement learning, reasonable for demonstrating and learning different sorts of corporations in genuine circumstances, to the issue of stock value forecast. The stock value forecast issue is considered as a Markov process which can be streamlined by reinforcement learning based algorithms. TD(0), a reinforcement learning which advances just from encounters, is embraced and work estimation by a fake neural organization is performed to get familiar with the upsides of states every one of which relates to a stock value pattern at a given time. A test result in view of the Korean financial exchange is introduced to assess the exhibition of the proposed strategy.

This paper (Shihavuddin et al. 2010) addresses data mining algorithms which have been tried on this accessible data to learn helpful patterns about the behavior of the stock market. The learned pattern holds the way to decipher the present and foresee the following stock cost. That loathed work utilizes Naïve Bayes Algorithm to arrange text news connected with FTSE100 given on these referenced sites and the classifier is prepared to become familiar with the development in the stock cost (up or down) from the news stories in the pages of that day.

In this paper, (C.-H. Lee 2015)proposed technique is executed with regards to naive Bayesian learning, and ideal loads of element values are determined utilizing a gradient approach. The presentation of naive Bayes learning with esteem weighting technique is contrasted with that of other cutting edge strategies for various datasets. The exploratory outcomes show that the worth weighting strategy could work on the exhibition of naive Bayes significantly. In this paper (Golmohammadi, Zaiane, and Diaz 2014) they utilize supervised learning algorithms to distinguish dubious exchanges corresponding to showcase control in financial exchange. They utilized a contextual investigation of controlled stocks during 2003. They embrace CART, contingent deduction trees, C5.0, Random Forest, Naïve Bayes, Neural Networks, SVM and kNN for characterization of controlled examples. Experimental outcomes show that Naïve Bayes beat other learning techniques accomplishing F 2 proportion of 53% (responsiveness and explicitness are 89% and 83% separately).

From the above conversation, a couple of articles guarantee that they give preferable execution over the proposed Novel Naive Bayes and K-Nearest Neighbor Algorithm for further developing the accuracy of stock market prediction. Likewise, the current value forecast requires no extra expense and accordingly got exceptional consideration as of late. In this way, we can gather that the proposed Naive Bayes and K-Nearest Neighbor Algorithm can be utilized to work on the accuracy of value expectation by managing the stock increase.

Stock market prediction has restricted value expectation capacity in view of future cost critical benefit which improves value forecast in future. Profound Learning calculation can address future stock expectations.

5. Conclusion

The work invovels Novel Naive Bayes Algorithm to find the Stock Price Prediction to be proved with better accuracy of 82.19% when compared to K-Nearest Neighbor Algorithm accuracy is 46.50% 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

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.

Chen, Shenglei, Geoffrey I. Webb, Linyuan Liu, and Xin Ma. 2020. "A Novel Selective Naïve Bayes Algorithm." Knowledge-Based Systems 192 (105361): 105361.

Chen, Wei, Haoyu Zhang, Mukesh Kumar Mehlawat, and Lifen Jia. 2021. "Mean–variance Portfolio Optimization Using Machine Learning-Based Stock Price Prediction." Applied Soft Computing 100 (106943): 106943.

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.

Ding, Guangyu, and Liangxi Qin. 2020. "Study on the Prediction of Stock Price Based on the Associated Network Model of LSTM." International Journal of Machine Learning and Cybernetics 11 (6): 1307–17.

Golmohammadi, Koosha, Osmar R. Zaiane, and David Diaz. 2014. "Detecting Stock Market Manipulation Using Supervised Learning Algorithms." In 2014 International Conference on Data Science and Advanced Analytics (DSAA). IEEE. https://doi.org/10.1109/dsaa.2014.7058109.

Hsieh, Tsung-Jung, Hsiao-Fen Hsiao, and Wei-Chang Yeh. 2011. "Forecasting Stock Markets Using Wavelet Transforms and Recurrent Neural Networks: An Integrated System Based on Artificial Bee Colony Algorithm." Applied Soft Computing 11 (2): 2510–25.

Huang, Wei, Yoshiteru Nakamori, and Shou-Yang Wang. 2005. "Forecasting Stock Market Movement Direction with Support Vector Machine." Computers & Operations Research 32 (10): 2513–22.

Islam, Md Shariful, Mohammad Abdus Salam, and Md Mahmud Hasan. 2015. "Factors Affecting the Stock Price Movement: A Case Study on Dhaka Stock Exchange." International Journal of Indian Culture and Business Management 10 (10). https://doi.org/10.5539/ijbm.v10n10p253.

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.

Lee, Chang-Hwan. 2015. "A Gradient Approach for Value Weighted Classification Learning in Naive Bayes." Knowledge-Based Systems 85 (September): 71–79.

Lee, Jae Won. 2002. "Stock Price Prediction Using Reinforcement Learning." In ISIE 2001. 2001 IEEE International Symposium on Industrial Electronics Proceedings (Cat. No.01TH8570). IEEE. https://doi.org/10.1109/isie.2001.931880.

M, Hiransha, Gopalakrishnan E.a., Vijay Krishna Menon, and Soman K.p. 2018. "NSE Stock Market Prediction Using Deep-Learning Models." Procedia Computer Science 132: 1351–62.

Mwaanga, Clement, and Nsama Njebele. 2017. "The Long-Run and Short-Run Relationship between the Exchange Rates and Stock Market Prices." Journal of Financial Risk Management 06 (04): 315–24.

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.

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.

Raza, Kamran. 2017. "Prediction of Stock Market Performance by Using Machine Learning Techniques." In 2017 International Conference on Innovations in Electrical Engineering and Computational Technologies (ICIEECT). IEEE. https://doi.org/10.1109/icieect.2017.7916583.

Shihavuddin, A. S. M., Mir Nahidul Ambia, Mir Mohammad Nazmul Arefin, Mokarrom Hossain, and Adnan Anwar. 2010. "Prediction of Stock Price Analyzing the Online Financial News Using Naive Bayes Classifier and Local Economic Trends." In 2010 3rd International Conference on Advanced Computer Theory and Engineering(ICACTE). IEEE. https://doi.org/10.1109/icacte.2010.5579624.

Sriram, Shreenidhi. 2020. "Stock Market Prediction Using Logistic Regression Analysis - A Pilot Study." International Journal for Research in Applied Science and Engineering Technology 8 (7): 2039–43.

Taunk, Kashvi, Sanjukta De, Srishti Verma, and Aleena Swetapadma. 2019. "A Brief Review of Nearest Neighbor Algorithm for Learning and Classification." In 2019 International Conference on Intelligent Computing and Control Systems (ICCS). IEEE. https://doi.org/10.1109/iccs45141.2019.9065747.

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.

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 Naive Bayes and K-Nearest Neighbor algorithm(KNN)

SAMPLES	NOVEL NAIVE BAYES ALGORITHM (ACCURACY)	K-NEAREST NEIGHBOR ALGORITHM (ACCURACY)
1	83.0	46.5
2	81.0	46.9
3	80.8	46.7
4	83.2	45.2
5	81.3	47.5
6	83.3	46.8

7	82.7	47.2
8	82.5	46.0
9	83.2	45.9
10	81.0	46.4

Table 2: Comparison of mean of Accuracy using Novel Support Vector Machine and Logistic Regression algorithms.

Matrics	**Algorithm**	**No of Samples**	**Mean**	**Std.Deviation**	**Std.Error Mean**
Accuracy	Naives Bayes algorithm	10	82.192	1.0628	0.3361
	KNN	10	46.503	0.6630	0.2097

Group Statistics

Group statistics comparison of accuracy for predecting Stock price using Naives Bayes Algorithm and K-Nearest Neighbor algorithm is done. Novel Naives Bayes algorithm has higher mean compared to KNN algorithm.

Naives Bayes algorithm = 82.192

KNN Algorithm= 46.503

Table 3: Independent sample T-test is performed for the two groups for significance and standard error determination $p<0.05$ for 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-tailed)**	**Mean Difference**	**std.Error Difference**	**Lower**	**Upper**
Accuracy	Equal Variance assumed	7.137	0.016	90.100	18	0.000	35.6890	0.3961	34.8568	36.5212

	Equal variance not assumed			90.100	15.083	0.000	35.6890	0.3961	34.8451	36.5329

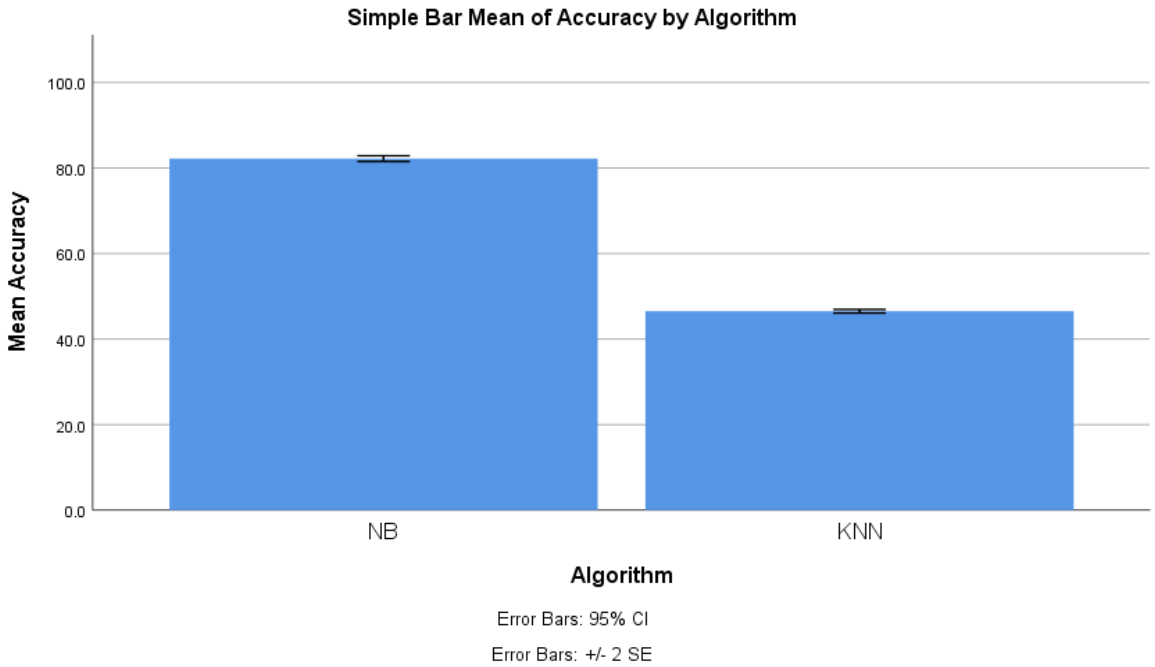

Fig-1: Sample mean of accuracy by using Naive Bayes algorithm and K-Nearest Neighbor algorithm.