IMPROVED ACCURACY IN STOCK PRICE PREDICTION SYSTEM USING A NOVEL SUPPORT VECTOR MACHINE ALGORITHM COMPARED TO K-NEAREST NEIGHBOR ALGORITHM

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 Support Vector Machine 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 Support Vector Machine Algorithm (N=10) and K-Nearest Neighbor (N=10) are simulated by adjusting the Novel Support Vector Machine 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 Support Vector Machine generates 84.67 percent accuracy, KNN produces 47.87 percent accuracy. The difference in statistical significance between Novel Support Vector Machine and KNN was discovered to be 0.508 ($p<0.05$). **Conclusion:** In terms of accuracy, the Support Vector Machine algorithm outperforms the K-Nearest Neighbor algorithm.

Keywords: Machine Learning, Stock price prediction, Support Vector Machine 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

The machine learning(Jordan and Mitchell 2015) model had a significant impact in this study of stock price prediction(Nayak, Pai, and Pai 2016), which produces forecasts based on current stock market values by training on prior values(Li et al. 2011). 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(Leitch and Paulson 1975). A stock price prediction is a trading platform where different investors sell and buy the shares consistently with stock availability(Gabriel and Ugochukwu 2012). Predicting the stock market with full accuracy is very difficult as external entities such as social(Z. Wang, Ho, and Lin 2018),political, psychological and economic have a great and substantial influence on it(Rao, Srinivas, and Mohan 2020). Machine learning is also often referred to as predictive analytics or predictive modeling(Dhall, Kaur, and Juneja 2020). The Machine Learning model had a significant impact in the following research work on Stock Price Prediction(Attanasio et al. 2019), which generates forecasts based on the present stock market's values by training on their prior values. These predictions will help to predict the stocks in future. It has continually been a warm spot for buyers and investing groups to comprehend the alternate regularity of the inventory market price and expect its trend(Gurav and Sidnal 2018).Because stock investing is such a large part of the financial market, a lack of proper understanding and detailed information will inevitably result in a loss of investment(Subhi Alzazah and Cheng 2021).A near gets anticipating the speculation can lead towards benefit(H. Wang et al. 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)

The Literature survey in Stock price prediction systems and many comparisons have also been done for this prediction.In the last 5 years, more than 63 papers have been published on IEEE xplore and google scholar on Stock price predictions which can be greatly invested in stocks for big companies. In this article investigation of Novel Support Vector Machine Algorithm(Gandhmal and Kumar 2019) and K-Nearest Neighbor Algorithm(Xiao, Zhong, and Zhong 2020). 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 Support Vector Machine Algorithm and K-Nearest Neighbor using conventional controllers like stock prediction controllers and Stock Price Controller (SPC). This article proposes a novel approach for that Support Vector Machine uses K-Nearest Neighbor to increase efficiency and accuracy of the system.

The research gap identified from the literature survey in the last 10 years, there have been many articles published in IEEE Xplore and some were published in Google Scholar for past few years, proposed machine learning algorithms such as K-Nearest Neighbors algorithm, Novel Linear Regression, Support vector machine algorithm, laboratory of neurogenetics and neuroscience algorithm, K-Nearest neighbors algorithm for Stock price prediction system.The Accuracy Improvement of the Naive Bayes algorithm with stock price prediction was not adequately implemented in a prior study to improve the stock price prediction's log loss rate. The aim of this study is to show that the Novel Support Vector Machine algorithm appears to perform better among all the algorithms. To overcome this issue a novel Support Vector Machine Algorithm is implemented to improve log loss rate of stock market prediction.

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 sample size has been calculated using the GPower software by comparing both of the controllers in Supervised Learning. 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 Support Vector Algorithm and K-Nearest Neighbor 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: To determine the accurate product value, reliable and accurate information on the financial report of the company, it is necessary to have competitive analysis strength and economic conditions in which they are interested. For long-term predictions, Fundamental analysis is useful and the advantages are due to their systematic approval and their ability to predict changes.
Technical Analysis: The idea behind technical analysis is that investors constantly changing attributes in response to different forces and factors make stock prices trends and movements. There are various different technical factors of quantitative parameters that can be used for analysis, such as trend indicators, lowest and highest daily values, daily ups and downs, stock volume, indices, etc. It is possible to extract rules from the data and the investors make future decisions based on these rules. Technical analysis data is preferable over fundamental analysis data as input to the system.

Support Vector Machine Algorithm
Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, primarily, it is used for Classification problems in Machine Learning.The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that we can easily put the new data point in the correct category in the future. This best decision boundary is called a hyperplane. SVM chooses the extreme points and vectors that help in creating the hyperplane. These extreme cases are called support vectors, and hence the algorithm is termed as Support Vector Machine.
Implementation of SVM :
Step 1: Load the important libraries
Step 2: Import dataset and extract the X variables and Y separately.
Step 3: Divide the dataset into train and test.
Step 4: Initializing the SVM classifier model
Step 5: Fitting the SVM classifier model
Step 6: Coming up with predictions
Step 7: Evaluating model's performance

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 8 GB. The system type used was a 64bit OS, X64 based processor with SDD of 256 GB. 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 Support Vector Machine is calculated using separate T-test analysis for both approaches.

3. Results

Table 1 shows the simulation results of proposed Novel Support Vector Machine 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 Support Vector Machine Algorithm was 84.67% and the K-Nearest Neighbor algorithm was 47.87%.

Table 2 represents the T-test comparison of both Novel Support Vector Machine 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 SVM algorithm produces a significant difference than the Logistic Regression algorithm with a value of 0.508 .

Table 3 represents the Mean of Novel Support Vector Machine Algorithm which is better compared with the K-Nearest Neighbor algorithm with a standard deviation of 1.23630 and 1.48216 respectively. From the results, Novel Support Vector Machine Algorithm (84.67%) gives better accuracy than the Nearest Neighbor algorithm (47.87%). Figure 1 gives the comparison chart of the Novel Support Vector Machine Algorithm of KNN algorithms in terms of mean and accuracy. The mean accuracy of the SVM algorithm is better than KNN algorithm. Figure 2 shows the error mean of SVM algorithm (0.39) and K-Nearest Neighbor algorithm (0.46).

4. Discussions

SVM and KNN Algorithms are implemented and compared for Stock price prediction to improve the accuracy by stock pricing. From obtained results it is concluded that the SVM algorithm provides better accuracy results compared to the KNN algorithm.

In this paper (Fenghua et al. 2014) they implemented singular spectrum analysis and Support vector machine algorithm to make price prediction, the empirical evidence shows that SVM produces better outcomes when compared to singular spectrum analysis.

In this paper (Rajeswar, Ramalingam, and SudalaiMuthu 2022)they proposed Linear regression and K-Nearest Neighbor to foresee the exhibition of time series scope of 15 minutes dataset.The financial exchange value practices of stock trade, instability, assortment of speculation, and exchanging volume values yielded great execution of the different elements of assessment. It yielded an accuracy result of 82.58%. The yielded outcomes can assist with foreseeing the stock trade cost in Initial Public Offering.

From the above discussion, only a few articles ensure that they provide better performance than the proposed SVM and KNN algorithm for improving accuracy of stock market prediction. Also, the present price prediction requires no additional cost and therefore received intense attention in recent years. So, we can infer that the proposed SVM and KNN algorithm can be used to improve the accuracy of price prediction by regulating the stock gain.

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

5. Conclusion

The work invovels Novel Support Vector Machine algorithm to find the Stock Price Prediction to be proved with better accuracy of 84.67% when compared to K-Nearest Neighbor algorithm accuracy is 47.87% 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

Attanasio, Giuseppe, Luca Cagliero, Paolo Garza, and Elena Baralis. 2019. “Combining News Sentiment and Technical Analysis to Predict Stock Trend Reversal.” In *2019 International Conference on Data Mining Workshops (ICDMW)*. IEEE. https://doi.org/10.1109/icdmw.2019.00079.

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.

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.

Dhall, Devanshi, Ravinder Kaur, and Mamta Juneja. 2020. "Machine Learning: A Review of the Algorithms and Its Applications." In *Lecture Notes in Electrical Engineering*, 47–63. Lecture Notes in Electrical Engineering. Cham: Springer International Publishing.

Fenghua, W. E. N., Xiao Jihong, H. E. Zhifang, and Gong Xu. 2014. "Stock Price Prediction Based on SSA and SVM." *Procedia Computer Science* 31: 625–31.

Gabriel, Ajao Mayowa, and Wemambu Mary Ugochukwu. 2012. "Volatility Estimation and Stock Price Prediction in the Nigerian Stock Market." *International Journal of Financial Research* 3 (1). https://doi.org/10.5430/ijfr.v3n1p2.

Gandhmal, Dattatray P., and K. Kumar. 2019. "Systematic Analysis and Review of Stock Market Prediction Techniques." *Computer Science Review* 34 (100190): 100190.

Gurav, Uma, and Nandini Sidnal. 2018. "Predict Stock Market Behavior: Role of Machine Learning Algorithms." In *Intelligent Computing and Information and Communication*, 383–94. Advances in Intelligent Systems and Computing. Singapore: Springer Singapore.

Jordan, M. I., and T. M. Mitchell. 2015. "Machine Learning: Trends, Perspectives, and Prospects." *Science* 349 (6245): 255–60.

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.

Leitch, R. A., and A. S. Paulson. 1975. "Estimation of Stable Law Parameters: Stock Price Behavior Application." *Journal of the American Statistical Association* 70 (351a): 690–97.

Li, Xiaodong, Chao Wang, Jiawei Dong, Feng Wang, Xiaotie Deng, and Shanfeng Zhu. 2011. "Improving Stock Market Prediction by Integrating Both Market News and Stock Prices." In *Lecture Notes in Computer Science*, 279–93. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg.

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.

Nayak, Aparna, M. M. Manohara Pai, and Radhika M. Pai. 2016. "Prediction Models for Indian Stock Market." *Procedia Computer Science* 89: 441–49.

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.

Rajeswar, Surya, Pon Ramalingam, and T. SudalaiMuthu. 2022. "Comparative Analysis of Stock Market Price Behaviors Using Machine Learning Techniques." In *INTERNATIONAL CONFERENCE ON ADVANCES IN MATERIALS, COMPUTING AND COMMUNICATION TECHNOLOGIES: (ICAMCCT 2021)*. AIP Publishing. https://doi.org/10.1063/5.0071068.

Rao, Polamuri Subba, K. Srinivas, and A. Krishna Mohan. 2020. "A Survey on Stock Market Prediction Using Machine Learning Techniques." In *Lecture Notes in Electrical Engineering*, 923–31. Lecture Notes in Electrical Engineering. Singapore: Springer

Singapore.

Subhi Alzazah, Faten, and Xiaochun Cheng. 2021. "Recent Advances in Stock Market Prediction Using Text Mining: A Survey." In *E-Business - Higher Education and Intelligence Applications*. IntechOpen.

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.

Wang, Haiyao, Jianxuan Wang, Lihui Cao, Yifan Li, Qiuhong Sun, and Jingyang Wang. 2021. "A Stock Closing Price Prediction Model Based on CNN-BiSLSTM." *Complexity* 2021 (September): 1–12.

Wang, Zhaoxia, Seng-Beng Ho, and Zhiping Lin. 2018. "Stock Market Prediction Analysis by Incorporating Social and News Opinion and Sentiment." In *2018 IEEE International Conference on Data Mining Workshops (ICDMW)*. IEEE. https://doi.org/10.1109/icdmw.2018.00195.

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.

Xiao, Qingtao, Xin Zhong, and Chenghua Zhong. 2020. "Application Research of KNN Algorithm Based on Clustering in Big Data Talent Demand Information Classification." *International Journal of Pattern Recognition and Artificial Intelligence* 34 (06): 2050015.

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 Support Vector Machine and K-Nearest Neighbor

SAMPLES	NOVEL SUPPORT VECTOR MACHINE ALGORITHM (ACCURACY)	K-NEAREST NEIGHBOR ALGORITHM (ACCURACY)
1	84.50	47.89
2	83.65	48.94
3	86.32	46.45
4	85.34	48.46
5	84.78	49.82

6	85.23	46.36
7	82.56	48.38
8	84.43	49.67
9	83.47	45.32
10	86.47	47.45

Table 2: Comparison of mean of Accuracy using Novel Support Vector Machine and K-Nearest Neighbor algorithms.

Matrics	**Algorithm**	**No of Samples**	**Mean**	**Std. Deviation**	**Std.Error Mean**
Accuracy	Support vector machine	10	84.6750	1.23630	0.39095
	K-Nearest Neighbor	10	47.8740	1.48216	0.46870

Group statistics

Group statistics comparison of accuracy for predecting Stock price using Support vector machine and K-Nearest Neighbor is done. SVM algorithm has higher mean compared to KNN algorithm.

Support vector Machine algorithm = 84.6750
K-Nearest Neighbor Algorithm= 47.8740

Table 2. This Independent sample t-test obtained the significance as 0.508 (p>0.05) (2-tailed), significance, mean difference, std. error difference, and lower and upper interval difference. Independent samples t-test is applied for comparison of SVM and KNN.

		Leven's test for equality of variance		T-Test for Equality of Means						
		f	sig	t	df	sig	Mean Difference	Std. Error Difference	95% Confidence Interval Of Difference	
									Lower	upper
Accuracy	Equal variance assumed	0.457	0.508	60.294	18	0.000	36.80100	0.61035	35.51871	38.08329
	Equal Variance not assumed			60.295	17.439	0.000	36.80100	0.61035	35.51574	38.08626

Fig-1: Sample mean of accuracy by using SVM algorithm and KNN algorithm.

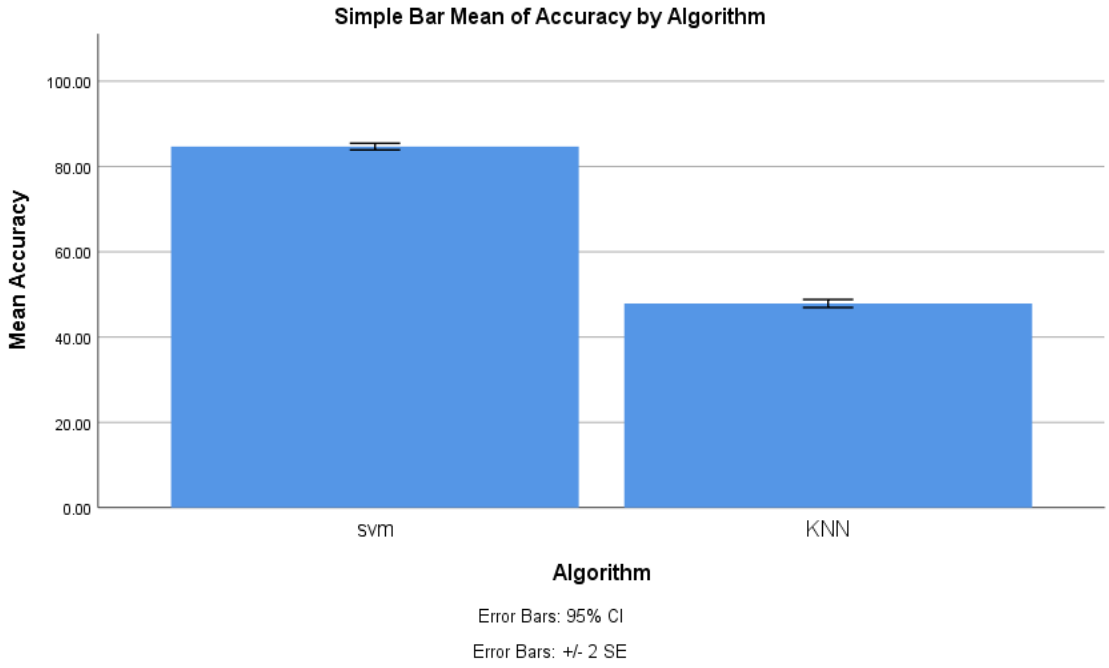

Fig. 1. Comparison of Novel Support Vector Machine Algorithm and K-Nearest Neighbor in terms of mean and accuracy. The mean accuracy of the Novel Support Vector Machine algorithm is better than the KNN algorithm. X-axis: SVM vs KNN, Y-axis: Mean accuracy of detection± 2SE