Original Article

Prediction Of Insufficient Accuracy For Kidney Cancer Renal Cell Carcinoma Using Naive Bayes In Comparison With K-Nearest Neighbour Algorithm With Improved Accuracy

G Nanda Vardhan Reddy[1], Saravanan.M.S[2]

[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 nandhavardhang18@saveetha.com

[2]Project Guide, Corresponding author, Department of Computer Science and Engineering,
Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences ,
Saveetha University, Chennai, Tamilnadu, India. Pincode:602105. saravananms@saveetha.com
DOI: 10.47750/pnr.2022.13.S04.277

Abstract

Aim: The main objective of this article is comparing K-Nearest Neighbour (KNN) with Naive Bayes to find out the best classification algorithm among these two based on accuracy. **Materials and Methods:** The study included 450 samples, two sets of algorithms, a G-power value of 80%, and kidney photos from a range of websites. The study's cutoff was 0.05 percent, and its mean, standard deviation, and confidence intervals were all 95%. Because the KNN algorithm has previously achieved 91.05 percent accuracy in predicting the renal failure rate for individuals, this research needs to develop a higher accuracy for kidney failure prediction using the Novel Naïve Bayes machine learning algorithm. **Results:** The accuracy for detecting renal failure using the Naïve Bayes algorithm was determined to be 94.23% in this study, with a significant value of 0.001 (p 0.05) for two-tailed testing. **Conclusion:** This study's findings demonstrate that the Naïve Bayes algorithm performs noticeably better on kidney images than the KNN algorithm.

Keywords: Machine Learning, Novel Naive Bayes, K-Nearest Neighbour, Tumor, Kidney Images, Carcinoma.

INTRODUCTION

90 percent of malignant kidney tumours are Renal Cell Carcinoma (RCC) and it affects both men and women (Gupta and Kanwar 2021). There are different types of kidney cancers like Urothelial Carcinoma, Sarcoma, Wilms Tumor etc. Dangerous growth of the kidney tumor and renal pelvis represents almost 4% of disease cases and more than 2% malignancy passes in india. At the age of 64 people are suffering from diagnosis, and mostly people are diagnosed between 65 and 74. There are about 29 million people living with kidney cancer worldwide (Oecd and OECD 2019).

Renal cell carcinoma (RCC) has expanded its occurrence at a pace of 2% each year in the most recent 65 years just as its mortality in the past two decades (Kitano, Ito, and Kim 2017). There are very much reported differences in histologic appearances of chRCC and renal oncocytoma: chRCC shows unmistakable cell line, raisinoid cores and perinuclear radiance, while oncocytoma shows settled engineering with myxoid or hyalinized stroma and cells with eosinophilic or granular cytoplasm and little round cores (Das et al. 2020). This research study's main objective is to identify various kidney problems and treat them using machine learning techniques. Applications of this study include understanding renal cell cancer and numerous age-related risk factors (Rodrigo 2017).

Studies on machine learning have been used to predict renal cell carcinoma and as a reference for kidney cancer. In the research study by (Riaño, Wilk, and ten Teije 2019), various classifiers were employed, including Naïve Bayes, KNN, SVM, DT, linear regression, and CNN (Deo 2015). The comparison of the algorithms is intended to determine which algorithm is best for predicting kidney cancer instances for cutting-edge tumour detections. Smaller datasets respond nicely to the KNN algorithm (Wu, Shen, and Sabuncu 2016). Naïve Bayes and KNN were utilised by Urbanos in another research work to train their models. They applied a DT classifier for kidney cancer to a dataset with 2400 CT scans and 1140 X-ray pictures that was made available to the public (Xing et al. 2007). Patients are screened with it for. It is used to screen patients so that decisions can be made quickly. This research study publishes and presents more than 42 indexed citations (Chandola et al. 2021). Another study (Vandana and Alva 2021) employed an AI method based on KNN to identify kidney cancer patients using CT scan and X-ray data (Cleff 2019). The majority of pertinent articles lack an accuracy rating for the revolutionary Naïve Bayes classifier for cutting-edge illness detections that can predict renal cell carcinoma (Szolovits 2019).

The research gap identified is that several works have demonstrated that the performance KNN provides less accuracy in prediction of kidney cancer renal cell carcinoma. Therefore, the primary goal of this research study is to expand the predicting model and improve the accuracy of kidney cancer prediction using the Novel Naïve Bayes learning method.

MATERIALS AND METHODS

The Artificial Intelligence Laboratory was used for this research project. There are two different categories of classification algorithms used in this study. The G-power version 3.1.10 was used for Naïve Bayes method is in Group 1 and the KNN algorithm is in Group 2 (Irizarry 2019). For this experiment, the G-power was set at 80% with a 0.05 threshold value and a 95% confidence range.

In The most well-known website for data scientists and machine learning practitioners, Kaggle.com, provided the dataset for this study. It enables users to browse and search through the different datasets. It offers a free personal Jupyter notebook environment online. The kidney image dataset includes null, missing, and duplicate values in addition to information about the signs and symptoms of kidney cancer.

Novel Naïve Bayes is unsupervised learning. Both classification and regression are accomplished using the cutting-edge support vector machine methodology. A categorical decision variable is used by the Naïve Bayes. The Gini Index is computed when the Naïve Bayes algorithm is used. When all splits have been analysed and the gini score has been determined, the dataset is divided by creating a root node. The terminal node in the recursive splitting strategy is used in both the final prediction and the tree-building process.

The Naïve Bayes algorithm applies the following pseudocode on the dataset. In the following pseudocode, the datasets will be used as input and the output will be determined by accuracy parameters.

Pseudocode of Naive Bayes

```
id3(examples,attributes)
Node = Naive Bayes Node(example)
Dictionary = summarizeExamples(examples,target attributes) FOR key in
dictionary:
        If dictionary[key]==total number of examples Node.label = key
Return node
If attributes is empty or number of examples < minimum allowed per branch:
Node.label = most common value in examples
Return node bestA = the attribute with the most information gain
Node.Naive Bayes = bestA Each
possible value v of bestA:
If subset is not empty:
Node.add branch(id3(subset,target attribute, attributes-bestA))
Return node
```

K-Nearest Neighbour is among the Machine Learning computations that have been accepted. It can be used to address issues with relapse and characterization type precognition problems. KNN is a non-parametric learning computation since it ignores all underlying information.

Pseudocode of K-Nearest Neighbour

```
Classify (X,Y,x) || x:training data , y: class labels of X, x:unknown sample
FOR i = 1 to m do
Compute distance d(X,x)
END FOR
Compute set I containing indices for the KNN - smallest distance d(xi,x)
Return majority label for{Y;where i ∈ 1}
```

For computations, a 50 GB hard disc and an Intel i5 processor are the absolute minimum requirements. The 4.1.15 version of Anaconda Spyder must be used on Windows and the algorithm must be run through the programme.

Statistical Analysis

Finally, the results from the two groups will be combined, and algorithms employing the 26 versions of the Statistical Package for Social Sciences will be used (SPSS). Both the KNN and Naïve Bayes algorithms were examined using the independent sample t-test. The dependent variables in our kidney cancer study are fatigue, blood and urine, eyesight issues, etc., while the independent variable in the dataset is the nation (Godoy and Guerin 2020). One of the permitted Machine Learning computations is the use of KNN. It can be applied to solve characterisation and relapse prediction problems. KNN is a non-parametric learning technique because it lacks any underlying data.

RESULTS

Table 1 displays the findings of a comparison between the Naïve Bayes machine learning algorithm and the KNN approach utilising 51 samples and various datasets for 30% testing and 70% training. The dataset contains

570 rows, and after 10 rounds, the accuracy of both techniques was assessed. Finally, the Naïve Bayes algorithm performs significantly better than the KNN technique.

The KNN method yielded standard deviations of 2.06 and standard error of 0.59 after statistical analysis of the 10 samples in Table 2 was complete, whereas the Naïve Bayes approach produced standard deviations of 1.88 and standard error of 0.65. Our hypothesis is correct because the 2-tailed significant value is less than 0.05 (p=0.001). As demonstrated in Table 3, when the input values (independent variables), which are associated output values, change.

Tables 2 and 3 display the outcomes of a group statistics independent sample t-test. These findings support the validity of our hypothesis and have two-tailed significant values (p=0.001) below 0.05. It is therefore established that the Naïve Bayes and KNN algorithms are very different from one another. The accuracy of the two algorithms was compared using an independent sample t-test, and a statistically significant difference (p 0.05) was found. The accuracy of the Naïve Bayes model was 94.23%. The proposed naïve Bayes classifier outperformed KNN in comparison to the performance of the other algorithm.

The preprocessing is used to clean the data to apply the for the machine learning algorithm, due to the original image containing more errors and null values, it has been preprocessed. The image has been resized and made ready for the next stage, as the network allows. The algorithm for preprocessing the images can be implemented using Spyder functions. The human kidney tumor segmentation shows in Fig. 1.

Two algorithms' accuracy was compared using an independent sample t-test with two-tailed significant "p" values, and a difference was noted at p<0.001. The accuracy rates for the KNN method were 91.05 percent, and the Naïve Bayes model was 94.23 percent accurate.

Figure 2 illustrates a Performance Analysis comparison between Naïve Bayes and KNN. The KNN algorithm represents a black curve, while the naïve Bayes algorithm represents a red curve. Naïve Bayes accuracy of 94.60 percent is considerably higher than KNN accuracy of 91.90 percent.

In Fig. 3, where Naïve Bayes and KNN object counts are compared, Naïve Bayes is represented by the red curve and KNN by the black curve. Naïve Bayes accuracy rate of 90.70% is higher than KNN accuracy rate of 88.80%.

In Fig. 4, which compares the mean accuracy of the Naïve Bayes and KNN algorithms, it is revealed that the Naïve Bayes method is much superior to the KNN algorithm.

DISCUSSION

KNN is greatly outperformed by Novel Naïve Bayes. The Naïve Bayes performs and scores much better than KNN in terms of speed and performance. P value 0.001 is less than 0.05 (p0.05) to indicate the significance of the 2-tailed test. As a result, the research study estimated that the Naïve Bayes algorithm has 94.23 percent accuracy by comparing it to the KNN method, which has 91.05 percent accuracy, and that the Naïve Bayes algorithm is considerably superior to the KNN algorithm with a significance value of (p0.05).

We used 10 samples as input for the machine learning algorithms. Naïve Bayes and KNN algorithm approaches have all been used to identify the kidney cancer prediction. The KNN algorithm outperforms the Naïve Bayes approach when compared to other algorithms (Chollet 2017). In this study small number of test cases were employed to get at the other research findings (Hyett, Parker, and Dhall 2016).

To oppose this study some other research findings used image based classification of kidney cancer renal cell carcinoma but we are using a graph image of carcinoma dataset for analysis. In other existing studies, they only applied for specific types of carcinoma but in this research study (Kevin Zhou, Greenspan, and Shen 2017) various types of renal cell carcinoma are used for learning. The graphs images are used in the existing study.

The Naive Bayes method is utilized, within the constraints of the research study, to locate a kidney images dataset with various features, like ECG, renal diseases and comparing the two groups for best performance. New features will be able to accurately and sensitively predicting the other diseases in the future studies.

The Naïve Bayes method is utilised, within the constraints of the research study, to locate a kidney pictures dataset with various features, such as renal diseases, ECG, and compares two groups for better performance. In future studies, new features will be able to accurately and sensitively predict the other diseases.

CONCLUSION

In this research, kidney cancer prediction is performed using the kidney cancer symptoms dataset for Naive Bayes and KNN. The accuracy value of the Naive Bayes classifier is 94.23% whereas the accuracy value of KNN is 91.05%. The quality of kidney cancer prediction and accuracy using Naive Bayes appears to be better than KNN.

DECLARATION

Conflict of Interests

There are no competing interests with this project.

Authors Contributions

All data collection, analysis, and manuscript writing were done by author GNVR. Conceptualization, data validation, and manuscript peer review were all completed by author SMS.

Acknowledgement

The infrastructure required to carry out this work successfully was provided by the Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences (formerly Saveetha University), for which the authors would like to express their heartfelt gratitude.

Funding

We appreciate the financial support from the organisations listed below for helping us complete the study.

1. Azacus Techno Solutions, Chennai.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

1. Brownlee, Jason. 2016. Machine Learning Mastery With Weka: Analyze Data, Develop Models, and Work Through Projects. Machine Learning Mastery.
2. Chandola, Yashvi, Jitendra Virmani, H. S. Bhadauria, and Papendra Kumar. 2021. Deep
3. Learning for Chest Radiographs: Computer-Aided Classification. Elsevier.
4. Chollet, Francois. 2017. Deep Learning with Python. Simon and Schuster.
5. Cleff, Thomas. 2019. Applied Statistics and Multivariate Data Analysis for Business and Economics: A Modern Approach Using SPSS, Stata, and Excel. Springer.
6. Das, Asit Kumar, Janmenjoy Nayak, Bighnaraj Naik, Soumi Dutta, and Danilo Pelusi. 2020. Computational Intelligence in Pattern Recognition: Proceedings of CIPR 2020. Springer Nature.
7. Deo, Rahul C. 2015. "Machine Learning in Medicine." Circulation 132 (20): 1920–30.
8. Gupta, Shruti, and Shamsher Singh Kanwar. 2021. "The Influence of Dysbiosis on Kidney Stones That Risk up Renal Cell Carcinoma (RCC)." Seminars in Cancer Biology. https://doi.org/10.1016/j.semcancer.2020.06.011.
9. Hyett, Matthew P., Gordon B. Parker, and Abhinav Dhall. 2016. "The Utility of Facial Analysis Algorithms in Detecting Melancholia." Advances in Face Detection and Facial Image
10. Analysis. https://doi.org/10.1007/978-3-319-25958-1_13.
11. Irizarry, Rafael A. 2019. Introduction to Data Science: Data Analysis and Prediction Algorithms with R. CRC Press.
12. Kevin Zhou, S., Hayit Greenspan, and Dinggang Shen. 2017. Deep Learning for Medical Image Analysis. Academic Press.
13. Kitano, Shigehisa, Ayumu Ito, and Youngji Kim. 2017. "Immunotherapy for Renal Cell Cancer (RCC)." Renal Cell Carcinoma. https://doi.org/10.1007/978-4-431-55531-5_12.
14. Moldagulova, Aiman, and Rosnafisah Bte Sulaiman. 2016. "CLASSIFICATION OF TEXTUAL
15. DOCUMENTS IN R USING KNN ALGORITHM." PONTE International Scientific
16. Researchs Journal. https://doi.org/10.21506/j.ponte.2016.12.35.
17. Oecd, and OECD. 2019. "Breast Cancer Surgery Type and Setting (2017) and Incidence (2012) per 100 000 Women." https://doi.org/10.1787/30a299f1-en.
18. Riaño, David, Szymon Wilk, and Annette ten Teije. 2019. Artificial Intelligence in Medicine:
19. 17th Conference on Artificial Intelligence in Medicine, AIME 2019, Poznan, Poland, June 26–29, 2019, Proceedings. Springer.
20. Rodrigo, Luis. 2017. Ascites: Physiopathology, Treatment, Complications and Prognosis. BoD – Books on Demand.
21. Szolovits, Peter. 2019. Artificial Intelligence In Medicine. Routledge.
22. Szymański, Piotr, and Tomasz Kajdanowicz. 2017. "Is a Data-Driven Approach Still Better Than
23. Random Choice with Naive Bayes Classifiers?" Intelligent Information and Database Systems. https://doi.org/10.1007/978-3-319-54472-4_74.
24. Vandana, B. S., and Sathyavathi R. Alva. 2021. "Deep Learning Based Automated Tool for Cancer Diagnosis from Bone Histopathology Images." 2021 International Conference on Intelligent Technologies (CONIT). https://doi.org/10.1109/conit51480.2021.9498367.
25. Wu, Guorong, Dinggang Shen, and Mert Sabuncu. 2016. Machine Learning and Medical
26. Imaging. Academic Press.

TABLES AND FIGURES

Table 1. With ten sample inputs, the Naïve Bayes and KNN algorithms are compared. Naïve Bayes has the highest performance of 94.23 percent, while KNN has the highest performance of 91.05 percent. Using a dataset size of 570 and 70% of training data and 30% of testing data.

Sample (N)	**Dataset size / Rows in %**	**Naive Bayes algorithm Accuracy in %**	**KNN algorithm Accuracy in %**
1	570	94.23	91.05
2	490	93.60	90.78
3	410	92.90	89.67

4	360	91.70	88.54
5	310	90.59	87.82
6	250	89.78	86.67
7	200	88.54	85.59
8	160	87.73	84.67
9	110	86.54	83.35
10	70	85.91	82.25

Table 2. Naïve Bayes and KNN algorithm statistical results For the Naïve Bayes and KNN algorithms, 10 iterations are acquired for the mean accuracy value, standard deviation, and standard error mean. It is discovered that the Naïve Bayes (91.54 percent) method outperforms the KNN (88.18 percent) algorithm.

Algorithms (Accuracy)	**Sample (N)**	**Mean**	**Standard Deviation**	**Standard Error Mean**
Naive Bayes algorithm	10	91.54	1.88	0.59
KNN algorithm	10	88.18	2.06	0.65

Table 3. The Naïve Bayes and KNN algorithms' independent sample t-tests yield 2 tailed significant values (p=0.001). As a result, the significance level of the Naïve Bayes and KNN algorithms is less than 0.05 with a 95% confidence range.

Accuracy	**Levene's Test for Equality of Variances**	**T-test of Equality of Means**					**95% of the confidence interval of the Difference**
		t	**df**	**Sig**	**Mean**	**Std Error**	

	F	Sig.			(2-tailed)	Difference	Difference	Lower	Upper
Equal Variance Assumed	.07	.78	3.80	18	0.001	3.36	0.88	1.50	5.21
Equal Variance Not Assumed	-	-	3.80	17.84	0.001	3.36	0.88	1.50	5.21

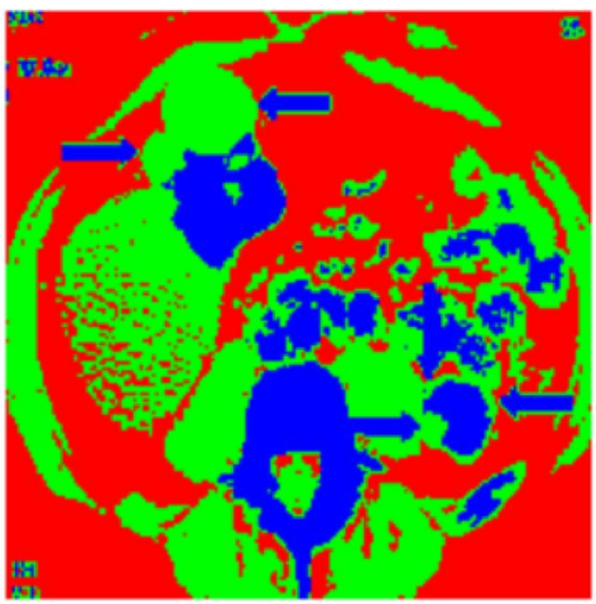

Fig. 1. Human Kidney Tumor Transparent segmentation.

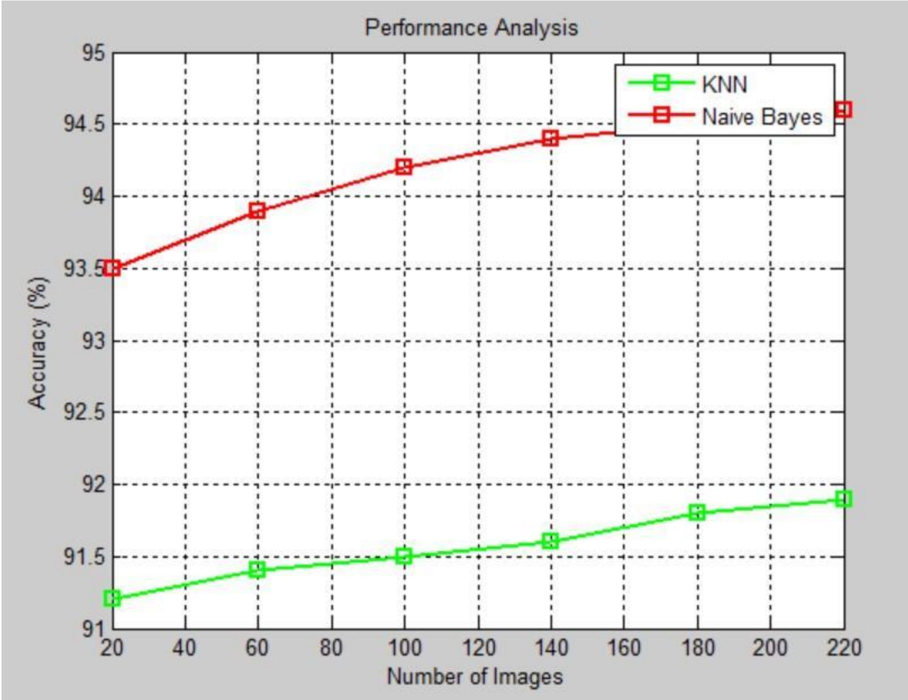

Fig. 2. Comparison between Naive Bayes and KNN. The green color curve represents the KNN (91.90) and the red curve color represents the Naive Bayes (94.60) algorithm.

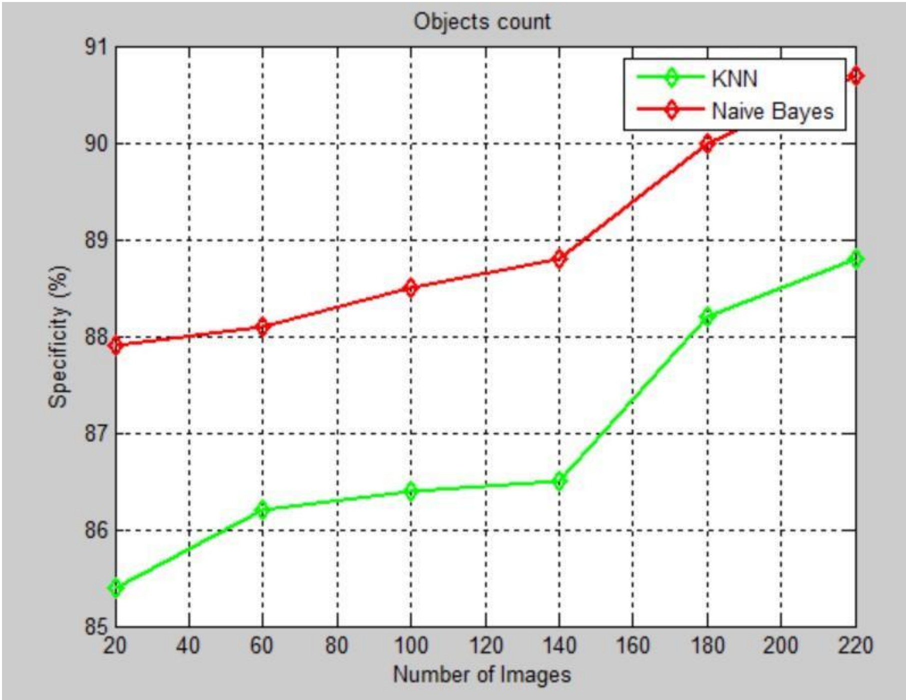

Fig. 3. Comparison between Naïve Bayes and KNN. The KNN (88.80) algorithm represents with green curve and Naïve Bayes (90.70) algorithm represents with red curve.

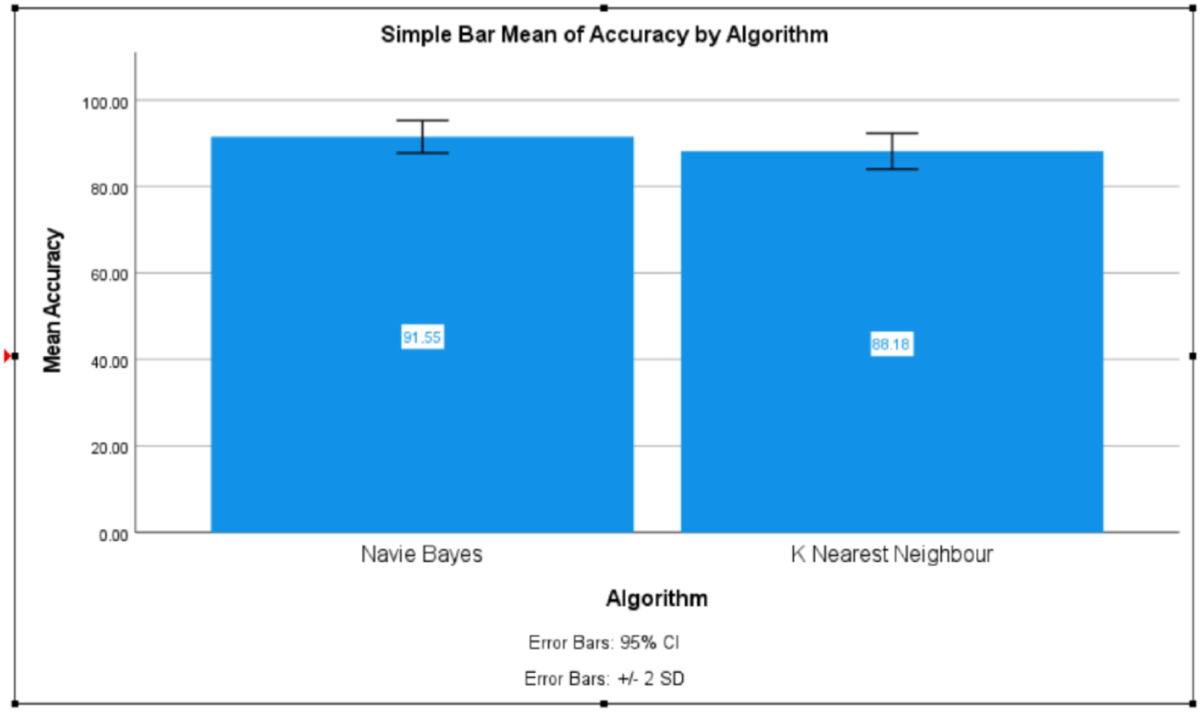

Fig. 4. In terms of mean accuracy, the Naïve Bayes and KNN algorithms are compared. Naïve Bayes standard deviation is lower than that of the KNN algorithm, and its mean accuracy is higher than that of KNN. X-axis: (GROUPS) Naïve Bayes vs KNN algorithm, and Y axis: Mean prediction accuracy 2 SD.