Original Article

Prediction of Insufficient Accuracy for Mushroom Classification whether Poisonous or Eatable Food using K-Nearest Neighbour by comparing Naive Bayes Training to Improve Accuracy

Maheswara 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.
[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.

Abstract

Aim: The main objective of the research study is to improve the good quality production for the better accuracy of eatable or poisonous mushrooms. It ought to be checked for consumable mushrooms. Exact assurance and appropriate distinguishing proof of species are the main safe approach to guarantee that the eatable mushrooms are poisonous, and defend against potential mishaps of burning through harmful one. **Materials and Methods:** The review utilized 51 samples with two groups of calculations with the G-power worth of 80% and the Mushroom were collected from different web sources with late review discoveries and edge 0.05%, certainty span 95% mean and standard deviation. To anticipate the Mushroom exactness rate for currently the Naive Bayes Training calculation has found 89.84% of precision, therefore this study needs to find better accuracy for accuracy decrease prediction with the Novel K-Nearest Neighbour (KNN) algorithm, Machine Learning calculation. **Results:** This exploration concentrated on finding 86.04% of precision for noxious identification utilizing the KNN calculation with a statistically significant difference between the two groups ($p=0.011$; $p<0.05$) with 95% confidence interval. **Conclusion:** This review reasons that the KNN calculation on exactness is essentially better compared to the Naive Bayes Training calculation.

Keywords: Machine Learning, Data Mining, Novel K-Nearest Neighbour, Naive Bayes Training, Poisonous, Mushrooms.

DOI: 10.47750/pnr.2022.13.S04.058

INTRODUCTION

Every year, mushroom poisoning kills or affects a large number of individuals all around the world. A person's most valuable asset is his or her health. Keeping this in mind, we've put together a quick overview of mushroom poisoning research. The term mushrooms has been diversely portrayed by a couple of makers. They are the substantial, fungus bearing fruiting combinations of developments, which usually appear over the ground after storms. They are for the most part fill in forest areas seeing someone on woody bits of trees either as parasite, saprophyte or as symbionts in the dirt (Engel 2019). Mushrooms and parasites generally speaking are non-green animals lacking chlorophyll (Lima et al. 2021). They can't make their own food from essential inorganic materials, like water, carbon dioxide, and nitrates, using energy from the sun. The signs of mushroom hurting can vacillate from slight nausea/vomiting (Jane 2021), regurgitating to death (Taylor 2021). If we eat our favourite mushroom without first determining if it is edible or Poisonous, we risk death.

This research related work was presented and published in more than 41 indexed journals. Every year Data is doubled, however all of the beneficial statistics appears to be decreased. Area of data mining has arisen over decades to be able to cope with the problem. It has now no longer grown to be the most effective and vital studies area, however it has additionally grown to be one with a massive capacity in the actual world. The Multilayer Perceptron is a modeling and forecasting device that makes use of a neural community to version your information (Lintas et al. 2017). It may be used to categorize styles or to expect values out of your information. Because it makes use of a supervised learning technique, it calls for your information to incorporate goals for education in the community. Classification is one of the obligations accomplished via the information mining process (National Research Council et al. 2015). It is classified as one of the prediction strategies for a massive hidden information.

A version produced via information education has to have the ability to differentiate the class of a brand new information mushroom in those families (Tesoriero 2018). Mushroom, Agaricaceae, belongs to participants of a circle of relatives of fungi with gills (Charles 2020). It is constantly classified into two groups; suitable for harmful or not. The time period mushroom is constantly used to consult suitable for eatable species, even as the time period toadstool is used for poisonous species. However, the toadstool is constantly changed via means of poisonous mushrooms (Orywal et al. 2021).

Our team has extensive knowledge and research experience that has translate into high quality publications(Bhansali et al. 2021; Jayanth et al. 2021; Sudhakar, Ravel, and Perumal 2021; Sathiyamoorthi et al. 2021; Deepanraj et al. 2021; Raju et al. 2021; Arun Prakash et al. 2020; Kamath et al. 2020; Shanmugam et al. 2021; Rajasekaran et al. 2020; Adhinarayanan et al. 2020; Rajesh et al. 2020; Aurtherson et al. 2021). The research gap has been identified that the poisonous prediction is not done for several varieties of mushroom species, at the same time regional plantations are not considered. Hence the aim of this study is to increase the accuracy of predicting the likelihood of poisonous or eatable mushrooms and improve the prediction model using the K-Nearest Neighbour.

MATERIALS AND METHODS

This exploration study was completed at the Machine Learning Laboratory, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai. This examination concentrates on utilizing two groups of arrangement calculations utilized for the review. Group 1 and Group 2 are the KNN calculation and the Naive Bayes Training calculation respectively (Jyothi et al. 2020). Each sample size was anticipated utilizing the G power device with rendition and bringing about 51 sample sizes with 80% of G-power esteems and the limit two followed critical worth is set to 0.05 and the certainty stretch as 95%. The mushroom dataset which is to be credited for the proposed work is collected from kaggle.com , one of the more well known web-based networks for information researchers and Machine Learning experts (Ilias and Georgios 2019).

The Mushroom harmful or not dataset is to be imputed for the proposed work is collected from kaggle.com, one of the more popular online communities for data scientists and data mining practitioners. It allows users to search and find different datasets that they require. The dataset used here consists of 14 attributes and contains 12 features that can be used to predict mortality by harmful or not mushrooms. The mushrooms dataset has 5125 rows which consists of data for the symptoms that are related to poisonous which also includes duplicate, null and missing values.

K-Nearest Neighbour falls under the category of supervised algorithms. KNNs can be used for both classification and regression. In the classification KNN the decision variable is categorical. For implementing a KNN, the cost function used to evaluate the binary splits called the Gini Index should be calculated. The split creation is done with help of calculating the gini score, splitting the dataset and evaluating the all splits and this is done by creating a root node. It can start building the tree by first deciding when to stop 3 of 12 the growth of the tree by creating the terminal node by setting the maximum tree depth and second, by using recursive splitting. The terminal node is used for the final prediction and recursive splitting is a method used to build the tree.

The help has been picked on the grounds that it addresses a structure both intriguing from an Machine Learning point of view and from an implanted frame works viewpoint. A Novel K-Nearest Neighbour is a direct or non-straight classifier, which is a numerical capacity that can recognize two various types of items. These articles fall into classes, which isn't to be confused with a Python class Training a KNN can be shown with the accompanying pseudo code.

Pseudocode of the KNN algorithm

```
Inputs : Determine the various training and test data.
Outputs: Determine the calculated accuracy.
Select the optimal value of poisonous and eatable for KNN.
while(stopping condition is not met)do
        Implement KNN train steps for each data point.
        Implement KNN clasity for testing data points.
end while
Return accuracy
```

The following pseudocode is for the Naive Bayes algorithm to apply on the mushrooms dataset and also works with the tree model. The pseudocode will take the datasets as input and the final output of the pseudocode will be sent through the parameters Accuracy and the classification.

Pseudocode of the Naive Bayes Training algorithm

Step1: Initialize all parameters (Bo. B_1, etc.)
Step 2: Calculate (predict) dependent variable (h(x))
Step 3: Calculate poisonous function (poisonous(ha(z), y))
Step 4: Calculate gradient for poisonous function
Step 5: Update all parameters
Step 6: Repeat steps 2 to 5.

Statistical Analysis

At long last, the outcomes collected from the collecting one and collecting two calculations will be applied utilizing Statistical Package for Social Sciences (SPSS) with adaptation 26 (Rajathi and Chandran 2019). The autonomous example t-test was performed to analyze the exhibitions of the Naive Bayes Training and Novel K-Nearest Neighbour calculations. The independent variable is the forest attribute in the dataset and other twenty attributes such as odor, population, bruises etc are dependent variables for our study for mushroom prediction.

RESULTS

The K-Nearest Neighbour algorithm and Naive Bayes algorithm are compared with 51 samples by applying various 70% of training and 30% of testing datasets by varying the number of records of the dataset and the outcomes are depicted in Table 1 and the dataset consists of 100 rows where the accuracy of both the KNN and Naive bayes algorithm are obtained for 10 iterations. It is observed that the Novel K-Nearest Neighbour algorithm is significantly better than the Naive Bayes Training algorithm.

From Table 1, the statistical analysis of KNN Algorithm and Naive Bayes. Mean accuracy value, Standard deviation and accuracy for KNN and Naive Bayes algorithms are obtained for 10 iterations. It is observed that the KNN algorithm performed better than the Naive Bayes algorithm.

From Table 2, the Independent sample t-test for significance and standard error are determined. By statistical analysis KNN gives 1.42 standard deviation with 0.45 standard error and Naive bayes gives 4.12 standard deviations with 1.306 standard error the significant value is less than 0.5 that shows the proposed work gives better results.

In Table 3, observed after performing statistical analysis of 10 samples, KNN obtained 5.82 standard deviations with 1.38 standard error while the Naive bayes algorithm obtained 5.82 standard deviations with 1.38 standard error. The statistically significant difference between the two groups ($p=0.011$; $p<0.05$) showed that our hypothesis holds a good value. For changes in the input values (independent variable) the corresponding output values are changed (dependent variables).

The independent sample t-test was used to compare the accuracy of two algorithms and a statistically significant difference was noticed with $p<0.05$. The KNN model obtained 89.6% accuracy. When compared with the other algorithm's performance, the proposed KNN classifier achieved better performance than Naive Bayes.

The Red dots shown poisonous mushrooms and Green dots shown its eatable mushrooms. The symptom shown here is the rest of the models accuracy comparison as shown in Fig. 1.

The Red dots showed poisonous mushrooms and Green dots showed its eatable mushrooms as shown in Fig. 2.

The data taken for this research study will undergo preprocessing, a step to remove the unwanted, missing values and null values, it has been preprocessed. The results of mushroom analysis classification has accuracy for KNN algorithm having 90.57% and Naive bayes having 89.87% across the samples as shown in Fig. 3, and it shows that KNN is significantly better than the Naive bayes algorithm.

Finally, compare the mean accuracy of the Novel K-Nearest Neighbour and the Naive Bayes algorithm as shown in Fig. 4, and it shows that the Novel K-Nearest Neighbour is significantly better than the Naive Bayes algorithm.

DISCUSSION

In this article, we proposed a modified mushroom destructiveness unmistakable evidence method. The KNN strategy proposed by Zhou has an affirmation Accuracy of more than 98% with statistically significant difference between the two groups ($p=0.011$; $p<0.05$). Taking into account the incredible precision essentials, we separated three model plan models. Vital backslide yielded collecting results by looking at the reasonable components essential to perceive noxiousness. The exactness of the KNN technique is better than that of determined backslide. Appearing differently in relation to KNN and determined backslide, KNN achieved better results similar to precision. Therefore, KNN is a nice technique for normally recognizing whether a mushroom is harmful (Beaufront 2019).

Different typical mushroom toxicity affirmation procedures are as of now being utilized. These procedures use different responsibilities for choosing toxicity, yet they have different cutoff points, similar to low precision, unsuitable acknowledgment of dark toxic substances, the prerequisite for an extreme exploratory environment, and sufficient master data and complex test testing strategies. To avoid the constraints of these systems and apply them to little model data assessment, we used Data Mining. Rather than significant neural associations, which require mind blowing effort in hyperparameter tuning, KNN is much less difficult to get ready and can be applied to different sorts of data in different regions. The KNN estimation partakes in the going with benefits: (1) it has a fundamental plan; (2) it will in general be applied to datasets of different sizes; (3) the testing strategies and dealing with are direct; and (4) for our preliminaries, mushroom harmfulness is seen quickly. Component based learning and iterative classifiers in the KNN procedure have the best presentation among the three systems for data mining or machine learning used in this article. This customized ID procedure is suitable for nonprofessional conspicuous confirmation and for dark mushroom assortments (Gupta 2016).

Among the three data mining strategies, Novel K-Nearest Neighbour yielded the best accuracy. In any case, the robustness of the classifier ought to be improved. An avocation for this good may come from the setback of components, the component checked stem-surface-above-ring, or it may come from the real estimation (Gupta 2016). In like manner, further creating trustworthiness is a first concern when endeavoring to deal with the precision of the classifier. Since Novel K-Nearest Neighbour can be used for different kinds of datasets and affirmation reliant upon picture features is more useful than with various classifiers , this strategy for perceiving whether a mushroom is poisonous can be contacted. Regardless, there is at this point no dataset of mushroom poisonous or eatable (Peng et al. 2020).

CONCLUSION

The main aim of the study is to measure the accuracy in the classification of mushrooms data. This research study applied the KNN algorithm for detection of mushrooms poisonous or eatable achieved have been compared with the Naive Bayes Training algorithm. The results obtained show that the KNN algorithm has found 88.54% of accuracy on mushrooms data than the 82.72% of the Naive Bayes Training algorithm.

DECLARATIONS

Conflict of Interests

No conflict of interest in this manuscript.

Authors Contributions

Author MR was involved in data collection, data analysis, and manuscript writing. Author SMS was involved in conceptualization, data validation, and critical review of the manuscripts.

Acknowledgements

The authors would like to express their sincere 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. AAIC Technologies Private Limited, Hyderabad.
2. Saveetha University.
3. Saveetha Institute of Medical and Technical Sciences.
4. Saveetha School of Engineering.

REFERENCES

1. Adhinarayanan, Rajesh, Aravindh Ramakrishnan, Gopal Kaliyaperumal, Melvinvíctor De Poures, Rajesh Kumar Babu, and Damodharan Dillikannan. 2020. "Comparative Analysis on the Effect of 1-Decanol and Di-N-Butyl Ether as Additive with diesel/LDPE Blends in Compression Ignition Engine." Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, June, 1–18.
2. Arun Prakash, V. R., J. Francis Xavier, G. Ramesh, T. Maridurai, K. Siva Kumar, and R. Blessing Sam Raj. 2020. "Mechanical, Thermal and Fatigue Behaviour of Surface-Treated Novel Caryota Urens Fibre–reinforced Epoxy Composite." Biomass Conversion and Biorefinery, August. https://doi.org/10.1007/s13399-020-00938-0.
3. Aurtherson, P. Babu, Bhanu Teja Nalla, Karthikeyan Srinivasan, Kulmani Mehar, and Yuvarajan Devarajan. 2021. "Biofuel Production from Novel Prunus Domestica Kernel Oil: Process Optimization Technique." Biomass Conversion and Biorefinery, May. https://doi.org/10.1007/s13399-021-01551-5.
4. Beaufront, Louis Couturat Otto. 2019. International-English Dictionary. Wentworth Press.
5. Bhansali, Karan J., Kamlesh R. Balinge, Subodh U. Raut, Shubham A. Deshmukh, M. Senthil Kumar, C. Ramesh Kumar, and Pundlik

R. Bhagat. 2021. "Visible Light Assisted Sulfonic Acid-Functionalized Porphyrin Comprising Benzimidazolium Moiety for Photocatalytic Transesterification of Castor Oil." Fuel 304 (November): 121490.
6. Charles, Vera K. 2020. Introduction to Mushroom Hunting. Read Books Ltd.
7. Deepanraj, B., N. Senthilkumar, D. Mala, and A. Sathiamourthy. 2021. "Cashew Nut Shell Liquid as Alternate Fuel for CI Engine—optimization Approach for Performance Improvement." Biomass Conversion and Biorefinery, February. https://doi.org/10.1007/s13399-021-01312-4.
8. Engel, Michael S. 2019. "Termite Evolution: A Primal Knock on Wood or a Hearty Mouthful of Dirt." Current Biology: CB.
9. Gupta, P. K. 2016. Fundamentals of Toxicology: Essential Concepts and Applications. Academic Press.
10. Ilias, Chamatidis, and Spathoulas Georgios. 2019. "Machine Learning for All: A More Robust Federated Learning Framework." Proceedings of the 5th International Conference on Information Systems Security and Privacy. https://doi.org/10.5220/0007571705440551.
11. Jane, Dr Elizabeth. 2021. Mushroom and Pain Management: Extensive Guide to Use Psilocybin to Manage Pains.
12. Jayanth, Bellappu Venkat, Melvin Victor Depoures, Gopal Kaliyaperumal, Damodharan Dillikannan, Dilipsingh Jawahar, Kumaran Palani, and Ganesha Prasad Meravanigee Shivappa. 2021. "A Comprehensive Study on the Effects of Multiple Injection Strategies and Exhaust Gas Recirculation on Diesel Engine Characteristics That Utilize Waste High Density Polyethylene Oil." Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, June, 1–18.
13. Jyothi, S., D. M. Mamatha, Suresh Chandra Satapathy, K. Srujan Raju, and Margarita N. Favorskaya. 2020. Advances in Computational and Bio-Engineering: Proceeding of the International Conference on Computational and Bio Engineering, 2019, Volume 2. Springer Nature.
14. Kamath, Manjunath, Subha Krishna Rao, Jaison, Sridhar, Kasthuri, Gopinath, Sivaperumal, and Shantanu Patil. 2020. "Melatonin Delivery from PCL Scaffold Enhances Glycosaminoglycans Deposition in Human Chondrocytes – Bioactive Scaffold Model for Cartilage Regeneration." Process Biochemistry 99 (December): 36–47.
15. Lima, G. A., B. F. S. Barbosa, R. G. A. C. Araujo, B. R. Polidoro, G. V. Polycarpo, D. C. Zied, J. D. Biller, et al. 2021. "And Mushrooms as Alternative Additives to Antibiotics in Diets for Broilers Challenged with Spp." British Poultry Science 62 (2): 251–60.
16. Lintas, Alessandra, Stefano Rovetta, Paul F. M. Verschure, and Alessandro E. P. Villa. 2017. Artificial Neural Networks and Machine Learning – ICANN 2017: 26th International Conference on Artificial Neural Networks, Alghero, Italy, September 11-14, 2017, Proceedings. Springer.
17. National Research Council, Institute of Medicine, Board on Children, Youth, and Families, and Committee on the Science of Children Birth to Age 8: Deepening and Broadening the Foundation for Success. 2015. Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation. National Academies Press.
18. Orywal, Karolina, Katarzyna Socha, Patryk Nowakowski, Wojciech Zoń, Piotr Kaczyński, Barbara Mroczko, Bożena Łozowicka, and Maciej Perkowski. 2021. "Health Risk Assessment of Exposure to Toxic Elements Resulting from Consumption of Dried Wild-Grown Mushrooms Available for Sale." PloS One 16 (6): e0252834.
19. Peng, Sheng-Lung, Son Le Hoang, G. Suseendran, and D. Balaganesh. 2020. Intelligent Computing and Innovation on Data Science: Proceedings of ICTIDS 2019. Springer Nature.
20. Rajasekaran, S., D. Damodharan, K. Gopal, B. Rajesh Kumar, and Melvin Victor De Poures. 2020. "Collective Influence of 1-Decanol Addition, Injection Pressure and EGR on Diesel Engine Characteristics Fueled with diesel/LDPE Oil Blends." Fuel 277 (October): 118166.
21. Rajathi, A., and P. Chandran. 2019. SPSS For You. MJP Publisher.
22. Rajesh, A., K. Gopal, De Poures Melvin Victor, B. Rajesh Kumar, A. P. Sathiyagnanam, and D. Damodharan. 2020. "Effect of Anisole Addition to Waste Cooking Oil Methyl Ester on Combustion, Emission and Performance Characteristics of a DI Diesel Engine without Any Modifications." Fuel 278 (October): 118315.
23. Raju, P., K. Raja, K. Lingadurai, T. Maridurai, and S. C. Prasanna. 2021. "Glass/Caryota Urens Hybridized Fibre-Reinforced nanoclay/SiC Toughened Epoxy Hybrid Composite: Mechanical, Drop Load Impact, Hydrophobicity and Fatigue Behaviour." Biomass Conversion and Biorefinery, March. https://doi.org/10.1007/s13399-021-01427-8.
24. Sathiyamoorthi, Ramalingam, Gomathinayakam Sankaranarayanan, Dinesh Babu Munuswamy, and Yuvarajan Devarajan. 2021. "Experimental Study of Spray Analysis for Palmarosa Biodiesel-diesel Blends in a Constant Volume Chamber." Environmental Progress & Sustainable Energy 40 (6). https://doi.org/10.1002/ep.13696.
25. Shanmugam, Rajasekaran, Damodharan Dillikannan, Gopal Kaliyaperumal, Melvin Victor De Poures, and Rajesh Kumar Babu. 2021. "A Comprehensive Study on the Effects of 1-Decanol, Compression Ratio and Exhaust Gas Recirculation on Diesel Engine Characteristics Powered with Low Density Polyethylene Oil." Energy Sources, Part A: Recovery, Utilization, and Environmental Effects 43 (23): 3064–81.
26. Sudhakar, M. P., Merlyn Ravel, and K. Perumal. 2021. "Pretreatment and Process Optimization of Bioethanol Production from Spent Biomass of Ganoderma Lucidum Using Saccharomyces Cerevisiae." Fuel 306 (December): 121680.
27. Taylor, Dr Raphael. 2021. Magic Mushroom and Pain Management: Step by Step Guide on Using Magic Mushroom for Pain Relief and Everything You Need to Know.
28. Tesoriero, Heather Won. 2018. The Class: A Life-Changing Teacher, His World-Changing Kids, and the Most Inventive Classroom in America. Ballantine Books.

TABLES AND FIGURES

Table 1. Comparison between KNN and Naive Bayes Bayes algorithm with N=10 samples of the dataset with the highest performance of 90.57% and 89.87% in the sample (when N=1) using the dataset size = 100 and the 70% of training and 30% of testing data.

Sample (N)	Dataset size / Rows in %	KNN algorithm Accuracy in %	Naive Bayes algorithm Accuracy in %
1	100	90.57	89.87
2	90	90.09	89.23

3	80	89.76	88.67
4	70	89.31	88.18
5	60	88.93	87.87
6	50	88.27	87.29
7	40	87.95	86.76
8	30	87.35	85.98
9	20	86.87	85.21
10	10	86.32	84.34

Table 2. Statistical results of KNN and Naive Bayes Bayes algorithms. Mean accuracy value, standard deviation and standard error mean for KNN and Naive Bayes Bayes algorithms are obtained for 10 iterations. It is observed that the KNN (90.57%) algorithm performed better than the Naive Bayes Bayes (89.87%) algorithm.

Algorithms (Accuracy)	Sample (N)	Mean	Standard Deviation	Standard Error Mean
KNN algorithm	10	88.54	1.42	0.45
Naive Bayes Bayes algorithm	10	82.72	4.12	1.306

Table 3. The Independent sample t-test of the significance level KNN and Naive Bayes Bayes algorithms results with a significant value (p=0.011). Therefore both the KNN and the Naive Bayes Bayes algorithms have a significance level less than 0.05 with a 95% confidence interval.

Accuracy	Levene's Test for Equality of Variances		T-test of Equality of Means					95% of the confidence interval of the Difference	
	F	Sig.	t	df	Sig (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Equal Variance Assumed	7.94	0.011	4.21	18	0.001	5.82	1.38	2.918	8.723
Equal Variance Not Assumed	-	-	4.213	11.17	0.001	5.82	1.38	2.783	8.858

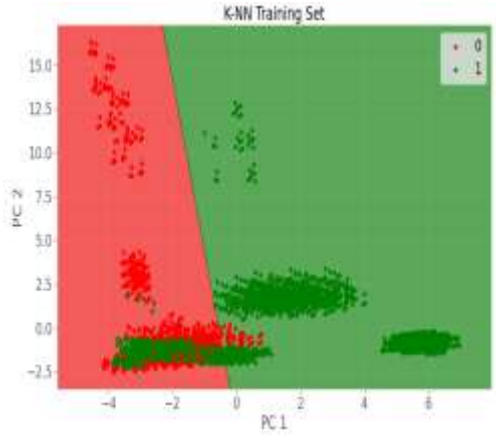

Fig. 1. Comparison of accuracy percentage of KNN (Poisonous in Red dots and Eatable in Green dots) across the samples.

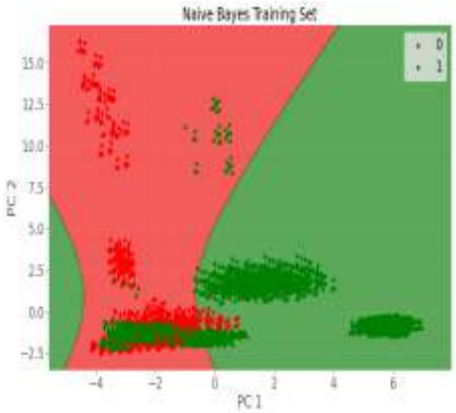

Fig. 2. Comparison of accuracy percentage of Naive Bayes (Poisonous in Red dots and Eatable in Green dots) across the samples.

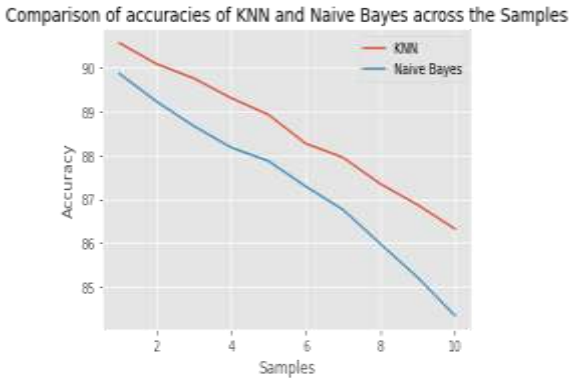

Fig. 3. Result of eatable mushroom classification using KNN (92.84) and Naive Bayes (89.87) across the samples.

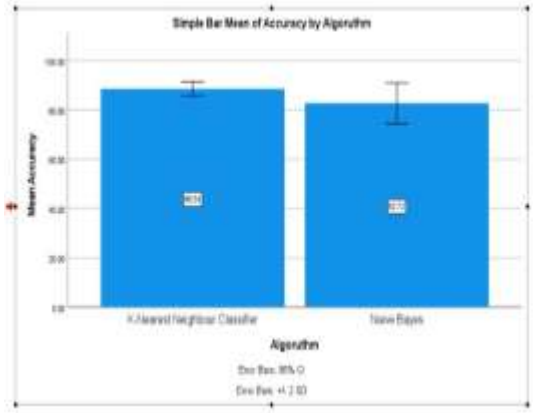

Fig. 4. Comparison of KNN algorithm and Naive Bayes Training in terms of mean accuracy. The mean accuracy of KNN is better than Naive Bayes Training and the standard deviation of KNN is slightly better than Naive Bayes Training algorithm. X-axis: (GROUPS) KNN vs Naive Bayes Training algorithm and Y axis: Mean accuracy of prediction ±2 SD.