Original Article

Prediction Of Insufficient Accuracy For Mushroom Classification Whether Poisonous Or Eatable Food Using Random Forest Training By Comparing Decision Tree 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.
maheswarareddyt18@saveetha.com

[2]Project guide Corresponding author, Department of Computer Science and Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Science, Saveetha University, Chennai, Tamilnadu, India. Pincode:602105.
saravananms@saveetha.com
DOI: 10.47750/pnr.2022.13.S04.276

Abstract

Aim: Mushrooms have different types and they contain principal supplements, for instance, proteins, nutrients, cell reinforcements, antioxidants and amino acids. There are a huge load of benefits of mushrooms. A wide range of mushrooms are not eatable. So before eating up mushrooms, it should be checked for consumable mushrooms. Careful affirmation and fitting distinctive confirmation of species are the super protected way to deal with ensuring harmful or not, and safeguard against expected accidents of consuming harmful one. **Resources and Techniques:** With 51 samples, two sets of calculations, and a G-power of 90.57 percent, the review was conducted. With late review finds, an edge of 0.05 percent, a certainty span of 95 percent mean, and a standard deviation, mushrooms were collected from a variety of online sources. As of right now, the Decision Tree Training calculation has 92.46 percent accuracy in projecting the Mushroom exactness rate. As a result, this study has to discover a unique Random Forest Training Machine Learning calculation that has a higher accuracy for accuracy reduction prediction. **Results:**The accuracy for toxic recognised proof was discovered to be 94.64 percent in this study employing the Novel Random Forest Training computation with a basic worth of two subsequent tests of 0.01 (p 0.05) and a 95 percent sameness stretch. **Conclusion:** This review reasons that the Novel Random Forest Training calculation on exactness is essentially better compared to the Decision Tree Training calculation.

Keywords: Machine Learning, Data Mining, Novel Random Forest Training, Decision Tree Training, Poisonous, Mushrooms.

INTRODUCTION

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 [1]. 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 [2]. The

Multilayer Perceptron is a modeling and forecasting device that makes use of a neural community to version your information [3]. It may be used to categorise 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.

One of the tasks carried out by the information mining process is classification. It is categorised as one of the prediction techniques for a significant amount of concealed information. A version created through information education must be able to distinguish the type of a new information mushroom in those familiar contexts (National Academies of Sciences, Engineering, and Medicine et al. 2016). The mushroom, or Agaricaceae, is a member of a group of gill-equipped fungi's distant cousins (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 eating species, even as the time period toadstool is used for poisonous species [4]. However, the fungus is continually changing due to harmful substances (Anderson 2018). This scientific work was presented and published in over 41 peer-reviewed publications.

Our team has extensive knowledge and research experience that has translate into high quality publications [5–13]. Numerous investigations have indicated that the performance Decision Tree gives lower accuracy in mushroom prediction, which has led to the identification of a research requirement. This study aims to enhance the prediction model and increase the precision of predicting whether mushrooms are poisonous or edible.

MATERIALS AND METHODS

In this work, two different sorts of categorization methods are used. The Novel Random Forest Training is used to compute the social events in Group 1, whereas the Decision Tree Training is used to calculate the social events in Group 2 (Koning and Smith 2017). Every model size was projected using the G-power device with version 3.1.10 and reaching 20 model sizes with 80% of G-power respects. The cutoff two following fundamental worth is set to 0.05, the conviction stretch is set to 95%. The mushroom dataset utilised in the proposed work was made available by Kaggle.com, one of the more well-known online communities for data scientists and Machine Learning specialists (Koning and Smith 2017; Das et al. 2021).

The dataset used in this study has 14 traits and 12 features that can be used to predict death from hazardous or edible mushrooms. The mushrooms dataset contains 305 rows of data for symptoms associated with dangerous mushrooms, as well as duplicate, null, and missing values.

Decision Tree Training is the definition of a supervised algorithm. The choice variable in training is categorical. The Gini Index, the cost function used to analyse binary splits, should be established prior to performing Decision Tree Training. The dataset is divided, the Gini score is calculated, and each split is then evaluated by creating a root node. The tree can be constructed by first choosing the maximum tree depth, creating the terminal node, and then employing recursive splitting. Recursive splitting is used to build the tree, and the terminal node is used to make the final prediction.

The help vector machine has been picked on the grounds that it addresses a structure both fascinating from an Machine Learning point of view and from an inserted frameworks viewpoint. A RF is a straight or non-direct classifier, which is a numerical capacity that can recognize two various types of items.

Pseudocode of the Random Forest Training Algorithm

Inputs: Determine the different training and test data.
Outputs: Determine the calculated accuracy.
Choose the best eatable and toxic value for RF.
While(The stopping condition has not been met.)do
Implement RF train steps for each data point.

Use RF clasity to test data points.
End while
Return accuracy

A decision tree is a supervised learning method that can be applied to classification and regression problems, however it is most typically employed to address classification problems. In this tree-structured categorization, leaf nodes indicate the conclusion, internal nodes reflect dataset attributes, and internal nodes reflect decision criteria.

Pseudocode of the Decision Tree Training algorithm

```
Import pandas as pd
Import Matplotlib.pyplot as plt
Import DecisionTreeClassifier
        import Decision Tree as Random Forest Training
          compare from sklearn.tree import DecisionTreeClassifier
           skip from sklearn.linear_model import Decision Tree
                Data extraction from sklearn import svm
                    initiate sklearn.metrics import accuracy score
                calculate sequence from sklearn.model_selection import train_test_split
             Find results from sklearn.feature_extraction.text import CountVectorizer
           count_vectorizer=CountVectorizer()
      cv=count_vectorizer.fit()
          cv.shape
             Random Forest Training=Random forest()
                Random Forest Training.fit(X_train,Y_train)
             prediction_Random Forest Training=Random Forest Training.predict(X_test)
        print(accuracy_score(prediction_Random Forest Training actiom_Random Forest Training,Y_test))
```

Statistical Analysis

Finally, the Statistical Package for Social Sciences (SPSS) with transformation 26 will be used to apply the results of the gatherings one and two computations. The outcomes of the Random Forest Training and Decision Tree Training computations were examined using the independent model t-test. The dataset's forest feature is the independent variable for our mushroom prediction study, while the dependent variables are the other 20 qualities including odour, population, bruising, and so forth.

RESULTS

The Random Forest method and Decision Tree algorithm are compared with 51 instances using a variety of 70 percent training and 30 percent testing datasets while changing the number of records in the dataset. Table 1 presents the outcomes. The accuracy of the RF and Decision Tree algorithms is assessed after 10 iterations, with the dataset having 100 rows. The Novel Random Forest method outperforms Decision TreeTraining.

Table 1 shows the statistical analysis of the RF Algorithm and Naive Bayes. Over the course of ten iterations, the mean accuracy value, standard deviation, and accuracy for the RF and Decision Tree algorithms are calculated. The RF algorithm outperformed the Decision Tree technique.

The Independent sample t-significance test's and standard error are displayed in Table 2. Statistical significance was defined as two-tailed significant p values less than 0.05, and 95% confidence intervals were

calculated. In a statistical analysis, Decision Tree has a standard deviation of 4.12 and a standard error of 1.306 compared to RF's 1.42 and 0.45, respectively. The proposed work delivers superior results because the two-tailed significant value is less than 0.5.

Following statistical analysis of 10 samples, the Decision Tree algorithm obtained 5.82 standard deviations with 1.38 standard error in Table 3, whereas RF also obtained the same results with 5.82 standard deviations. Our hypothesis is supported by the two-tailed significance value (p=0.045), which is less than 0.05. The connected output values vary when the input values (an independent variable) change (dependent variables).

The accuracy of the two algorithms was compared using the independent sample t-test, and a statistically significant difference was found at p 0.05. The RF model had an accuracy rating of 89.6%. The proposed RF classifier outperforms Decision Tree in comparison to the other methods.

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 collected for this research study will be preprocessed, which is a phase in which unwanted, missing, or null information is removed. As shown in Fig. 3, the RF algorithm's accuracy of mushroom analysis categorization is 92.87 percent and the Decision Tree technique's accuracy is 89.32 percent across all samples, demonstrating that the RF algorithm outperforms the Decision Tree approach.

Finally, it is evident that the Novel Random Forest surpasses the Decision Tree method when comparing the mean accuracy of the two algorithms, as shown in Fig. 4.

DISCUSSION

In this article, we proposed a changed mushroom ruinous tendency unquestionable proof strategy. The novel Random Forest Training system proposed by Zhou has an insistence accuracy of over 98%. Considering the fantastic accuracy fundamentals, we isolated three model arrangement models. Fundamental loss of faith yielded assembling results by taking a gander at the sensible parts fundamental to see harmfulness. The precision of the Random Forest Training method is superior to not really set in stone break faith. Showed up distinctively corresponding to not set in stone break faith, Random Forest Training accomplished better outcomes like ID accuracy. Consequently, Random Forest Training is a pleasant procedure for regularly perceiving whether a mushroom is noxious [14].

Distinctive commonplace mushroom poisonousness certification methods are at this point being used [15]. These systems utilize various responsibilities regarding picking harmfulness, yet they have distinctive cutoff focuses, like low accuracy, inadmissible affirmation of dull poisonous substances, the essential for a super exploratory climate, and adequate expert information and complex test testing procedures. To keep away from the requirements of these frameworks and apply them to minimal model information appraisal, we utilized Data Mining. Maybe than huge neural affiliations, which require mind blowing exertion in hyperparameter tuning, Random Forest Training is substantially less hard to prepare and can be applied to various kinds of information in various areas. The Random Forest Training assessment participates in the going with benefits: (1) it has a crucial arrangement; (2) it will overall be applied to datasets of various sizes; (3) the testing methodologies and management are immediate; and (4) for our starters, mushroom harmfulness is seen rapidly [16]. Part based learning and iterative classifiers in the Random Forest Training technique have the best show among the three frameworks for Data Mining utilized in this article. This modified ID strategy is reasonable for nonprofessional prominent affirmation and for dull mushroom assortments [14,17].

Among the three Machine Learning systems, Random Forest Training yielded the best exactness. Regardless, the power of the classifier should be improved. A diversion for this good might come from the difficulty of parts, the part checked stem-surface-above-ring, or it might come from the genuine assessment. In like way, further reliability is a first concern when attempting to manage the accuracy of the classifier. Since Random Forest Training can be utilized for various types of datasets and confirmation dependent upon picture highlights is more helpful than with different classifiers, this system for seeing whether a mushroom is poisonous can be reached with picture insistence. Notwithstanding, there is now no dataset of mushroom pictures [4,14,17].

CONCLUSION

The primary goal of the research is to assess the accuracy of mushroom data classification. This study compared the Random Forest Training algorithm to the Decision Tree Training algorithm for detecting toxic or non-harmful mushrooms. The results reveal that the Random Forest Training algorithm found 92.84 percent accuracy on mushrooms data, compared to 89.82 percent for the Decision Tree Training method.

REFERENCES

1. Shugan I, Chen Y-Y, Hsu C-J. The flood bore problem and the mushroom jet formation in the dam-break flow. doi:10.5194/egusphere-egu2020-7515

2. Ziółkowski C, Kelner JM, Krygier J, Chandra A, Prokeš A. Radio Channel Capacity with Directivity Control of Antenna Beams in Multipath Propagation Environment. Sensors . 2021;21. doi:10.3390/s21248296

3. Mumtaz S, Rodriguez J, Dai L. mmWave Massive MIMO: A Paradigm for 5G. Academic Press; 2016.

4. Menser GP. Hallucinogenic and Poisonous Mushroom Field Guide. Ronin Publishing; 2016.

5. Valsalan P, Sriramakrishnan P, Sridhar S, Latha GCP, Priya A, Ramkumar S, et al. Knowledge based fuzzy c-means method for rapid brain tissues segmentation of magnetic resonance imaging scans with CUDA enabled GPU machine. J Ambient Intell Humaniz Comput. 2020. doi:10.1007/s12652-020-02132-6

6. Dhaya chandhran KS, Jothilakshmi M, Mohan A, Tholkapiyan M. Performance evaluation and R-value analysis for thermally insulated PCM wall with fluted sheets. Materials Today: Proceedings. 2020;22: 912–918.

7. Saravanan A, Kumar PS, Yaashikaa PR, Karishma S, Jeevanantham S, Swetha S. Mixed biosorbent of agro waste and bacterial biomass for the separation of Pb(II) ions from water system. Chemosphere. 2021;277: 130236.

8. Indiramma P, Sudharani C, Needhidasan S. Utilization of fly ash and lime to stabilize the expansive soil and to sustain pollution free environment – An experimental study. Materials Today: Proceedings. 2020;22: 694–700.

9. Ezhilarasan D. Dapsone-induced hepatic complications: it's time to think beyond methemoglobinemia. Drug Chem Toxicol. 2021;44: 330–333.

10. Chupradit S, Jalil AT, Enina Y, Neganov DA, Alhassan MS, Aravindhan S, et al. Use of Organic and Copper-Based Nanoparticles on the Turbulator Installment in a Shell Tube Heat Exchanger: A CFD-Based Simulation Approach by Using Nanofluids. J Nanomater. 2021;2021. doi:10.1155/2021/3250058

11. Venu H, Appavu P. Combustion and emission characteristics of tamarind seed biodiesel-diesel blends in a compression ignition engine. Int J Ambient Energy. 2021;42: 1441–1446.

12. Prabhu A, Venkata Ramanan M, Jayaprabakar J, Harish V. Experimental investigation of emission characteristics on ricebran biodiesel–alcohol blends in a diesel engine. Int J Ambient Energy. 2021;42: 1123–1128.

13. Raja G, Chandramohan D, Gnanavel BK, Sathish T. Effect of inter-facial coupled contact forces in the multilayered pacemaker lead cable. PROCEEDINGS OF INTERNATIONAL CONFERENCE ON RECENT TRENDS IN MECHANICAL AND MATERIALS

ENGINEERING: ICRTMME 2019. AIP Publishing; 2020. doi:10.1063/5.0024971

14. Bellettini MB, Bellettini S, Fiorda FA, Pedro AC, Bach F, Fabela-Morón MF, et al. Diseases and pests noxious to Pleurotus spp. mushroom crops. Rev Argent Microbiol. 2018;50: 216–226.

15. Suetsugu K, Okamoto T, Kato M. Mushroom attracts hornets for spore dispersal by a distinctive yeasty scent. Ecology. 2019. doi:10.1002/ecy.2718

16. Richardson NE. Growing Mushroom Indoors (Psilocybin Mushroom):: A Straight-Forward and Reliable Guide to Growing Psilocybin Mushroom Fast; Plus Meaningful Tips That Will Assist You in the Course of Cultivation. Createspace Independent Publishing Platform; 2018.

17. Song W, Zhao L, Tao Y, Guo X, Jia J, He L, et al. The interruptive effect of electric shock on odor response requires mushroom bodies in Drosophila melanogaster. Genes Brain Behav. 2019;18: e12488.

TABLES AND FIGURES

Table 1. Comparison of the Random Forest and Decision Tree Bayes algorithms using N=10 dataset samples, with maximum performance of 92.84 percent and 89.32 percent in the sample (when N=1) using dataset size = 100 and 70% training and 30% testing data.

Sample (N)	**Dataset size / Rows in %**	**Random Forest algorithm Accuracy in %**	**Decision Tree algorithm Accuracy in %**
1	100	92.84	89.32
2	90	92.21	88.78
3	80	91.86	88.12
4	70	91.19	87.79
5	60	90.56	87.15
6	50	90.05	86.83
7	40	89.56	86.23
8	30	88.75	85.65
9	20	88.22	85.10
10	10	87.56	84.65

Table 2. Statistical outcomes of the Random Forest and Decision Tree Bayes algorithms The mean accuracy value, standard deviation, and standard error mean for the Random Forest and Decision Tree Bayes algorithms are calculated after 10 iterations. It was revealed that the Random Forest approach outperformed the Decision Tree Bayes algorithm (89.87 percent) (86.16 percent).

Algorithms (Accuracy)	Sample (N)	Mean	Standard Deviation	Standard Error Mean
Random Forest algorithm	10	89.87	2.15	0.68
Decision Tree algorithm	10	86.16	1.97	0.62

Table 3. The Independent sample t-test of the significance level Random Forest and Decision Tree Bayes algorithms results with two tailed significant values (p=0.01). Therefore both the Random Forest and the Decision Tree 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	0.147	0.706	4.02	18	0.01	3.71	0.923	1.77	5.653
Equal Variance Not Assumed	-	-	4.02	17.86	0.01	3.71	0.923	1.77	5.654

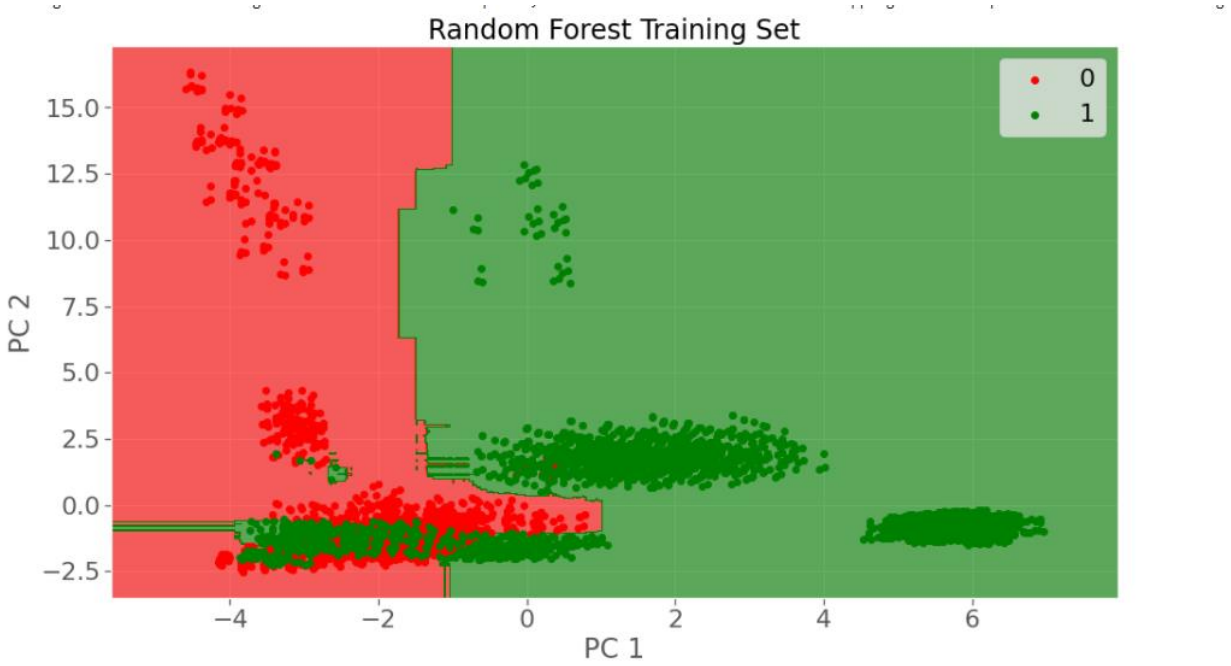

Fig. 1. Comparison of accuracies of Random Forest Training (Poisonous in Red dots and Eatable in Green dots) across the samples.

Fig. 2. Comparison of accuracies of Decision Tree Training (Poisonous in Red dots and Eatable in Green dots) across the samples.

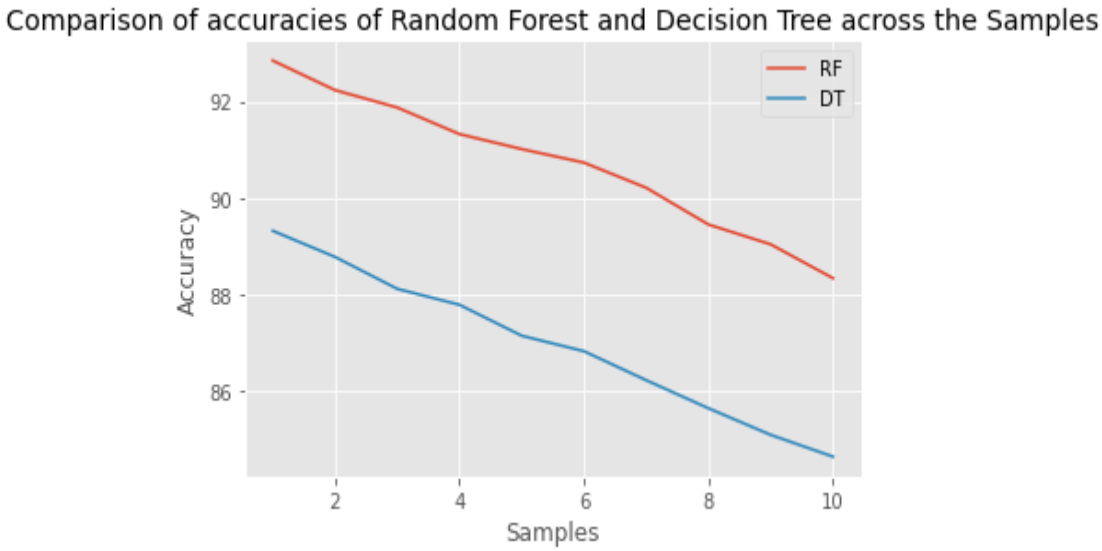

Fig. 3. Result of eatable mushroom classification using RF (92.84) and DT (89.32) across the samples.

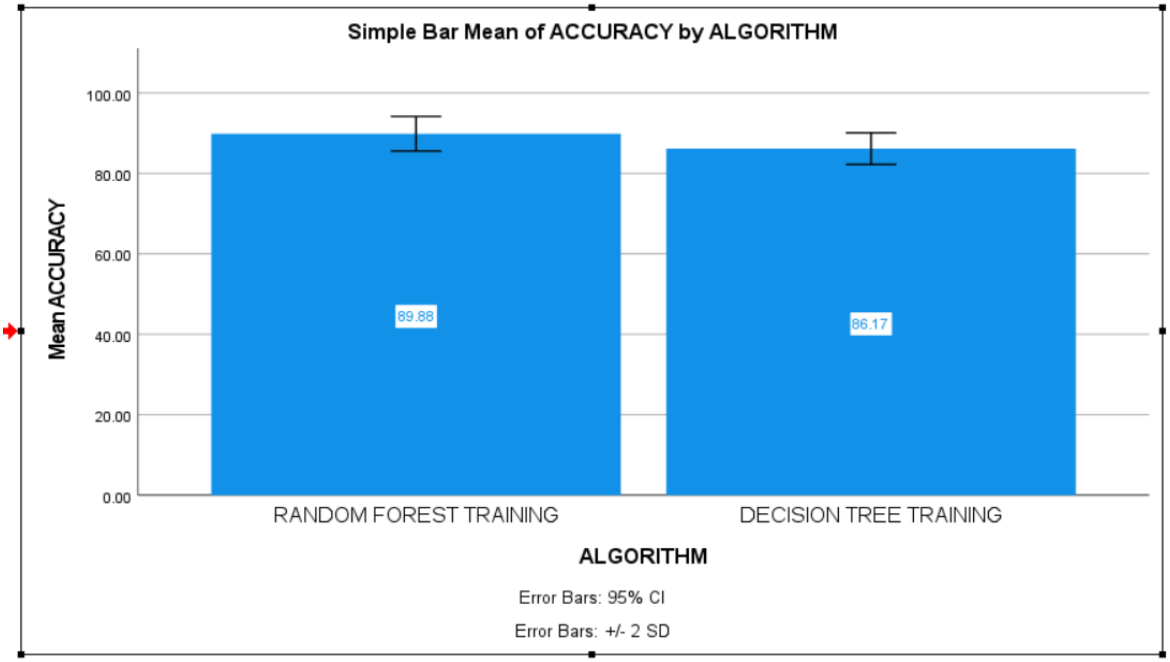

Fig. 4. Comparison of Random Forest Training algorithm and Decision Tree Training in terms of mean accuracy. The mean accuracy of Random Forest Training is better than Decision Tree Training and the standard deviation of Random Forest Training is slightly better than Decision Tree Training algorithm.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