**Title page :**

Paddy Crop Selection and Yield Forecasting Accuracy Using Random Forests over Decision Tree

**A.T.KONDA REDDY1 , AFREEN BANU2**

A.T.KondaReddy1

Research Scholar,

Department of Computer Science Engineering,Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin Code: 602105.

[ambavaramtirumalakondareddy0642.sse@saveetha.com](mailto:ambavaramtirumalakondareddy0642.sse@saveetha.com)

Afreen Banu2

Research Guide, Corresponding Author,

Department of Networking Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamil Nadu, India, Pin Code: 602105.

[afreenbanu.sse@saveetha.com](mailto:afreenbanu.sse@saveetha.com)

**keywords:**Paddy crop,crop analysis,Random Forests,Decision treeAlgorithm,Crop analysis,selection,Agricultural data,Machine learning,Decision support system,Agricultural sustainability,Predictive modelling.

**ABSTRACT:**

**Aim:** The goal is to evaluate and compare the efficacy in crop selection and yield prediction by using Random Forests and Decision tree Algorithm to improve the crop selection and yield prediction in paddy crops. Through this comparative evaluation, our goal is to discern which set of rules yields in crop to improve the benefits, contributing valuable insights to the optimization of crop choice techniques.**Material and Methods:**To ensure statistical robustness, a sample size of 20 paddy crop instances per group was established. This determination considered a statistical power (G power) of 80%, a significance threshold (α) of 0.05, and a confidence interval (CI) of 95%. This sample size calculation aimed to provide adequate statistical power to detect significant differences in crop selection and yield prediction between the Random Forests and Decision Algorithms.**Results:**The study comparing the predictive performance of Random Forest and Decision tree Algorithms for paddy crop selection and yield prediction revealed noteworthy disparities. Utilising IBM SPSS, key metrics including accuracy, precision, recall, and F1 score were evaluated. The Independent Sample t-test indicated a statistically significant differentiation between the two algorithms, with a calculated p-value of 0.001 (p<0.05). Specifically, the Random Forest algorithm exhibited a statistically higher accuracy level compared to Decision tree Algorithm. These findings offer valuable insights into the relative effectiveness of the algorithms in paddy crop crop selection and yield prediction, offering guidance for stakeholders to optimise agricultural practices**.Conclusion:**Through our exploration of paddy crop selection and yield prediction employing the Random Forest Algorithm and Decision tree Algorithm, we've got exposed significant strides in accuracy enhancement. The Random Forest Algorithm, with its superior performance, emerges as a beacon for unique crop selection and yield prediction, promising improved yield consequences. Our findings underscore the pivotal role of superior algorithms in shaping the destiny of agriculture. By leveraging techniques like Random Forest and Decision tree Algorithm in destiny research, we are able to further refine predictive skills, in the end empowering stakeholders with the insights needed to pressure sustainable farming practices and maximise crop selection and yield capability.

**INTRODUCTION:**

Decision Trees and Random Forests are machine learning Algorithms that can be used for yield forecasting and crop selection in paddy crop in agriculture. However, Random Forests offer several advantages over Decision Trees, in comparison of accuracy of [(Pradeep et al. 2019)](https://paperpile.com/c/5fkdxM/DknS)crop[(Yang et al. 2024)](https://paperpile.com/c/5fkdxM/bQv3) selection and [(Mousavi et al. 2023)](https://paperpile.com/c/5fkdxM/Szis)yield prediction.

Paddy[(Su et al. 2024)](https://paperpile.com/c/5fkdxM/O1In) is a staple crop in lots of elements of the world, and its cultivation is affected by different factors[(Kim et al. 2013)](https://paperpile.com/c/5fkdxM/CxcY), such as [(Pradeep et al. 2019)](https://paperpile.com/c/5fkdxM/DknS)soil type[(Mousavi et al. 2023)](https://paperpile.com/c/5fkdxM/Szis), weather[(Yang et al. 2024)](https://paperpile.com/c/5fkdxM/bQv3) conditions, and water availability. To ensure a hit paddy crop yield, it is crucial to select the proper crop range primarily based at the actual-time conditions.Real-time [(Yoshida 1981)](https://paperpile.com/c/5fkdxM/3Ojn)paddy[(Meena 2019)](https://paperpile.com/c/5fkdxM/TrS9) crop choice includes making knowledgeable decisions primarily based on up-to-date information and facts. This can consist of monitoring [(Zhan et al. 2024)](https://paperpile.com/c/5fkdxM/69S4)weather patterns[(Fang, Chang, and Lin 2023)](https://paperpile.com/c/5fkdxM/P3Ae), soil moisture stages, and available water sources. With the help of current era and records analysis tools, farmers can make informed choices about which [(Shu et al. 2024)](https://paperpile.com/c/5fkdxM/KtXY)paddy[(Zhao et al. 2024)](https://paperpile.com/c/5fkdxM/5DtV) crop variety to choose for his or her precise place and situations.

Paddy, also known as rice, is a staple crop in many parts of the world. It is a type of grass that grows in flooded fields and is used to produce rice grains. [(Zhao et al. 2024)](https://paperpile.com/c/5fkdxM/5DtV)Paddy cultivation requires specific conditions, including adequate soil moisture, water availability, and suitable weather conditions. Farmers must consider these factors when selecting paddy crop varieties to ensure a successful yield. Real-time paddy crop selection involves making informed decisions based on up-to-date data and analysis, which can help optimize crop yields and reduce the risks associated with crop failures. By using modern technology and data analysis tools, farmers can monitor weather patterns, soil moisture levels, and available water resources to make informed decisions about which paddy crop varieties to plant and how to manage their crops throughout the growing season**.**

Paddy[(Pradeep et al. 2019)](https://paperpile.com/c/5fkdxM/DknS) crop selection and yield forecasting are important elements of agriculture which could assist farmers optimize crop yields and decrease risks associated with crop failures. In latest years, facts evaluation gear consisting of system studying have emerge as increasingly famous in these fields, providing greater correct predictions and selection-making skills.

Real-time paddy crop choice and yield forecasting involve the usage of up-to-date information and information to make knowledgeable choices approximately which paddy [(Jin et al. 2024)](https://paperpile.com/c/5fkdxM/9t7a)crop types[(Sawatraksa et al. 2023)](https://paperpile.com/c/5fkdxM/FCsz) to plant[(Sun et al. 2022)](https://paperpile.com/c/5fkdxM/JsZN) and while to plant them. This can encompass monitoring weather styles, soil moisture stages, and to be had water assets.

Machine gaining knowledge of algorithms which includes Decision Trees and Random Forests may be used to enhance the accuracy of yield forecasting. Decision Trees are a sort of supervised system gaining knowledge of set of rules that can be used for both type and regression tasks. They work by using recursively partitioning the input information into smaller and smaller subsets until a preventing criterion is met.

Random Forests, alternatively, are an ensemble getting to know approach that combines a couple of Decision Trees to enhance the accuracy and robustness of the predictions. By averaging the predictions of a couple of Decision Trees, Random Forests can reduce the effect of overfitting and enhance the overall accuracy of the predictions.

In the context of paddy crop choice and yield forecasting, Random Forests may be used to research massive datasets of ancient and actual-time statistics to predict the most efficient paddy crop varieties and yields for precise locations and conditions. By using actual-time facts and analysis, farmers can make informed choices approximately which paddy crop types to plant and how to control their plants during the growing season.Selecting the right paddy crop variety is critical for optimizing crop yields and lowering dangers associated with crop failures. By leveraging real-time records and analysis, farmers could make knowledgeable choices approximately which paddy crop varieties to plant, while to plant them, and how to manage their plants in the course of the developing season.

One of the main blessings of Random Forests is that they're an ensemble getting to know technique that combines multiple Decision Trees to improve the accuracy and robustness of the predictions. By averaging the predictions of more than one Decision Trees, Random Forests can reduce the effect of overfitting and offer a extra correct and reliable prediction.

In the context of Paddy crop choice and yield forecasting, Random Forests may be used to research massive datasets of historical and actual-time statistics to expect the foremost paddy crop varieties and yields for particular locations and conditions. By the usage of real-time statistics and analysis, farmers can make informed choices about which paddy crop varieties to plant and how to control their plants during the growing season.

Real-time information analysis equipment can help farmers make more informed decisions approximately paddy crop selection and yield forecasting. By using device learning algorithms like Random Forests, farmers can leverage the energy of huge statistics to optimize crop yields and decrease risks associated with crop failures.

In the context of Paddy Crop Selection and Yield Forecasting, Decision Trees may be used to classify distinctive forms of paddy vegetation based on various attributes together with soil type, weather situations, and crop records. However, Decision Trees may be prone to overfitting, where the model or version becomes too complex and performs plays poorly on unseen statistics where compared to both Algorithms.

Paddy crop selection involves figuring out the most suitable variety of rice for cultivation based on various factors consisting of soil kind, climate conditions, pest resistance, and marketplace (sometimes it may differ like depending upon the government rate for farmers to overcome to achieve their goals). Decision Tree is a widely used algorithm for this venture, but it could suffer from overfitting and shortage generalisation ability.

To overcome these limitations, Random Forests is very much useful to achieve prediction accuracy in crop selection and yield prediction . It is an ensemble studying approach that combines a couple of decision timber to make predictions. Each tree inside the wooded area is constructed on a random subset of the education information and functions. This randomness enables to reduce overfitting and improve the version's accuracy.

Paddy Crop Selection and Yield Forecasting Accuracy the use of Random Forests over Decision Trees is a effective technique for maximizing crop yields and profits while ensuring meals safety. By combining the strengths of Decision Trees and ensemble studying, Random Forests can offer extra accurate and sturdy predictions, making them a precious tool for farmers, agronomists, and comparison of both Algorithm in IBM SPSS by using dataset.

The advantage of using Random Forests over Decision Trees lies in its ability to handle large datasets with high dimensionality while maintaining good prediction performance. Moreover, it provides feature importance measures that help identify which factors contribute most significantly to crop selection or yield forecasting.

Decision Trees and Random Forests are both famous machine learning algorithms used for type and regression duties. Decision Trees are a kind of flowchart-like structure, wherein every internal node represents a feature (or characteristic), each branch represents a choice rule, and each leaf node represents an final results.

In conclusion, utilizing Random Forests over Decision Trees enhances paddy crop selection by using considering multiple factors or elements simultaneously and improves yield forecasting accuracy through mitigating overfitting issues. This approach proves precious in optimizing agricultural planning and selection making for sustainable farming practices in practical manner to achieve their sustainable farming farms for farmers.

**MATERIALS AND METHODS**

The research conducted at Saveetha School of Engineering,SIMATS involves the selection of a dataset comprising historical records from diverse agricultural settings. There are two groups identified.Group 1 KNN Algorithm and Group 2 is Random forest Algorithm A size of 20 (N=20) samples, representative of different geographical locations and varying agricultural practices, are collected and calculated from SPSS analysis of SSE. These samples include data on paddy crop selection and yield prediction , weather patterns, soil characteristics, and historical farming techniques,and the dataset is collected from kaggle.com.

**RANDOM FOREST ALGORITHM:**

The Random Forest algorithm is an ensemble learning technique that combines the predictions of multiple individual models to improve overall performance and generalisation. Specifically, it constructs a collection of decision trees during training and outputs the mode of the classes (classification) or the mean prediction (regression) of the individual trees. The Random Forest algorithm was introduced by Leo Breiman and Adele Cutler in 2001 and has since become a widely used and powerful machine learning algorithm.The Random Forest algorithm is an ensemble learning method that operates by constructing a multitude of decision trees during training and outputs the class that is the mode of the classes (classification) or the mean prediction (regression) of the individual trees. It was introduced by Leo Breiman in 2001 and has since become one of the most popular and powerful algorithms in machine learning.Following are the steps to perform the algorithm:

Step 1: Import necessary libraries.

Step 2: Load or prepare your database.

Step 3: Split the dataset into training and testing sets.

Step 4: Create and train the Random Forest model.

Step 5: Make predictions on the test set.

Step 6: Evaluate the model's performance.

Step 7: Feature Importance (Optional).

Step 8: Fine-tune the model (Optional).

**DECISION TREE ALGORITHM:**

The decision tree algorithm is a supervised machine learning algorithm used for classification and regression tasks. It builds a tree-like model of decisions and their possible consequences. The algorithm iteratively splits the dataset into subsets based on specific features to create decision rules that can predict the target variable.A decision tree consists of nodes and branches. The nodes represent features or attributes, and the branches represent possible outcomes or values for those attributes. The tree is built recursively by splitting the data based on different attributes at each node, with the goal of creating homogeneous subsets of data in terms of the target variable.The process starts with a root node that represents the entire dataset. At each step, an attribute is selected to split the data into two or more subsets based on some criteria, such as entropy or Gini impurity. This splitting process continues until a stopping condition is met, such as reaching a maximum depth or when further splits do not provide significant improvement.Ross Quinlan is a well-known researcher in the field of machine learning and data mining, and his contributions to decision trees have had a significant impact on the field.

Step 1: Import necessary libraries.

Step 2: Load or prepare your database.

Step 3: Split the dataset into training and testing sets.

Step 4: Create and train the decision tree algorithm model.

Step 5: Make predictions on the test set.

Step 6: Evaluate the model's performance.

Step 7: Feature Importance (Optional).

Step 8: Fine-tune the model (Optional).

Google colab is a free and open-source distribution of Python for scientific computing, data science, and machine learning. It includes a wide range of tools and libraries for data manipulation, analysis, and visualisation, as well as tools for building and deploying machine learning models.

**Statistical Analysis:**

Using the SPSS statistical package, the analysis of mean accuracy by using Random forest algorithm and Decision tree algorithm was carried out by applying an independent sample t-test to obtain the accuracy of 94.4%. The speed and file size are independent variables and type of the file is dependent variables.

**Results:**

Our study's results are summarized using a comparative bar graph, illustrating the performance of the Random Forest and Decision Tree algorithms in paddy crop selection and yield forecasting accuracy to improve the accuracy in yield . The x-axis represents distinct algorithm groups, distinguishing between Random Forest and Decision Tree for easy comparison. Meanwhile, the y-axis measures predictive accuracy in percentage.

Utilizing the Random Forest algorithm, the mean accuracy achieved is 94.4090%, with a standard deviation of 1.97338 and a standard error of the mean of 0.62404. This demonstrates the superior predictive capability of Random Forest in accurately paddy crop selection and yield forecasting.

On the other hand, employing the Decision Tree algorithm yields a mean accuracy of 84.1610%, with a standard deviation of 1.12583 and a standard error of the mean of 0.32602. While slightly lower than Random Forest, Decision Tree still exhibits substantial predictive performance in paddy crop selection .This visible illustration presents an impartial and impactful evaluation of the relative overall performance of these algorithms in paddy crop selection and yield forecasting accuracy. This research will helps to select the paddy crop selection and yield forecasting for farmers and future researches to involves their data to analysis for their research.

**Discussions:**

In our comparative evaluation of machine learning algorithms for paddy crop selection and yield forecasting accuracy, we focused on two prominent models: Random Forest and Decision tree algorithm . The results revealed that the Random Forest algorithm exhibited superior performance, achieving an accuracy of 94.4090%, compared to Decision tree Algorithm, which attained 84.1610%. Additionally, the standard deviation for the Decision tree algorithm was 1.12583, with a standard error mean of 0.35602. Conversely, for the Random Forest algorithm, the mean was 94.4090, with a standard deviation of 1.97338 and a standard error mean of 0.62404.

In x-axis represents number of groups and y-axis represents the % of measured accuracy of both groups in data ,usually the data is in completely based upon the data set and code for given data in google colab it gives a accuracy values and it completely based upon the given information of the google colab.

The investigation underscores the effectiveness of employing the Random Forest algorithm for capturing intricate patterns within the dataset, resulting in more precise forecasts of paddy crop selection and yield forecasting. The heightened accuracy of Random Forests demonstrates its resilience against overfitting and its ability to discern complex inter relationships among various input features.

The bar graph vividly depicts the superiority of Random Forests over Decision tree Algorithm, with the former outperforming by a margin of 10.2480%. This significant difference underscores the robustness of the Random Forest algorithm in capturing complex relationships within the dataset, leading to more accurate predictions of paddy crop selection and yield forcasting.

The findings of this study advocate for the adoption of Random Forests for analyzing paddy crop crop selection and yield forecasting the data analysis , offering a more accurate and reliable alternative to Decision tree algorithm. As precision agriculture continues to play a pivotal role in crop selection management to farmers and researchers, leveraging advanced algorithms like Random Forests can significantly contribute to optimizing crop yields and ensuring the sustainability of agricultural practices for farmers and future data analysis.

The discussion surrounding this contrast delves into the nuanced intricacies of the algorithms. Random Forests, as an ensemble learning approach, excels in handling diverse and correlated features, which contributes to its superior performance. The graph serves as a visual aid to explain the impact of algorithm selection on the analysis of paddy crop selection and yield forecasting accuracy, underscoring the importance of selecting models that align with the characteristics of the agricultural dataset.

In conclusion, the bar graph reinforces our findings, providing a clear visual representation of the substantial difference in accuracy between Random Forest and Decision tree Algorithm. This discussion enhances our understanding of algorithmic choices in the analysis of paddy crop crop selection and yield forecasting accuracy analysis but also serves as a practical guide for stakeholders seeking to adopt the most effective machine learning techniques in precision agriculture for high-accuracy data analysis

**Conclusion:**

Random Forest Algorithm has 94.4090 % accuracy and Decision tree Algorithm has 84.1610% accuracy. The difference in accuracy between Random Forest algorithm and Decision tree algorithm has set of rules is 94.4090 % and 84.1610% respectively. The clean difference in accuracy is shown in the bar graph below. In conclusion, our study on the analysis of Paddy Crop using Machine learning of Algorithms shows that Random Forest algorithm is more accurate than Decision tree algorithm.

The outcomes propose that Random Forest's potential to handle complicated relationships within the dataset contributed to its higher predictive accuracy values in given data . This finding underscores the importance of set of rules selection in crop analysis.

In sensible terms, the bar graph serves as a honest instance of the enormous distinction in performance among Random Forest and Decision tree Algorithm. This end affords a concise perception into the effectiveness of machine learning algorithms in optimizing paddy crop selection and yield forecasting , with Random Forest emerging as the extra dependable choice in our study.random forest can gives a perfect comparison in between Decision tree Algorithm and it can give a mean values.

**DECLARATIONS:**

**Conflicts of Interest**

No conflict of interest in this manuscript.

**Author’s contributions**

Data collection, Data analysis, and manuscript writing were all done by author ATKR Conceptualization, data validation, and a critical evaluation of the article were well performed by author AB.

**Acknowledgment**

Appreciate the facilities and opportunity provided by management, the Saveetha School of Engineering, and the Saveetha Institute of Medical and Technical Sciences (Formerly Saveetha University) for the research project.

**Funding**

Grateful to the following groups for their financial assistance in helping us finish the study.

* The Big Event, Coimbatore.
* I collected some data from my village agriculture officer(RBK).
* Saveetha School of Engineering.
* Saveetha Institute of Medical and Technical Sciences.
* Saveetha University.

**REFERENCE:**

[Fang, Pingping, Jiang Chang, and Guanghua Lin. 2023. “Adaptation of Agriculture to Extreme Weather Events: Evidence from Apple Farmers’ Organic Fertilizer Use in China.” *Environmental Science and Pollution Research International*, August. https://doi.org/](http://paperpile.com/b/5fkdxM/P3Ae)[10.1007/s11356-023-29221-1](http://dx.doi.org/10.1007/s11356-023-29221-1)[.](http://paperpile.com/b/5fkdxM/P3Ae)

[Jin, Yi, Zhongyi Cheng, Yan He, Jianming Xu, and Jiachun Shi. 2024. “Dynamic Response of Cadmium Immobilization to a Ca-Mg-Si Soil Conditioner in the Contaminated Paddy Soil.” *The Science of the Total Environment* 908 (January): 168394.](http://paperpile.com/b/5fkdxM/9t7a)

[Kim, Han-Yong, Jonghan Ko, Suchel Kang, and John Tenhunen. 2013. “Impacts of Climate Change on Paddy Rice Yield in a Temperate Climate.” *Global Change Biology* 19 (2): 548–62.](http://paperpile.com/b/5fkdxM/CxcY)

[Meena, Ram Swaroop. 2019. *Nutrient Dynamics for Sustainable Crop Production*. Springer Nature.](http://paperpile.com/b/5fkdxM/TrS9)

[Mousavi, Seyed Rohollah, Vahid Alah Jahandideh Mahjenabadi, Bahman Khoshru, and Meisam Rezaei. 2023. “Spatial Prediction of Winter Wheat Yield Gap: Agro-Climatic Model and Machine Learning Approaches.” *Frontiers in Plant Science* 14: 1309171.](http://paperpile.com/b/5fkdxM/Szis)

[Pradeep, N., Sandeep Kautish, C. R. Nirmala, Vishal Goyal, and Sonia Abdellatif. 2019. *Modern Techniques for Agricultural Disease Management and Crop Yield Prediction*. IGI Global.](http://paperpile.com/b/5fkdxM/DknS)

[Sawatraksa, Nateetip, Poramate Banterng, Sanun Jogloy, Nimitr Vorasoot, and Gerrit Hoogenboom. 2023. “Crop Model Determined Mega-Environments for Cassava Yield Trials on Paddy Fields Following Rice.” *Heliyon* 9 (3): e14201.](http://paperpile.com/b/5fkdxM/FCsz)

[Shu, Xiaowei, Xiaoxiang Zhang, Shushen Wang, Tong Fu, Zhouyu Ding, Ying Yang, Zihan Wang, et al. 2024. “Simplified Panicle Fertilization Is Applicable to Cultivars, but Splits Are Preferred in Rice for a Higher Paddy Yield under Wheat Straw Return.” *Frontiers in Plant Science* 15 (January): 1273774.](http://paperpile.com/b/5fkdxM/KtXY)

[Sun, Boteng, Wei Zhou, Shilin Zhu, Song Huang, Xun Yu, Zhenyuan Wu, Xiaolong Lei, et al. 2022. “Universal Detection of Curved Rice Panicles in Complex Environments Using Aerial Images and Improved YOLOv4 Model.” *Frontiers in Plant Science* 13 (November): 1021398.](http://paperpile.com/b/5fkdxM/JsZN)

[Su, Yanqiu, Qiuyun Shi, Ziyuan Li, Hongmei Deng, Qian Zhou, Lihuan Li, Lanyin Zhao, Shu Yuan, Qi Liu, and Yanger Chen. 2024. “Rhodopseudomonas Palustris Shapes Bacterial Community, Reduces Cd Bioavailability in Cd Contaminated Flooding Paddy Soil, and Improves Rice Performance.” *The Science of the Total Environment*, March, 171824.](http://paperpile.com/b/5fkdxM/O1In)

[Yang, Xuziqi, Zekai Hua, Liang Li, Xingheng Huo, and Ziqiang Zhao. 2024. “Multi-Source Information Fusion-Driven Corn Yield Prediction Using the Random Forest from the Perspective of Agricultural and Forestry Economic Management.” *Scientific Reports* 14 (1): 4052.](http://paperpile.com/b/5fkdxM/bQv3)

[Yoshida, Shouichi. 1981. *Fundamentals of Rice Crop Science*. Int. Rice Res. Inst.](http://paperpile.com/b/5fkdxM/3Ojn)

[Zhan, Cun, Chuan Liang, Lu Zhao, Shouzheng Jiang, and Yaling Zhang. 2024. “Differential Responses of Crop Yields to Multi-Timescale Drought in Mainland China: Spatiotemporal Patterns and Climate Drivers.” *The Science of the Total Environment* 906 (January): 167559.](http://paperpile.com/b/5fkdxM/69S4)

[Zhao, Yijia, Chunming Yu, Ping Chen, Pan Mou, Jikang Chen, Gang Gao, Xiaofei Wang, Aiguo Zhu, and Kunmei Chen. 2024. “Study on Remediation of Cadmium Contaminated Paddy Field by Ramie (Boehmeria Nivea L.) Floating Island and Its Supporting Technology.” *Environmental Research* 242 (February): 117798.](http://paperpile.com/b/5fkdxM/5DtV)

**TABLES AND FIGURES**

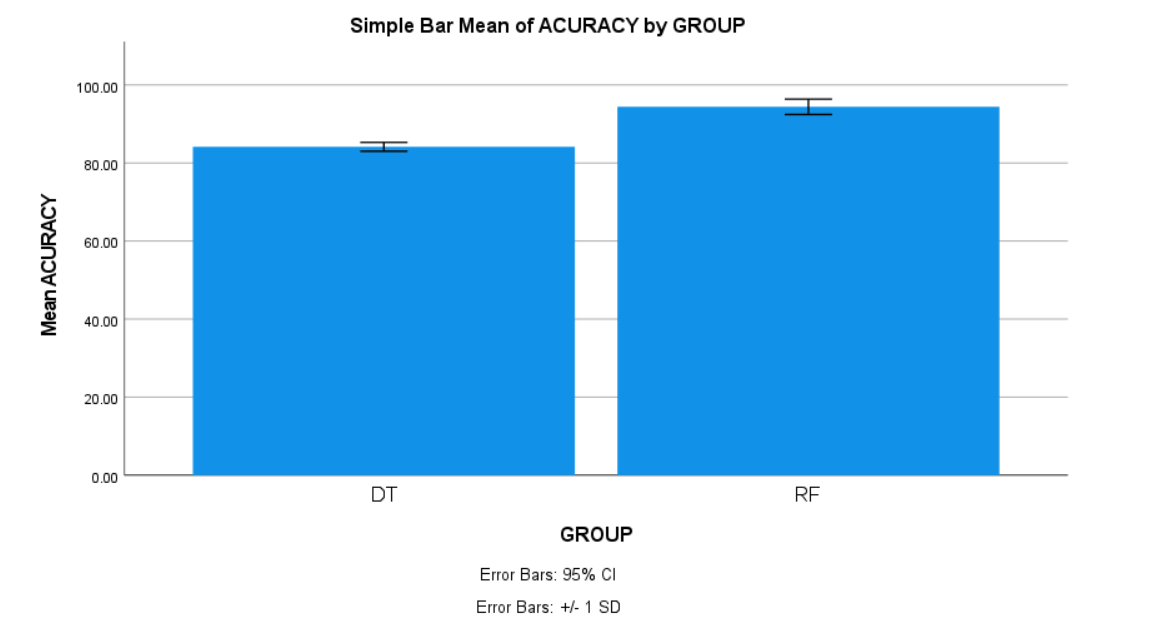
**Table 1**:The number of epochs taken for the RF and Decision tree Algorithms. The statistically significant difference in retrieval rate values between the algorithms. For the proposed algorithm RF, the mean value is 94.4090 and the standard deviation is 1.97338, with a standard error mean of 0.62404. Whereas for the comparison algorithm Decision, the mean value is 84.1610 with a standard deviation of 1.12583 and a standard error mean of 0.35602

|  | **Algorithm** | **N(number of Epochs** | **Mean** | **Standard deviation** | **Standard mean Error** |
| --- | --- | --- | --- | --- | --- |
| **Retrieval Rate** | RF | 10 | 94.4090 | 1.97338 | 0.62404 |
| **Retrieval Rate** | DT | 10 | 84.1610 | 1.12583 | 0.35602 |

**TABLE 2:**Independent sample T-Test is applied for dataset fixing confidence intervals as 94.4090% (Random Forest Algorithm appears to perform better than Decision Algorithm). And the significant difference between these two algorithms is 0.001 (p<0.05).

|  |  | **Leven’s test for equality of variables** | |  |  |  | **Test for equality of means** | | **95% Confidence Interval of the Difference** | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  | F | Sig | t | df | Sig  2-tailed | Mean  Difference | Std.error Difference | Lower | Upper |
| **Retrieval Rate** | **Equal variances Assumed** | 2.073 | .167 | 14.264 | 18 | <.001 | 10.24800 | .71845 | 8.73859 | 11.75741 |
| **Retrieval Rate** | **Equal Variances Not Assumed** |  |  | 14.264 | 14.297 | <.001 | 10.24800 | .71845 | 8.73859 | 11.78592 |

**Simple Bar mean of Retrieval Time By Group**

****

**FIG.1.**Bar chart showing the comparison of mean accuracy and standard errors for Random Forest and Decision tree Algorithms. The Random Forest algorithm demonstrates superior performance compared to the Decision tree algorithm in terms of mean accuracy, which is 94.4090, with a standard deviation of 1.97338. The significant difference between these two algorithms is 0.001 (p<0.05).

X-Axis: Random Forest Algorithm VS Decision tree Algorithm

Y-Axis: Mean retrieval rate of Accuracy detection.

Confidence interval: 94.4090%