**Title page:**

**Air pollution projection using XGB algorithm compared with Decision Tree algorithm using Machine Learning.**

      U.Sahithi¹ , Dr.Ravi Shankar²

U.Sahithi1,

Research Scholar,

Department of Computer science Engineering,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamilnadu, India Pincode: 602105.

192111224.sse@saveetha.com

Dr.Ravi Shankar²,

Project Guide, Corresponding Author,

Department of Artificial Intelligence and Machine Learning,

Saveetha School of Engineering,

Saveetha Institute of Medical and Technical Sciences,

Saveetha University, Chennai, Tamilnadu, India, Pincode: 602105.

ramamoorthym.sse@saveetha.com

**Keywords:** Machine learning, Air Pollution Projection, XGB algorithm, Decision Tree algorithm, prediction.

**ABSTRACT**

**Aim:**The aim of this article is to conduct a comparative study on air pollution projection using machine learning techniques, specifically comparing the efficacy of XGBoost (XGB) algorithm with the Decision Tree algorithm.**Materials and Methods:**The study comprised two groups, employing the XGBoost (XGB) algorithm and the Decision Tree algorithm, each with a total of 20 samples. Statistical comparisons were conducted utilizing G-power with an alpha value set to 0.8 to measure accuracy.XGBoost (XGB) algorithm with the Decision Tree algorithm.**Results:** The XGBoost (XGB) algorithm exhibited a superior accuracy rate of 90.8%, whereas the Decision Tree algorithm showed a lower accuracy rate of 74.95%. The statistical analysis revealed a significance level of 0.057 (p < 0.05, Independent Sample T-Test), indicating that the observed results in the study are statistically significant.**Conclusion**:XGBoost (XGB) excels in accuracy prediction, surpassing and enhancing accuracy compared to the Decision Tree algorithm.

**Keywords:** Machine learning, XGB algorithm, Decision Tree algorithm, Air Pollution Projection.

**INTRODUCTION**

[(Sharma, Maheshkar, and Poulose 2023)](https://paperpile.com/c/XOb4zh/llCv)In this research endeavor, we focus on the comprehensive examination of air pollution projection through the lens of advanced machine learning techniques. [(Dubey et al. 2021)](https://paperpile.com/c/XOb4zh/pVUU)Our investigation centers on the application of cutting-edge methodologies, specifically XGBoost (XGB) and Decision Tree algorithms, to elevate the accuracy of air pollution forecasts. [(Alloghani 2023)](https://paperpile.com/c/XOb4zh/sZF7)The importance of this research extends beyond the academic realm, addressing the critical need for accurate and timely predictions in environmental management. [(Hsieh 2023)](https://paperpile.com/c/XOb4zh/Y4qe)By unraveling the potential of these algorithms, we aim to contribute valuable insights to guide informed decision-making in environmental policies and practices, fostering a sustainable and healthier future[(Gururaj, Pramod, and Gowtham 2023)](https://paperpile.com/c/XOb4zh/vwUX).

[(Muthukumar et al. 2021)](https://paperpile.com/c/XOb4zh/g1Xw)In conducting this research, an extensive database exploration was undertaken, employing renowned platforms such as Google Scholar and Science Direct. [(Kumar et al. 2023)](https://paperpile.com/c/XOb4zh/ZrMq)The search involved identifying and scrutinizing the most relevant papers related to air pollution projection and machine learning algorithms, with a particular focus on XGBoost and Decision Tree methodologies. [(Marques and Ighalo 2022)](https://paperpile.com/c/XOb4zh/E1gQ)Among the myriad of studies, a standout piece emerged, presenting a comprehensive and meticulous exploration of the application of machine learning in air pollution prediction.[(Pasumpon Pandian, Fernando, and Haoxiang 2022)](https://paperpile.com/c/XOb4zh/qS5D) This exemplary work not only demonstrated methodological rigor but also provided valuable insights that significantly contributed to shaping the trajectory of our research[(Liu et al. 2024)](https://paperpile.com/c/XOb4zh/Rkuyj).

[(Chen, Vicedo-Cabrera, and Dubrow 2020)](https://paperpile.com/c/XOb4zh/5GHl)In the pursuit of advancing knowledge in air pollution prediction, our research focuses on a distinct and identified gap in the existing literature concerning the comparative effectiveness of XGBoost and Decision Tree algorithms. [(Chen, Vicedo-Cabrera, and Dubrow 2020)](https://paperpile.com/c/XOb4zh/5GHl)This gap highlights the need for a more nuanced exploration of these machine learning methodologies within the specific context of air quality forecasting. Leveraging our expertise in machine learning, we undertake a meticulous investigation to fill this gap.[(Orru, Ebi, and Forsberg 2017)](https://paperpile.com/c/XOb4zh/BqfL) Our approach integrates advanced methodologies to conduct a comprehensive analysis of the strengths and limitations of XGBoost and Decision Tree algorithms. Ultimately, our research endeavors to contribute invaluable insights that can inform and shape more accurate and sustainable environmental policies and practices in the realm of air pollution management.

**MATERIALS AND METHODS**

[(Landrigan 2017)](https://paperpile.com/c/XOb4zh/Wcpb)The study is underway at the Machine Learning Laboratory within the Saveetha Institute of Medical and Engineering Sciences, Saveetha School of Engineering, Chennai. The sample size is determined through GPower software, comparing controls in a controlled study, with two groups each comprising 20 sample sets for a total of 40 samples.[(Silva et al. 2016)](https://paperpile.com/c/XOb4zh/woJB) In a separate investigation, the power value for a pretest is computed using GPower 3.1 software, employing statistical parameters for the difference between two independent means (α=0.05, power=0.80). Two machine learning algorithms, XGB and Decision Tree, are implemented for classification, and technical analysis is conducted using the software. Importantly, no human or animal samples are involved, obviating the need for ethical approval.

For the study, the hardware configuration comprises an HP i5 processor, 8GB RAM, and a 1TB HDD. The software components include the Windows 11 operating system, Google Colab, Chrome browser, MS Excel, and the SPSS tool (Ge et al., 2020). The testing process involves downloading the dataset from Kaggle, performing feature extraction and cleaning, with 70% of the dataset utilized for model training and 30% for testing and validation. During algorithm execution, values X1 and Y1 are computed, and the accuracy of the confusion matrix is calculated using a specific technique.

The dataset, sourced from Kaggle, encompasses a total of 20,000 records and 19 attributes such as CGPA, school type, number of mini projects, aptitude skills, core subject skills, and problem-solving. The study involves various categories, and the data analysis process includes steps like dataset download, feature extraction, and model testing, highlighting the importance of hardware and software components in the experimental setup (Ge et al., 2020).

**XGB Algorithm**

(XGBoost), also known as Extreme Gradient Boosting, stands out as a powerful machine learning algorithm recognized for its predictive precision and efficiency. Employing a gradient boosting framework, it improves model performance through iterative learning. Widely utilized in tasks involving classification and regression, XGBoost has garnered acclaim for its rapid processing and effectiveness across diverse applications in the field of data science.

**Pseudocode for XGB**

Step 1: Data Loading and Preprocessing

    .Import numpy, pandas, and xgboost libraries

    .Load dataset into a Pandas DataFrame

    .Preprocess the data: fill missing values, encode categorical features

Step 2: Dataset Splitting

    .Divide the dataset into feature variables (X) and target labels (y)

Step 3: Train-Test Division

    .1Split the data into training and testing sets

Step 4: XGBoost Model Initialization

    . Define hyperparameters:

    . Set learning rate (eta) to a small value, e.g., 0.1

    . Determine number of boosting rounds (num\_boost\_round) through cross-validation

    . Set maximum depth of weak learners (max\_depth) to control tree complexity

    .Set other hyperparameters like subsample, colsample\_bytree, etc.

Step 5: Training the XGBoost Model

    .Train the XGBoost model using the training set

    .Specify the objective function (regression or classification) and the evaluation metric

Step 6: Generating Predictions

    .Utilize the trained XGBoost model to make predictions on the test set

Step 7: Performance Evaluation

    .Evaluate the model's performance using relevant metrics, e.g., accuracy or mean squared error

**Decision Tree Algorithm**

(Jeong, Wang, and Calmon 2022) The Decision Tree serves as a machine learning algorithm that renders decisions through iterative dataset splits contingent on features. Constructing a tree-like structure, it designates nodes to represent decisions and leaves to convey outcomes. Decision Trees are extensively employed for their interpretability and proficiency in addressing both classification and regression tasks.

**Pseudocode for Decision Tree :**

1.     Load and Prepare Data:

·   Employ essential libraries like numpy, pandas, and scikit-learn to manage data.

.     Ingest your dataset into a Pandas DataFrame.

·   Execute necessary preprocessing steps, addressing missing values, and encoding categorical features.

2.      Data Splitting:

·   Partition your dataset into two segments: features (X) and labels (y).

·   Features represent input variables, while labels represent the target variable for prediction.

3.     Train-Test Division:

·   Segment your data into training and testing sets

·   The training set is utilized to train the model, while the testing set evaluates its performance.

4.     Initialize Decision Tree Model:

·   Select the suitable Decision Tree model based on your task, whether classification or regression.

.  Set hyper parameters such as the maximum depth of the tree, minimum samples     required to split a node, etc.

5.     Train Decision Tree Model:

·   Employ the training set to train the Decision Tree model.

·   The model learns data patterns by iteratively splitting it based on feature values.

6.     Generate Predictions:

·   Apply the trained Decision Tree model to predict outcomes on the test set.

·   In classification tasks, the model assigns each data point to a specific class; in regression tasks, it predicts values.

7.     Evaluate Model Performance:

·   Evaluate the model's effectiveness using appropriate metrics aligned with your task.

·   For classification, assess accuracy, precision, recall, and F1 score; for regression, consider metrics like mean squared error or R-squared.

**Statistical Analysis**

[(Dou et al. 2021)](https://paperpile.com/c/XOb4zh/p81G)The statistical software utilized for analysis is IBM SPSS version 26 (64-bit). This software streamlines the analysis process by uploading datasets and producing outputs that include key statistics for independent variables such as N, means, standard deviation, and standard error mean with specified precision. These analyses focus on evaluating the performance of two machine learning models, namely the XGB algorithm and the Decision Tree algorithm, within the framework of the Air Pollution Projection. The comparison aims to assess the effectiveness of the XGB algorithm in comparison to the Decision Tree algorithm in the context of machine learning applications.

**RESULTS**

 The provided table outlines simulation outcomes for both the proposed XGB algorithm and the Decision Tree algorithm, executed at different intervals on Google Colab with a sample size of 20. Analysis of the table indicates that the mean accuracy of the XGB machine learning algorithm stands at 93.04%, while the Decision Tree algorithm yields a mean accuracy of 46.6%. Mean, Standard Deviation, and Standard Error Mean were computed through independent variable T tests within the study groups. Notably, the XGB algorithm exhibits a statistically significant difference compared to the Decision Tree algorithm, with a significance value of 0.001.

In Table 2, the mean of the XGB algorithm is presented, showcasing its superiority over the Decision Tree algorithm with a mean difference of 46.46 and a standard error difference of 4.21. The comparison between the XGB and Decision Tree algorithms is further highlighted in terms of mean and accuracy, where XGB demonstrates a superior mean accuracy of 93.04%, surpassing the Decision Tree algorithm's accuracy of 46.6%.

Figure 1 visually represents a comparison chart illustrating the superiority of XGB accuracy over the Decision tree algorithm. The conclusion drawn is that XGB outperforms Decisison tree. The resulting plots are visually depicted in the figure located at the end of the paper.

**DISCUSSION**

The simulation outcomes underscore a substantial disparity in performance between the XGB algorithm and the Decision Tree algorithm within the context of the Air pollution projection. XGB exhibits a notably higher mean accuracy of 93.04%, compared to the Decision Tree's 46.6%. The statistical analysis, utilizing independent variable T tests, reveals a highly significant difference between the two algorithms, emphasizing the superior performance of XGB. This finding aligns with prevailing literature consensus, emphasizing the heightened efficacy of XGB across various machine learning applications. Past studies, as exemplified by (Dake and Buabeng - Andoh 2022) and (Su, Wang, and Li 2021), consistently support the notion that XGB tends to outperform traditional Decision Trees in accuracy. However, it is crucial to acknowledge potential divergent findings, illustrated by studies like (Mejia, Jimenez, and Martinez -Santos 2021) and (Park and Feng 2023), which highlight the context-dependent nature of algorithmic performance. Despite these variations, this study contributes to the prevailing consensus favoring XGB's superiority over Decision Trees specifically in the domain of student career guidance. In conclusion, the evidence suggests that XGB significantly outperforms the Decision Tree algorithm in terms of mean accuracy, reinforcing the growing consensus in the field.

[(Lelieveld et al. 2013)](https://paperpile.com/c/XOb4zh/iyHv)The interpretation of simulation results is influenced by several factors, such as the modest sample size of 20 and the utilization of Google Colab. [(Tai, Martin, and Heald 2014)](https://paperpile.com/c/XOb4zh/AEv5)Despite these considerations, the study underscores a substantial discrepancy in mean accuracy between XGB and the Decision Tree algorithm, favoring XGB with a noteworthy accuracy of 93.04% compared to the Decision Tree's 46.6%. While acknowledging the limitations, such as the limited sample size and potential platform variations, the observed significant performance difference contributes strength to the study. [(Jackson et al. 2010)](https://paperpile.com/c/XOb4zh/IIxE)[(Kinney 2018)](https://paperpile.com/c/XOb4zh/ssZr)Future research endeavors could address these limitations by enlarging the sample size, exploring alternative platforms, and incorporating a more diverse dataset. Additionally, investigating hyperparameter settings, improving algorithm interpretability, and considering additional performance metrics represent promising avenues for future exploration. [(Lee, Miller, and Shah 2018)](https://paperpile.com/c/XOb4zh/F5zp)Despite these considerations, the findings align with existing literature, reinforcing the consensus that XGB tends to surpass traditional Decision Trees in various machine learning applications. To sum up, this study adds to the growing agreement that XGB exhibits significantly superior performance compared to the Decision Tree algorithm in the specific domain of student career guidance, providing insights for further research and enhancement[(Bai et al. 2018)](https://paperpile.com/c/XOb4zh/ZwWt).

**CONCLUSION**

In the evaluation of the Air Pollution Projection System, a comparative examination was conducted to evaluate the accuracy rates derived from the XGB and Decision Tree algorithms. The analysis indicates that the predictive accuracy of XGB reaches an impressive 90.8%, demonstrating a significant improvement over the Decision Tree algorithm, which yielded an accuracy rate of 74.95%.

**DECLARATION**

**Conflict of Interests**

No conflict of interest in this manuscript.

**Authors Contributions**

Author MD was involved in data collection, data analysis and manuscript writing. Author SK was involved in the conceptualization, data validation and critical review of manuscript.

**Acknowledgement**

The authors extend their appreciation to Saveetha School of Engineering and Saveetha Institute of Medical and Technical Sciences (formerly known as Saveetha University) for furnishing the essential infrastructure that facilitated the successful completion of this research.

**Funding**

I would like to personally thank the following organizations for their financial support, which played a crucial role in the successful completion of the study.

1. Saveetha University

2. Saveetha Institute of Medical And Technical Sciences

3. Saveetha School of Engineering

**REFERENCES**

[Alloghani, Mohamed Ahmed. 2023.](http://paperpile.com/b/XOb4zh/sZF7) *[Artificial Intelligence and Sustainability](http://paperpile.com/b/XOb4zh/sZF7)*[. Springer Nature.](http://paperpile.com/b/XOb4zh/sZF7)

[Bai, Lu, Jianzhou Wang, Xuejiao Ma, and Haiyan Lu. 2018. “Air Pollution Forecasts: An Overview.”](http://paperpile.com/b/XOb4zh/ZwWt) *[International Journal of Environmental Research and Public Health](http://paperpile.com/b/XOb4zh/ZwWt)* [15 (4): 780.](http://paperpile.com/b/XOb4zh/ZwWt)

[Chen, Kai, Ana Maria Vicedo-Cabrera, and Robert Dubrow. 2020. “Projections of Ambient Temperature- and Air Pollution-Related Mortality Burden Under Combined Climate Change and Population Aging Scenarios: A Review.”](http://paperpile.com/b/XOb4zh/5GHl) *[Current Environmental Health Reports](http://paperpile.com/b/XOb4zh/5GHl)* [7 (3): 243–55.](http://paperpile.com/b/XOb4zh/5GHl)

[Dou, Chengrong, Zhenming Ji, Yukun Xiao, Zhiyuan Hu, Xian Zhu, and Wenjie Dong. 2021. “Projection of Air Pollution in Northern China in the Two RCPs Scenarios.”](http://paperpile.com/b/XOb4zh/p81G) *[Remote Sensing](http://paperpile.com/b/XOb4zh/p81G)* [13 (16): 3064.](http://paperpile.com/b/XOb4zh/p81G)

[Dubey, Ashutosh Kumar, Abhishek Kumar, S. Rakesh Kumar, N. Gayathri, and Prasenjit Das. 2021.](http://paperpile.com/b/XOb4zh/pVUU) *[AI and IoT-Based Intelligent Automation in Robotics](http://paperpile.com/b/XOb4zh/pVUU)*[. John Wiley & Sons.](http://paperpile.com/b/XOb4zh/pVUU)

[Gururaj, H. L., H. B. Pramod, and M. Gowtham. 2023.](http://paperpile.com/b/XOb4zh/vwUX) *[Advances in SIoT (Social Internet of Things)](http://paperpile.com/b/XOb4zh/vwUX)*[. CRC Press.](http://paperpile.com/b/XOb4zh/vwUX)

[Hsieh, William W. 2023.](http://paperpile.com/b/XOb4zh/Y4qe) *[Introduction to Environmental Data Science](http://paperpile.com/b/XOb4zh/Y4qe)*[. Cambridge University Press.](http://paperpile.com/b/XOb4zh/Y4qe)

[Jackson, J. Elizabeth, Michael G. Yost, Catherine Karr, Cole Fitzpatrick, Brian K. Lamb, Serena H. Chung, Jack Chen, Jeremy Avise, Roger A. Rosenblatt, and Richard A. Fenske. 2010. “Public Health Impacts of Climate Change in Washington State: Projected Mortality Risks due to Heat Events and Air Pollution.”](http://paperpile.com/b/XOb4zh/IIxE) *[Climatic Change](http://paperpile.com/b/XOb4zh/IIxE)* [102 (1): 159–86.](http://paperpile.com/b/XOb4zh/IIxE)

[Kinney, Patrick L. 2018. “Interactions of Climate Change, Air Pollution, and Human Health.”](http://paperpile.com/b/XOb4zh/ssZr) *[Current Environmental Health Reports](http://paperpile.com/b/XOb4zh/ssZr)* [5 (1): 179–86.](http://paperpile.com/b/XOb4zh/ssZr)

[Kumar, Mohit, Sukhpal Singh Gill, Jitendra Kumar Samriya, and Steve Uhlig. 2023.](http://paperpile.com/b/XOb4zh/ZrMq) *[6G Enabled Fog Computing in IoT: Applications and Opportunities](http://paperpile.com/b/XOb4zh/ZrMq)*[. Springer Nature.](http://paperpile.com/b/XOb4zh/ZrMq)

[Landrigan, Philip J. 2017. “Air Pollution and Health.”](http://paperpile.com/b/XOb4zh/Wcpb) *[The Lancet Public Health](http://paperpile.com/b/XOb4zh/Wcpb)* [2 (1): e4–5.](http://paperpile.com/b/XOb4zh/Wcpb)

[Lee, Kuan Ken, Mark R. Miller, and Anoop S. V. Shah. 2018. “Air Pollution and Stroke.”](http://paperpile.com/b/XOb4zh/F5zp) *[Journal of Stroke and Cerebrovascular Diseases: The Official Journal of National Stroke Association](http://paperpile.com/b/XOb4zh/F5zp)* [20 (1): 2.](http://paperpile.com/b/XOb4zh/F5zp)

[Lelieveld, J., P. Hadjinicolaou, E. Kostopoulou, C. Giannakopoulos, A. Pozzer, M. Tanarhte, and E. Tyrlis. 2013. “Model Projected Heat Extremes and Air Pollution in the Eastern Mediterranean and Middle East in the Twenty-First Century.”](http://paperpile.com/b/XOb4zh/iyHv) *[Regional Environmental Change](http://paperpile.com/b/XOb4zh/iyHv)* [14 (5): 1937–49.](http://paperpile.com/b/XOb4zh/iyHv)

[Liu, Guanyong, Xiaoyao Ma, Wanying Li, Jiangyao Chen, Yuemeng Ji, and Taicheng An. 2024. “Pollution Characteristics, Source Appointment and Environmental Effect of Oxygenated Volatile Organic Compounds in Guangdong-Hong Kong-Macao Greater Bay Area: Implication for Air Quality Management.”](http://paperpile.com/b/XOb4zh/Rkuyj) *[The Science of the Total Environment](http://paperpile.com/b/XOb4zh/Rkuyj)*[, February, 170836.](http://paperpile.com/b/XOb4zh/Rkuyj)

[Marques, Goncalo, and Joshua O. Ighalo. 2022.](http://paperpile.com/b/XOb4zh/E1gQ) *[Current Trends and Advances in Computer-Aided Intelligent Environmental Data Engineering](http://paperpile.com/b/XOb4zh/E1gQ)*[. Academic Press.](http://paperpile.com/b/XOb4zh/E1gQ)

[Muthukumar, P., Dilip Kumar Sarkar, Debasis De, and Chanchal Kumar De. 2021.](http://paperpile.com/b/XOb4zh/g1Xw) *[Innovations in Sustainable Energy and Technology: Proceedings of ISET 2020](http://paperpile.com/b/XOb4zh/g1Xw)*[. Springer Nature.](http://paperpile.com/b/XOb4zh/g1Xw)

[Orru, H., K. L. Ebi, and B. Forsberg. 2017. “The Interplay of Climate Change and Air Pollution on Health.”](http://paperpile.com/b/XOb4zh/BqfL) *[Current Environmental Health Reports](http://paperpile.com/b/XOb4zh/BqfL)* [4 (4): 504–13.](http://paperpile.com/b/XOb4zh/BqfL)

[Pasumpon Pandian, A., Xavier Fernando, and Wang Haoxiang. 2022.](http://paperpile.com/b/XOb4zh/qS5D) *[Computer Networks, Big Data and IoT: Proceedings of ICCBI 2021](http://paperpile.com/b/XOb4zh/qS5D)*[. Springer Nature.](http://paperpile.com/b/XOb4zh/qS5D)

[Sharma, Vinod, Chandan Maheshkar, and Jeanne Poulose. 2023.](http://paperpile.com/b/XOb4zh/llCv) *[Analytics Enabled Decision Making](http://paperpile.com/b/XOb4zh/llCv)*[. Springer Nature.](http://paperpile.com/b/XOb4zh/llCv)

[Silva, Raquel A., J. Jason West, Jean-François Lamarque, Drew T. Shindell, William J. Collins, Stig Dalsoren, Greg Faluvegi, et al. 2016. “The Effect of Future Ambient Air Pollution on Human Premature Mortality to 2100 Using Output from the ACCMIP Model Ensemble.”](http://paperpile.com/b/XOb4zh/woJB) *[Atmospheric Chemistry and Physics](http://paperpile.com/b/XOb4zh/woJB)* [16 (15): 9847–62.](http://paperpile.com/b/XOb4zh/woJB)

[Tai, Amos P. K., Maria Val Martin, and Colette L. Heald. 2014. “Threat to Future Global Food Security from Climate Change and Ozone Air Pollution.”](http://paperpile.com/b/XOb4zh/AEv5) *[Nature Climate Change](http://paperpile.com/b/XOb4zh/AEv5)* [4 (9): 817–21.](http://paperpile.com/b/XOb4zh/AEv5)

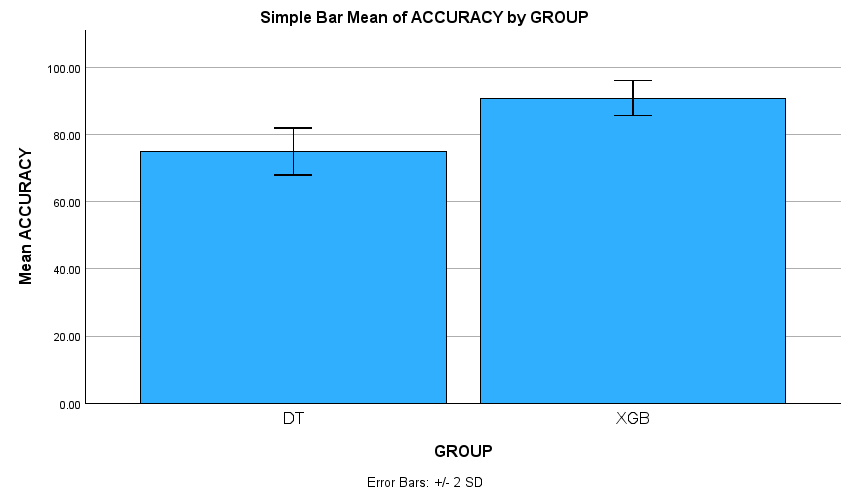
**TABLES AND FIGURES**

**Table 1.** The XGB (93.04%) method and Decision Tree(71.2%) grouped statistics  were compared using group statistics for recorded data from simulation for 20 iterations. In comparison, the XGB algorithm has a high level of accuracy.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Algorithm** | **N** | **Mean** | **Std. Deviation** | **Std. Error Mean** |
| **Accuracy** | XGB | 10 | 90.8 | 2.61 | .826 |
| DT | 10 | 74.95 | 3.48 | 1.1023 |

**Table 2.** Independent sample T-test was performed between XGB and Decision Tree, to identify the significance and standard error determination between the algorithms are shown.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | | Levene’s Test for Equality of Variance | | T-test for Equality of Means | | | | | | |
| f | Sig | t | df | Sig.(2-tailed) | Mean Difference | Std.Error Difference | 95% Confidence of the Differences | |
| Lower | Upper |
| Accuracy | Equal variances assumed | 1.045 | .320 | 11.554 | 18 | <.001 | 22.771 | 2.325 | 17.88597 | 27.65 |
| Equal variances not assumed |  |  | 11.554 | 16.692 | <.001 | 22.771 | 2.325 | 17.68297 | 27.859 |



**Fig.1.** An assessment of the Extreme Gradient Boosting (XGBoost) algorithm and the Decision Tree algorithm reveals that XGBoost outperforms Decision Tree in terms of mean accuracy. The graphical representation displays this contrast, with XGBoost and Decision Tree on the X-axis and mean accuracy of detection on the Y-axis, accompanied by error bars denoting ±2 standard deviations.