** A logo for a college

Description automatically generated**

**Kingdom of Saudi Arabia**

**Ministry of Education**

**King Faisal University**

**College of Computer Sciences & Information Technology**

**(This report is based on the research paper ' Energy time series forecasting-Analytical and empirical assessment of conventional and machine learning models by Hala Hamdoun, Alaa Sagheer, and Hassan Youness Journal of Intelligent & Fuzzy Systems, 40(6), 12477- 12502, 2021)**

**By**

**Saja Adel Alabdulmuhsen (220029219)**

**Fatimah Albaqshi (221419327)**

**Asaeil AlAskar (21930652)**

**Supervised by**

**Dr. Hala Hamdoun**

**December ,2024**

Contents

[**1.** **Introduction** 3](#_Toc185334957)

[**1.1.** **Performance Table** 3](#_Toc185334958)

[**1.2.** **ROC Curve Analysis** 4](#_Toc185334959)

[**1.3.** **Conclusion** 4](#_Toc185334960)

**List of figures:**

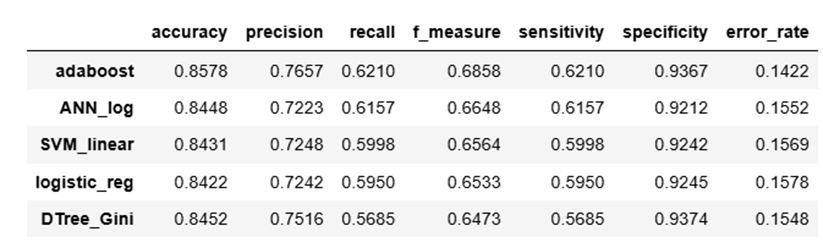
|  |  |  |
| --- | --- | --- |
| *Figure 1* | *Models results* | *Page 4* |
| *Figure 2* | *ROC Curve Analysis* | *Page 5* |

## **Introduction**

In thie project, we explore salary dynamics and their multifaceted influences. Personal attributes, such as age, workclass, education,and race significantly impact salaries. We investigate the interplay between personality traits and income, utilizing a diverse dataset from "Census Income" dataset with 14 distinct attributes. Our research aims to unveil industry-specific disparities, analyze the relationship between experience and income growth, and explore regional variations in compensation. We also address the impact of gender on pay differentials, as well as correlations with factors like race and education levels. Therefore, we aim to build a model that predicts future salaries based on personal attributes.

## **Performance Table**

The following table summarizes the performance metrics of various models. The Adaboost model achieved the best overall results, with the highest accuracy and lowest error rate.



*Figure 1: Models results*

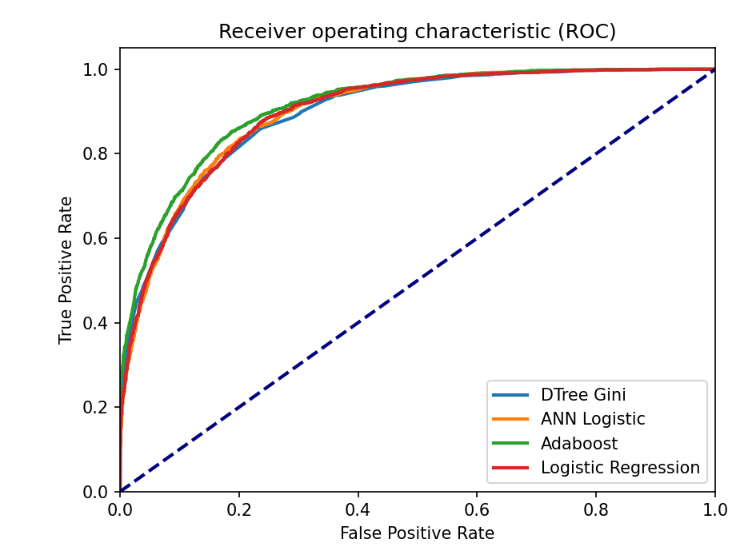
Table Explanation:

* Accuracy: Represents the proportion of correct predictions made by the model.
* Precision: Indicates the accuracy of positive predictions.
* Recall: Reflects the model’s ability to identify positive samples.
* F-measure: A weighted average of Precision and Recall, used to measure balanced performance.
* Sensitivity: The proportion of correctly identified positive samples.
* Specificity: The proportion of correctly classified negative samples.
* Error Rate: Represents the proportion of incorrect predictions.

From the table, it is evident that the Adaboost model achieved the best results, with the highest accuracy (0.8578) and the lowest error rate (0.1422).

## **ROC Curve Analysis**

The results are visualized through the ROC Curve in the figure below:



*Figure 2: ROC Curve Analysis*

Curve Analysis:

* The plot above of the receiver operating characteristic curve for the 4 models; Adaboost, ANN with logistics activation function, DTree Gini and logistic regression model.
* From figure, we can see that the ROC curve of the Adaboost model has the highest lift and is closest to the top left corner (TPR of 1 and FPR of 0) of the plot. The Adaboost model's curve clearly separates itself from the ROC curves of the other 3 models, which overlap with each other.

## **Conclusion**

The Adaboost model not only has the **highest accuracy**, but also has the **highest precision and F-measure** of all the models developed as a part of this analysis. The advantages of using Adaboost over other models is that they are very simple to implement. Since they are made up of weak individual learners, they are less susceptible to overfitting. However, Adaboost is sensitive to noisy data and outliers.