Adult Census Income Prediction

Objective

Developing a machine learning model that can accurately predict whether a person has an annual income of more than $50000USD, based on set of demographic and socio-economic features

Data Transformation

Handling Missing Values: We have rows consisting of missing values which should be removed from our dataset

Encoding : We have used Labelencoder to convert categorical values into numerical values

Scaling : We have used Standard Scalar to bring all columns under same scale and improve the model efficiency

Model Selection

We have used various classification models like Logistic Regression, Random Forests, Decision Tree Classifier and selects the model which gave best accuracy-score, f1-score without overfitting the model

Metrics

Accuracy=(Number of correctly classified data points)/(Total number of data points)

Precision=(True Positives)/(True Positives + False Positives)

Recall=(True Positives)/(True Positives+False Negatives)

F1-Score=(2*Precision*Recall)/(Precision+Recall)

Q & A

Q1) What is the Source of the data?

The source of the Adult Census dataset is United State Census Bureau

Q2)What was the type of the data?

The data was the combination of numerical and categorical values

Q3)How you treat Null values?

We have very small number of values, so we removed those rows from the dataset

Q4)How you convert categorical values into Numerical values?

We have used label encoding to convert categorical values to numerical values.

Q5)How you selected the best model?

We have trained our data using various classification models and we select the model which gave best f1-score.

Q6)In which platform you deployed your model?

We deployed our model in AWS Platform.