**Proposed code Result of** NetworkTrafficClassification

**Libraries Used:**

Numpy, pandas, matplotlib, seaborn, sklearn, imblearn etc.

**Dataset:**

* Dataset Name: **TotalFeatures-ISCXFlowMeter**
* Dataset link**:** <https://www.kaggle.com/datasets/antoniogarcanoguez/totalfeatures-iscxflowmeter>

**Loading the dataset:**

 Mounted Google Drive to access files.

 Loaded the dataset from **/content/drive/MyDrive/client /TotalFeatures- ISCXFlowMeter.csv** into a Data Frame

**Data preprocessing & EDA:**

 Displayed the first five rows of the dataset using df.head().

 Retrieved dataset dimensions (rows, columns) with df.shape.

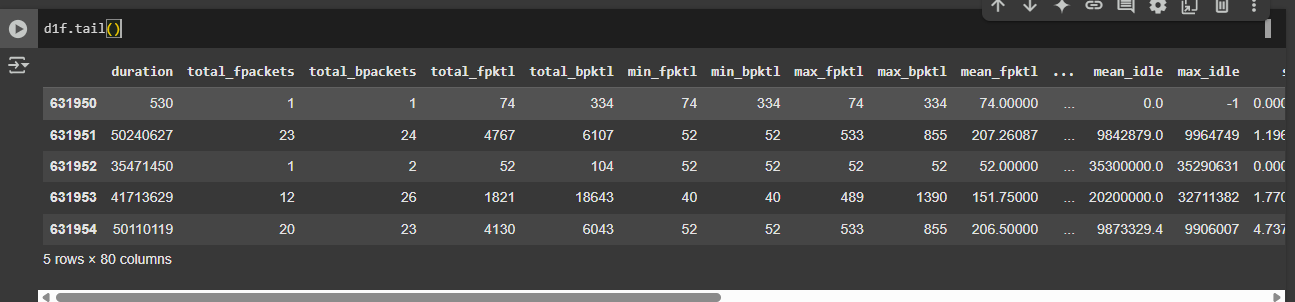
 Check unique Value in Columns. d1f['calss'].nunique()

 Checking the Columns name. d1f.columns.

**Label Encoding:**

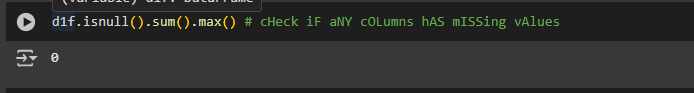
We have applied label Encoding to convert categorical values to numerical values (for our target feature -“calss”)

 Retrieves and displays the last five rows of the DataFrame **d1f**, helping to inspect recent entries or verify data integrity.

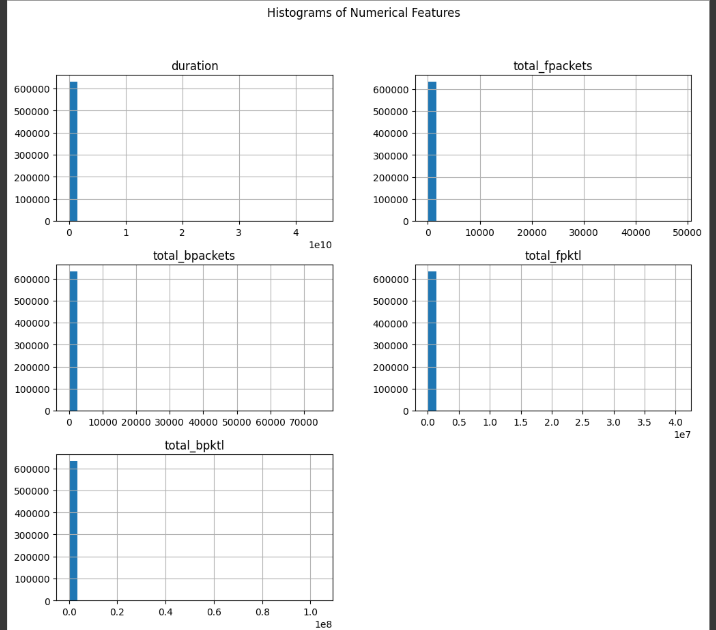


**Null Values:**

We have checked a null values in our dataset does not contain null values.

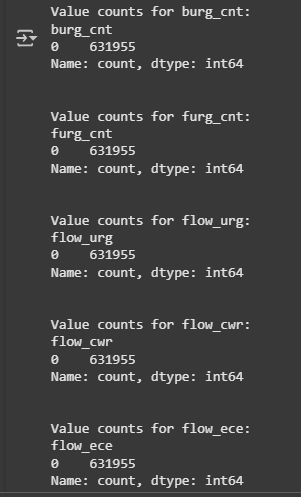


* The first five numerical columns from d1f, plots their histograms with 30 bins using Matplotlib, sets a title, and displays the plot.



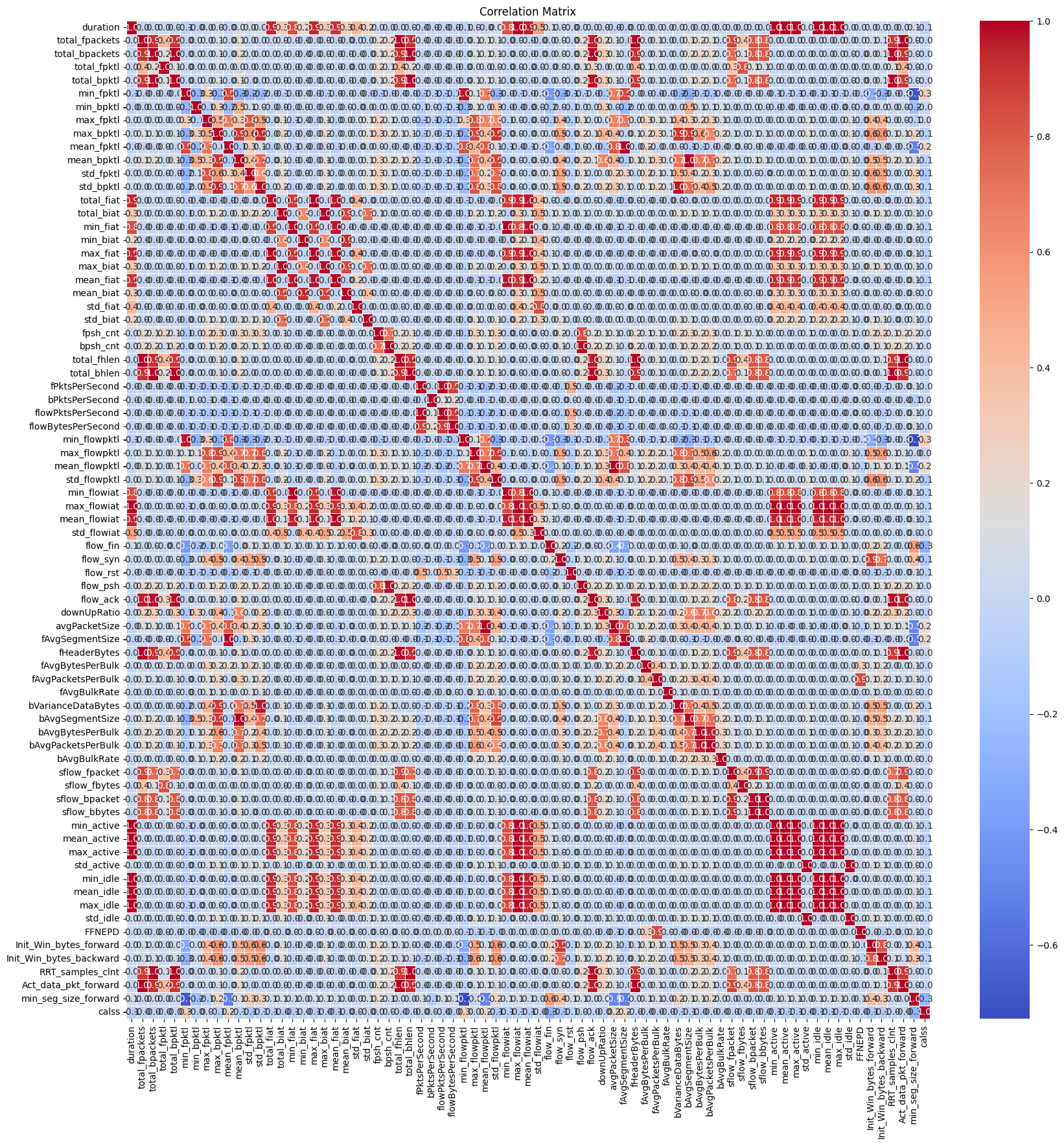
* We have counted the number of class in this four features-

burg\_cnt', 'furg\_cnt', 'flow\_urg', 'flow\_cwr', 'flow\_ece

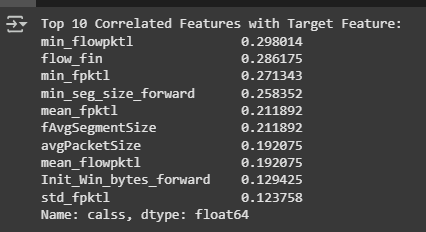


* We drop 4 features because they only contains 1 class Value.

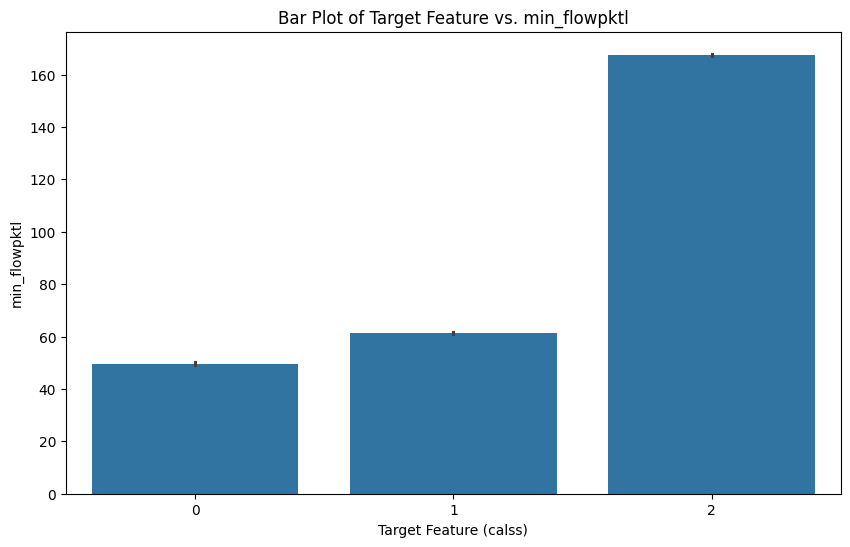
**Correlation matrix:**



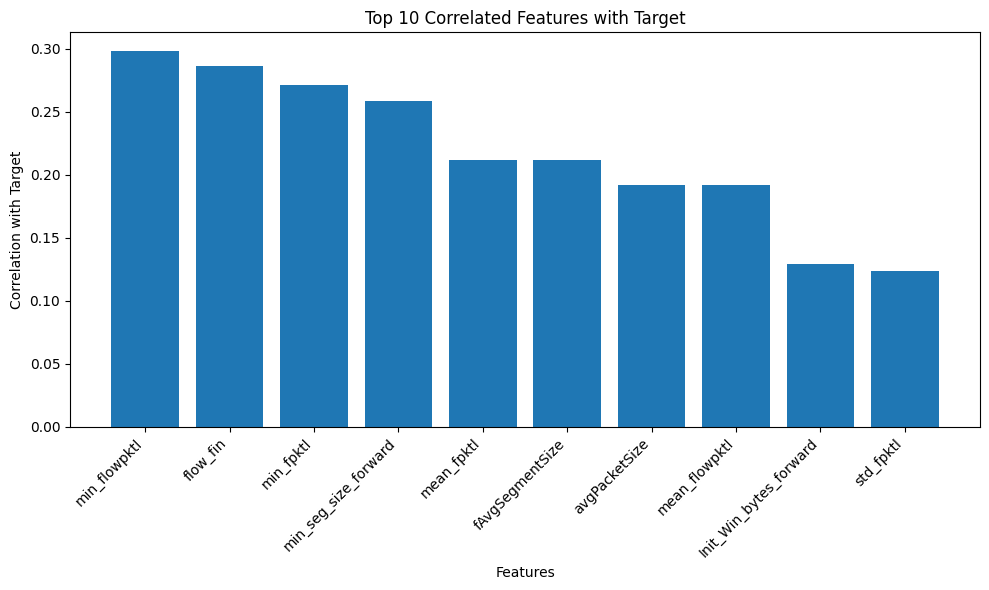
* calculates the absolute correlation of all features in d1f with the target feature 'calss', sorts them in descending order, selects the top 10



* We creates a bar plot showing the relationship between the target feature 'calss' and 'min\_flowpktl' using seaborn, labeling axes and adding a title for better visualization



* We creates a bar plot displaying the top 10 features most correlated with the target, labeling axes, rotating x-axis labels for readability, adjusting layout, and showing the plot.

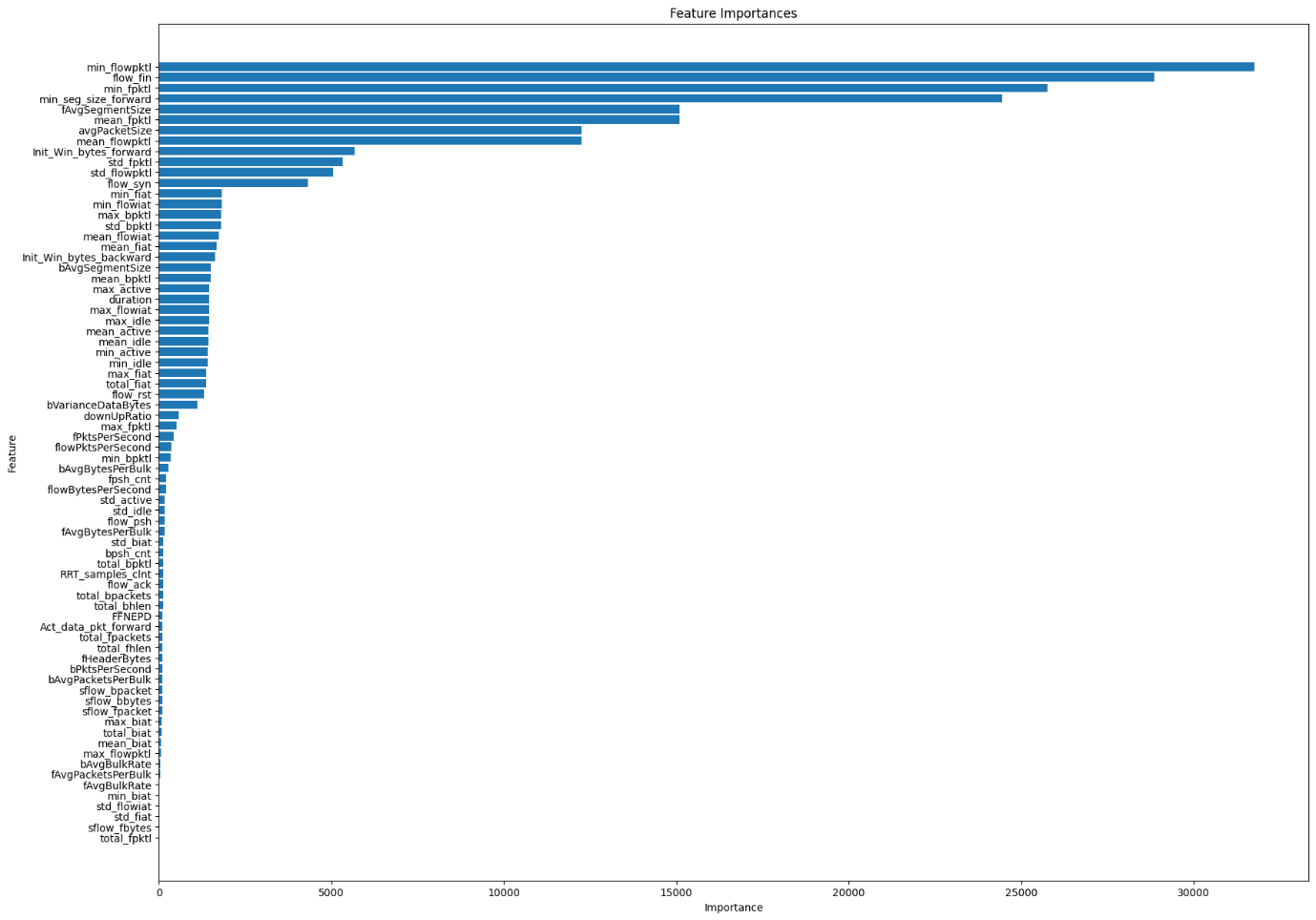


**Feature Selecting:**

It selects the top 50 most statistically important features from X based on the target y using SelectKBest with f\_classif,

**Features importance:**

It calculates feature importance scores (ANOVA F-scores) using SelectKBest, sorts them, and plots a horizontal bar chart showing the most important features at the top



**Checking data imbalance:**

calss

2 471597

1 155613

0 4745

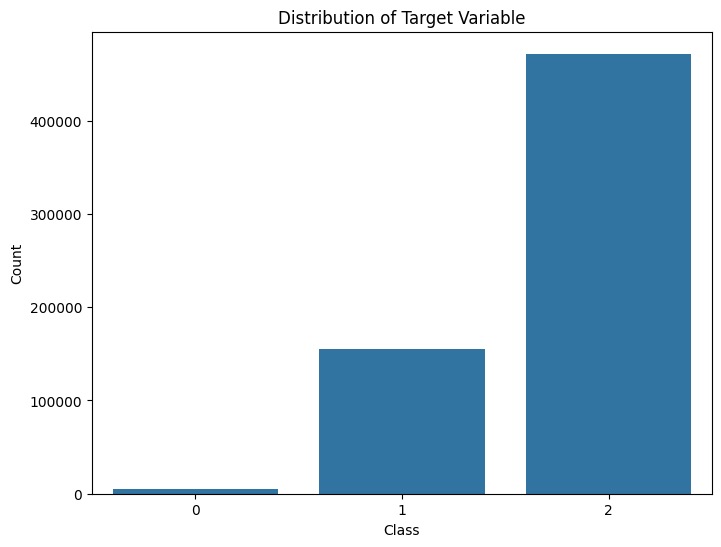
Name: count, dtype: int64

calss

2 74.625092

1 24.624063

0 0.750845



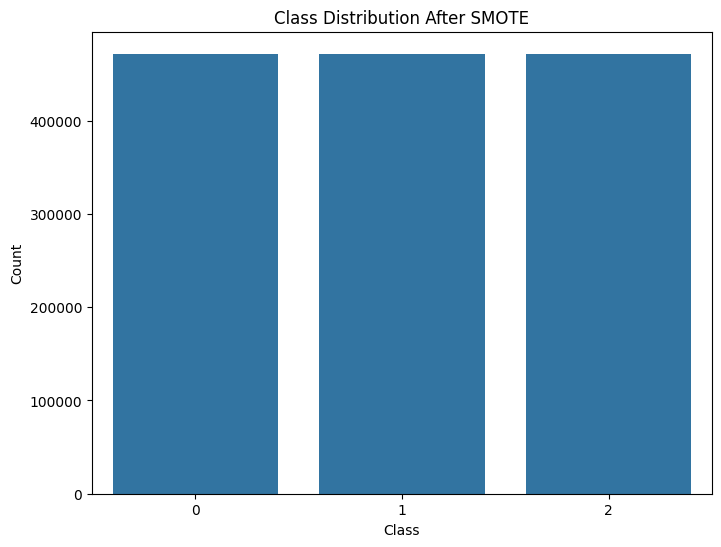
**Balance the data by using Smote**

calss

2 471597

1 471597

0 471597



**Data Splitting:**

We split data into training and testing sets in the ratio of 80:20 .

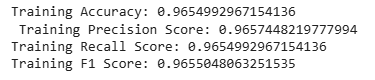
80% for training and 20% for testing.

**Model implementation:**

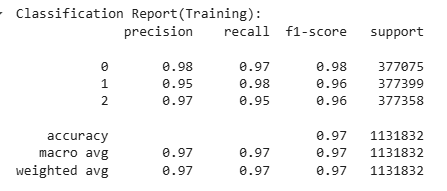
**Random Forest**

Parameters=(n\_estimators=500,max\_depth=25,n\_jobs=-1, random\_state=42)

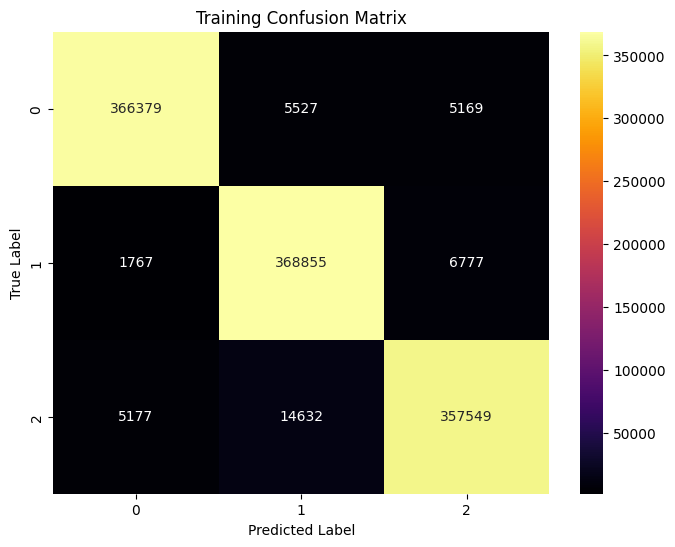
**Random forest Training Result**

****

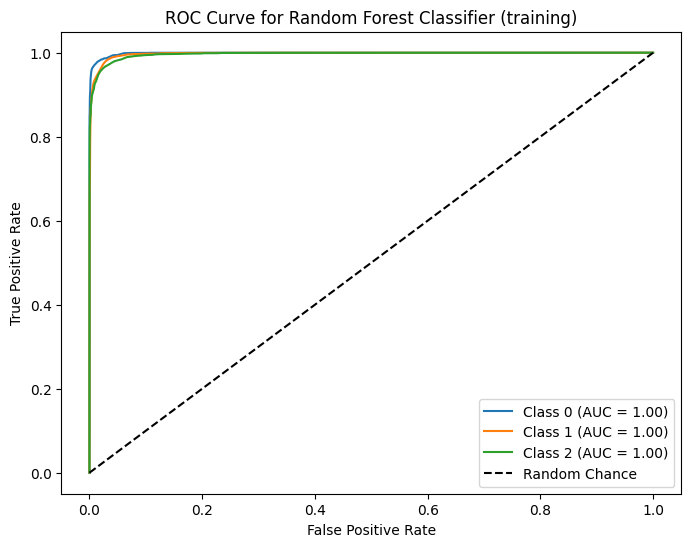
**Classification Report**

****

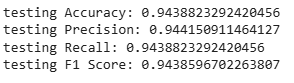
**Confusion Matrix**



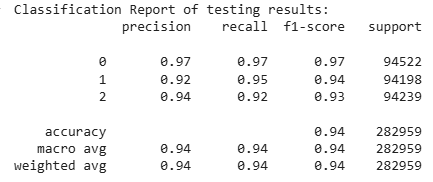
**Roc Curve**



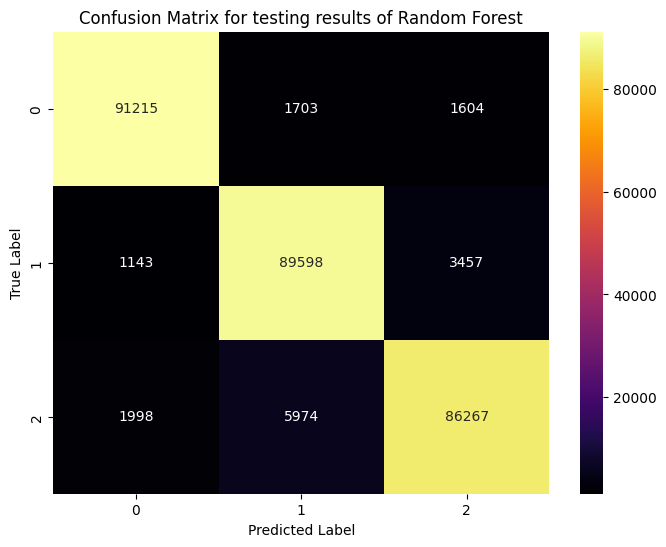
**RandomForest Test Result**

****

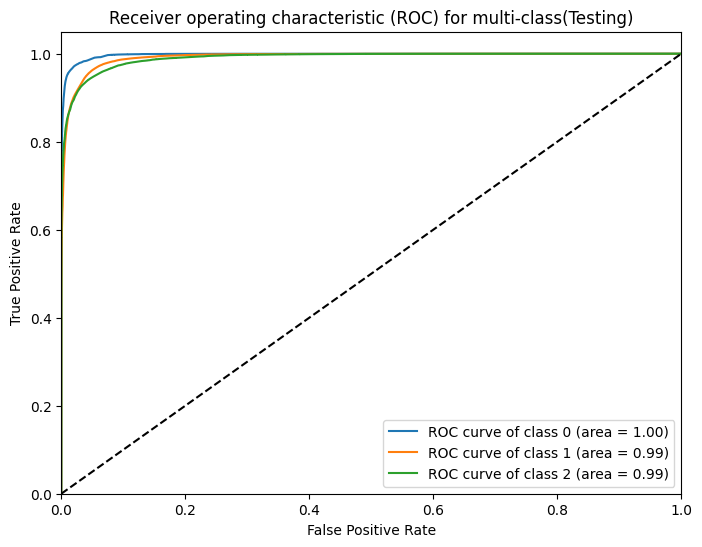
**Classification Report**

****

**Confusion Matrix**



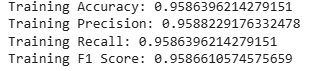
**Roc curve**



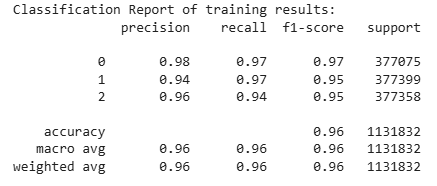
**Xgboost classifier**

Parameter: 

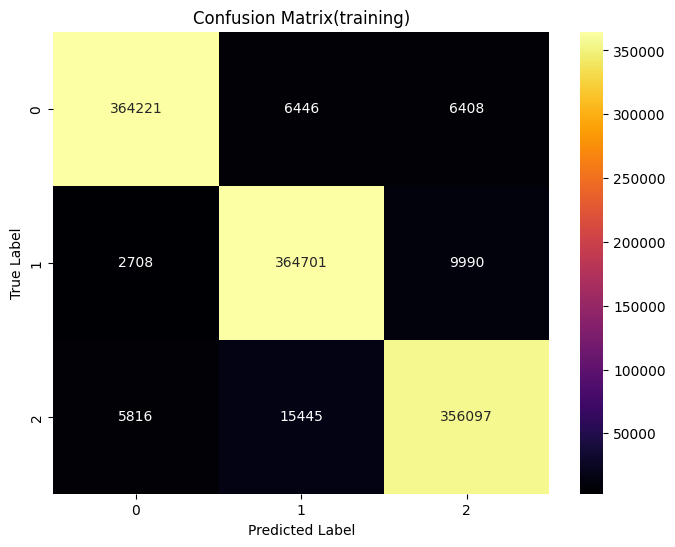
Training Result



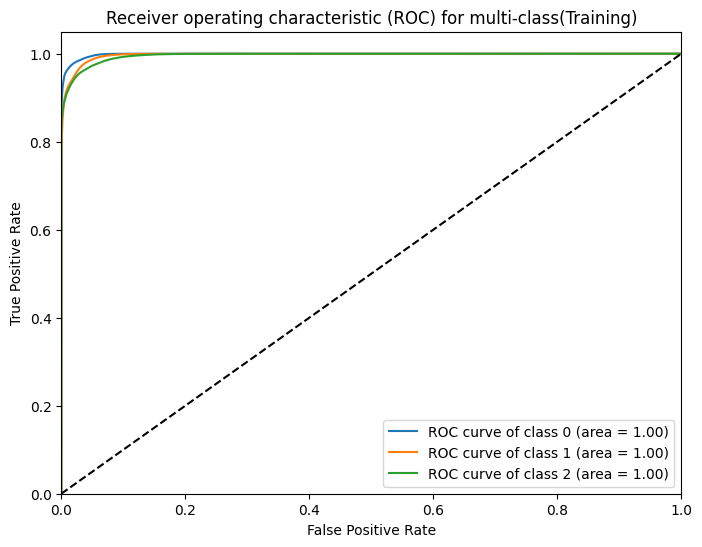
Classification Report



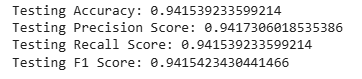
Confusion Matrix



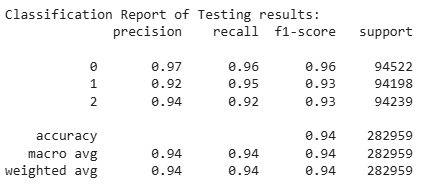
ROC Curve



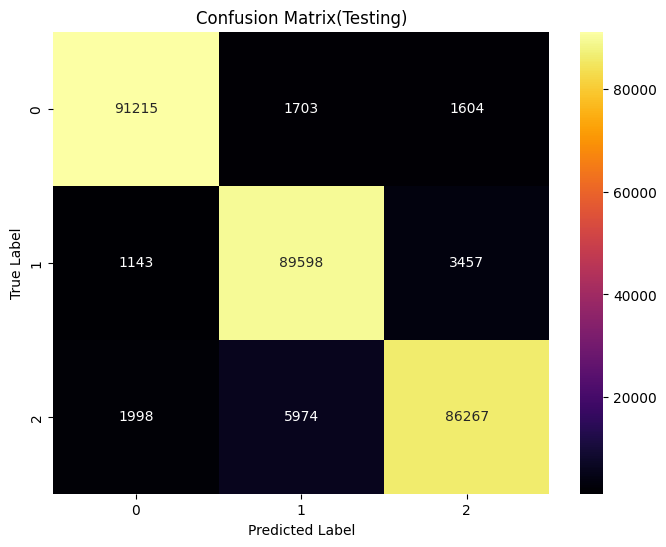
Xgboost Testing Result



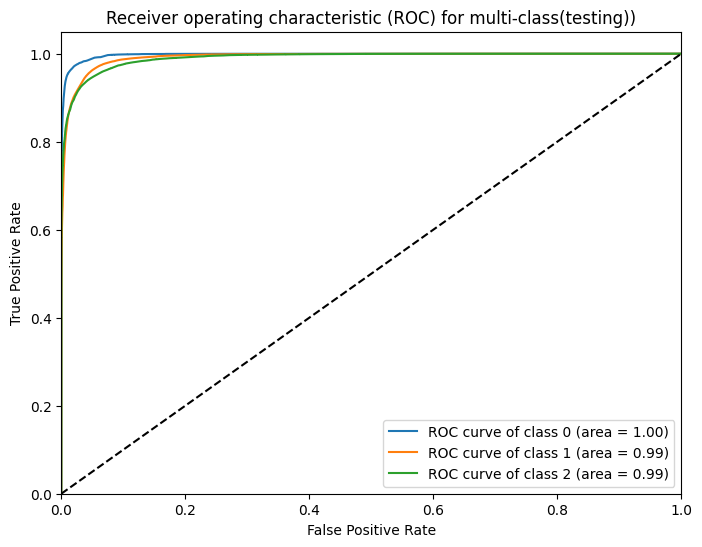
Classification Report



Confusion MAtrix



Roc curve



Comparision Graph:

