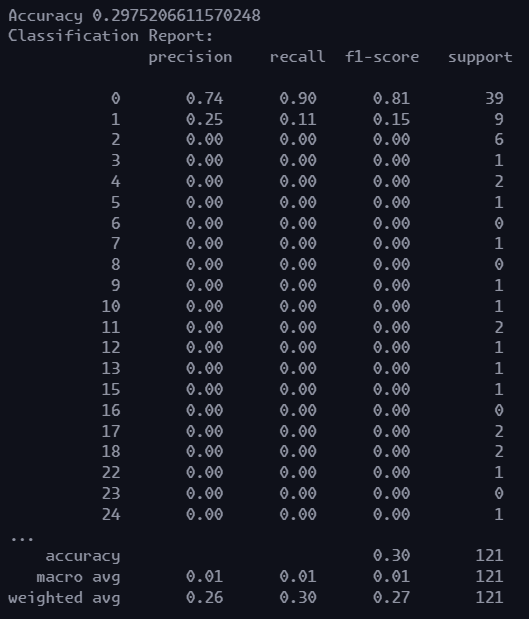
Input:

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'MKT\_CARRIER\_AIRLINE\_ID', 'OP\_CARRIER\_AIRLINE\_ID']

label **=** 'TOT\_DEN\_BOARDING'

model = RandomForestClassifier(n\_estimators=100, random\_state=42)

Output:

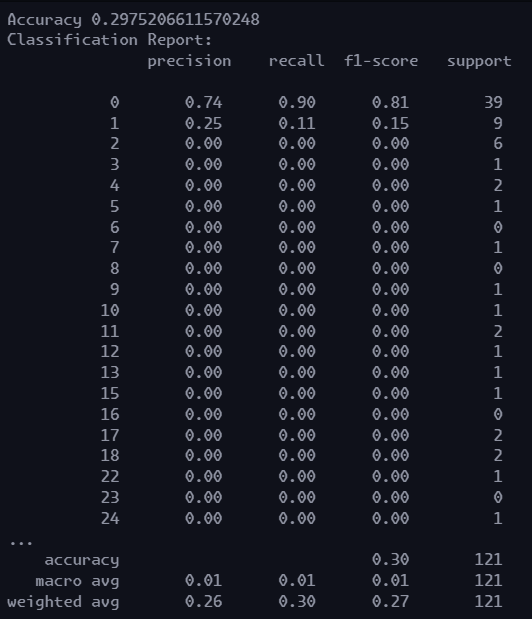


Change:

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'PAX\_UPGRADE', 'OP\_CARRIER\_AIRLINE\_ID']

label **=** 'TOT\_DEN\_BOARDING'

Output:

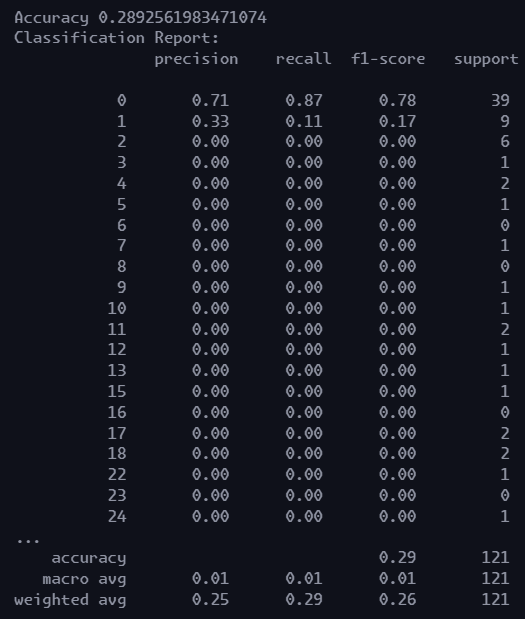


Changes:

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'MKT\_CARRIER\_AIRLINE\_ID', 'OP\_CARRIER\_AIRLINE\_ID', 'PAX\_DOWNGRADE']

label **=** 'TOT\_DEN\_BOARDING'

Output:



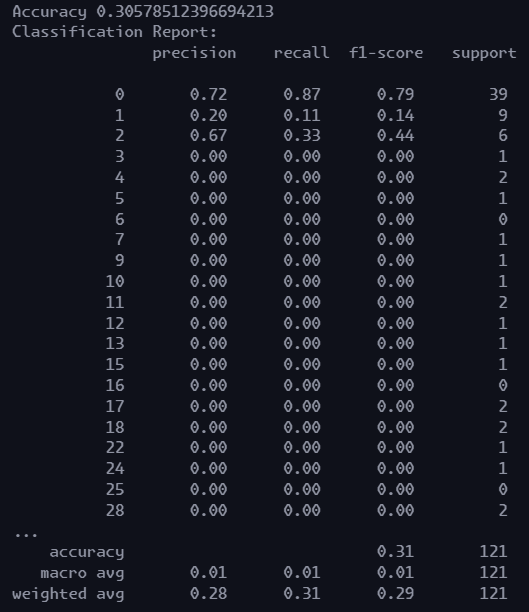
One reason passen Passengers would be denied on boarding would be because over booking. That means they’ll provide passengers high level passengers with alternative transport services And that is why I’ll be including alt transportation in the next model feature.

Changes:

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'MKT\_CARRIER\_AIRLINE\_ID', 'OP\_CARRIER\_AIRLINE\_ID', 'PAX\_DOWNGRADE', 'PAX\_ALT\_TRANS']

label **=** 'TOT\_DEN\_BOARDING'

Output:



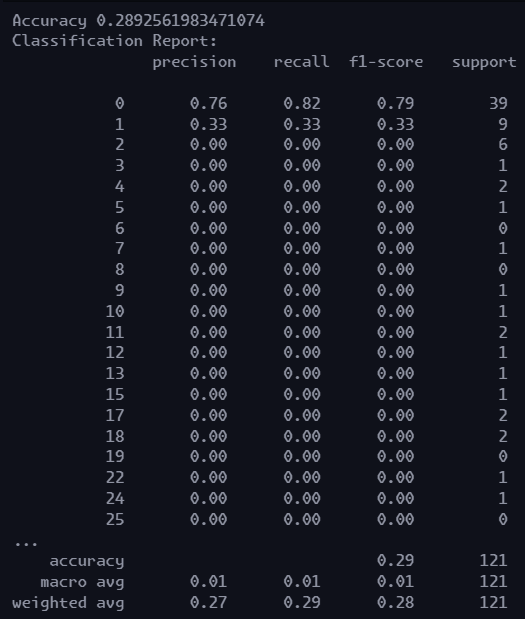
My assumption was somewhat correct

Now, lets remove something

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'PAX\_DOWNGRADE', 'PAX\_ALT\_TRANS']

label **=** 'TOT\_DEN\_BOARDING'

Output:



Accuracy dropped: The MARKET carrier ID, and Operation carrier ID contributed to accuracy

Now lets replace ID with name only

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'MKT\_CARRIER', 'OP\_CARRIER' , 'PAX\_DOWNGRADE', 'PAX\_ALT\_TRANS']

label **=** 'TOT\_DEN\_BOARDING'

But MKT\_CARRIER, and OP\_CARRIER are object type, so we’ll encode them first

*from* sklearn.preprocessing *import* LabelEncoder

*# Assuming df is your DataFrame*

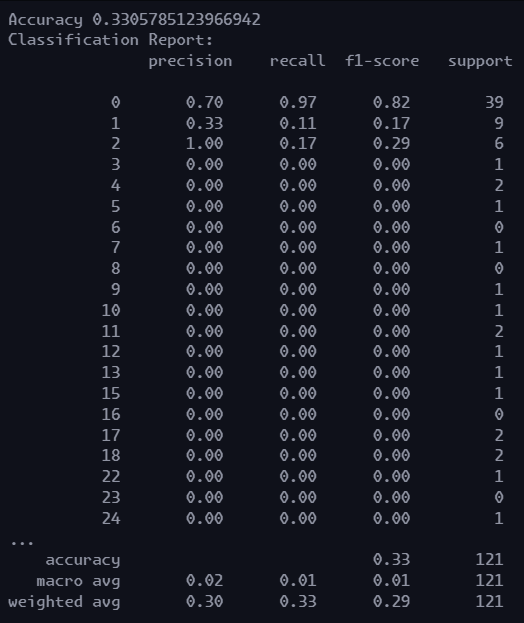
label\_encoder **=** LabelEncoder()

*# Convert MKT\_CARRIER and OP\_CARRIER to numerical values*

df['MKT\_CARRIER'] **=** label\_encoder.fit\_transform(df['MKT\_CARRIER'])

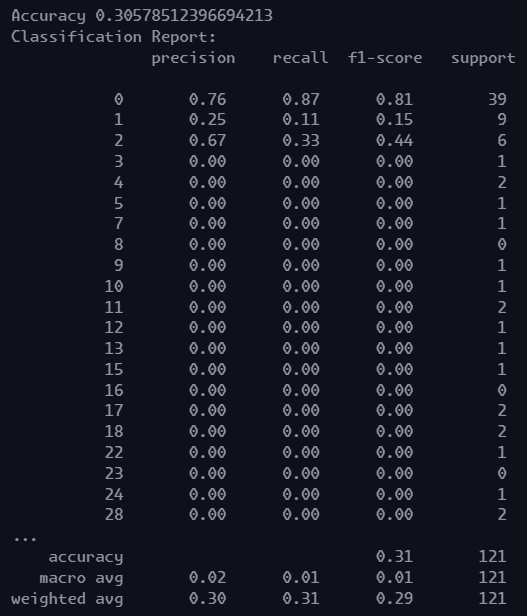
df['OP\_CARRIER'] **=** label\_encoder.fit\_transform(df['OP\_CARRIER'])

O/P:



Changes: added a new column in feature ‘TOT\_BOARDING’

O/P



Changes:

I created a new column called DENIAL\_RATIO which held the ratio of denied passengers to total passengers

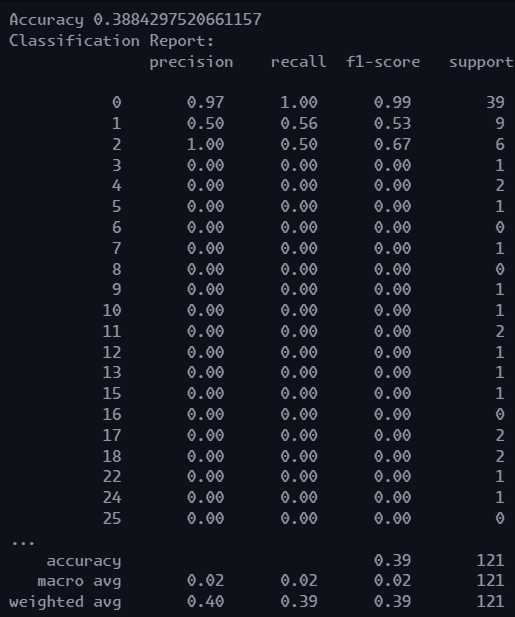
*# creating a new column, denial\_ratio*

df['DENIAL\_RATIO'] **=** df['TOT\_DEN\_BOARDING'] **/** df['TOT\_BOARDING']

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'MKT\_CARRIER', 'OP\_CARRIER' , 'PAX\_DOWNGRADE', 'PAX\_ALT\_TRANS', 'DENIAL\_RATIO']

label **=** 'TOT\_DEN\_BOARDING'

Output:



Changes

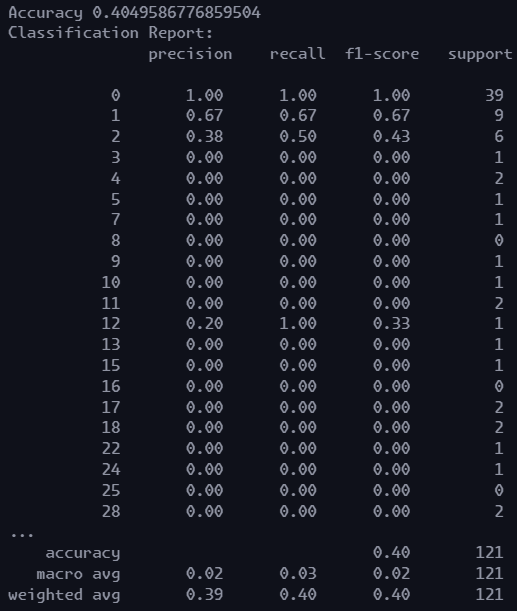
*# creating a new column, denial\_ratio*

df['DENIAL\_RATIO'] **=** df['TOT\_DEN\_BOARDING'] **/** df['TOT\_BOARDING']

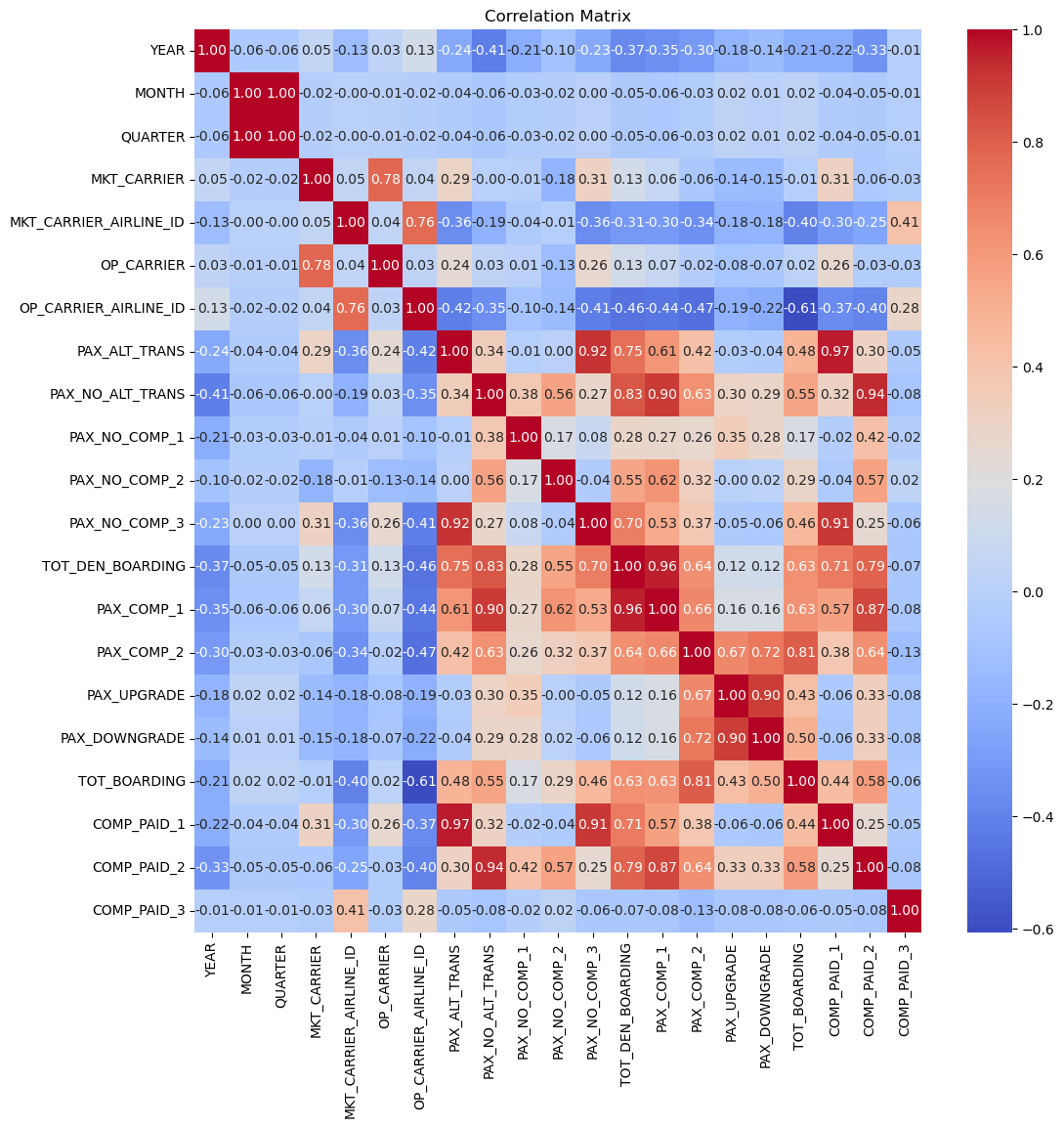
df['UPGRADE\_RATIO'] **=** df['PAX\_UPGRADE'] **/** df['TOT\_BOARDING']

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'MKT\_CARRIER', 'OP\_CARRIER' , 'PAX\_DOWNGRADE', 'PAX\_ALT\_TRANS', 'DENIAL\_RATIO', 'UPGRADE\_RATIO']

label **=** 'TOT\_DEN\_BOARDING'



I’ve should’ve done this before, but now to find out the relation between the label and features, I plotted a correlation matrix and this is the result

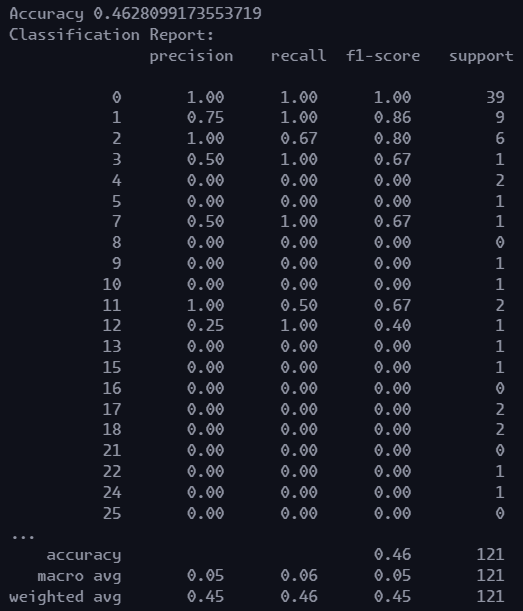


Correlation between PAX\_COMP\_2 and label is 0.96. let’s add that in feature

feature **=** ['YEAR', 'MONTH', 'QUARTER', 'MKT\_CARRIER', 'OP\_CARRIER' , 'PAX\_DOWNGRADE', 'PAX\_ALT\_TRANS', 'DENIAL\_RATIO', 'UPGRADE\_RATIO', 'PAX\_COMP\_1']

label **=** 'TOT\_DEN\_BOARDING'

Output:



Now what should we do?