ASSIGNMENT NO.6

Name:- Gawade Payal Dilip

Roll No:23107030

Class:-TY-AIDS(A)

In [51]:
```
import pandasaspd
import numpyasnp

# Text processing
from sklearn.feature_extraction.text import TfidfVectorizer

# Model from sklearn.model_selection import train_test_split from sklearn.naive_bayes import MultinomialNB from sklearn.linear_model import LogisticRegression from sklearn.ensemble import RandomForestClassifier # Evaluation from sklearn.metrics import accuracy_score, precision_score, recall_score, classifica
```

In [9]:
```
# Load dataset
df = pd.read_csv("Uber Customer Reviews.csv")
df
```

```
/tmp/ipykernel_10494/1145898452.py:2: DtypeWarning: Columns (3) have mixed types. Spec
ify dtype option on import or set low_memory=False.
  df = pd.read_csv("Uber Customer Reviews.csv")
```

Out[9]:

	source	review_id	user_name	review_title	review_description	rating	t
0	Google Play	18d6584c-d0e9-4833-a744-f607058aee97	Milky Way	NaN	Suddenly, the driver can't have my location an...	1	
1	Google Play	50a08f18-cece-4ddf-b617-028844c8aa28	Bradlee Severa	NaN	Very cordial.. And helped with a quick turnaro...	5	
2	Google Play	b0d8e75a-80a7-4dcd-abaf-72b046dbeeb7	Amit Aggarwal	NaN	Very good experience	5	
3	Google Play	502702a9-25ed-4373-a96c-7fa1f06caacd	Bryant Inman	NaN	All I use	5	
4	Google Play	f47a3fb6-23db-49bd-9e63-f33c8d724d07	Addie Whittaker	NaN	I have enjoyed traveling by Uber my drivers ha...	5	
...	...	...	...	...	...	...	
1069611	App Store	015547c9-1d97-4b92-8206-ef47a540b70b	Ad hater 20140323	Map problems	I tried to find away to report problems direct...	3	
1069612	App Store	e1125a24-a804-419e-8aa2-039e3f380d25	valeramos02	Quality decrease	I used to love Uber, specially the Uber pool s...	2	
1069613	App Store	132aac5d-10df-4207-a71d-01d81a4efde0	Janeé Brown	Uber pool walking blocks to get to the ride is...	If I wanted to take a bus to be dropped off on...	1	
1069614	App Store	99864769-f3f9-49fc-841e-3230a72fe18e	zachwiesler	TERRIBLE CUSTOMER SERVICE	Hello\n\nSTORY TIME\n\nI wanted to delete JUST...	1	
1069615	App Store	93f3188d-db2e-4532-bde3-6ec432558b5b	formerbaker1	Poor Customer Service	The past couple times I've ridden with Uber, I...	1	

1069616 rows × 13 columns

In [11]:

```
print(df.head())

# Check columns
print(df.columns)
```

```
          source                                review_id         user_name\
0  Google Play 18d6584c-d0e9-4833-a744-f607058aee97          Milky Way
1  Google Play 50a08f18-cece-4ddf-b617-028844c8aa28      Bradlee Severa
2  Google Play b0d8e75a-80a7-4dcd-abaf-72b046dbeeb7      Amit Aggarwal
3  Google Play 502702a9-25ed-4373-a96c-7fa1f06caacd       Bryant Inman
4  Google Play f47a3fb6-23db-49bd-9e63-f33c8d724d07  Addie Whittaker

  review_title                                   review_description  rating \
0          NaN Suddenly, the driver can't have my location an...       1
1          NaN Very cordial.. And helped with a quick turnaro...       5
2          NaN                                Very good experience       5
3          NaN                                             All I use       5
4          NaN I have enjoyed traveling by Uber my drivers ha...       5

   thumbs_up          review_date developer_response developer_response_date\
0        0.0  2023-08-10 17:48:51  0.0                NaN                    NaN
1        2023-08-10     17:38:35     0.0                NaN                    NaN
2        2023-08-10     17:38:17     0.0                NaN                    NaN
3        2023-08-10     17:37:45     0.0                NaN                    NaN
4        2023-08-10 17:36:56                              NaN                   NaN

          appVersion laguage_codecountry_code
0              NaN              en                in
1     4.485.10000               en                in
2     4.486.10002               en                in
3     4.467.10008               en                in
4    4.486.10002                en                in
Index(['source', 'review_id', 'user_name', 'review_title',
       'review_description','rating',          'thumbs_up',    'review_date',
       'developer_response','developer_response_date',          'appVersion',
       'laguage_code', 'country_code'],
      dtype='object')
```

In [21]:
```
df = df[['review_description', 'rating']]
df = df.dropna()
```

In [23]:
```
df = df[df['rating'] != 3]
```

In [25]:
```
def convert_sentiment(rating):
    if rating <= 2:
        return 0       # Negative
    else:
        return 1       # Positive

df['sentiment'] = df['rating'].apply(convert_sentiment)

print(df['sentiment'].value_counts())
```

```
sentiment
1    693940
0    325590
Name: count, dtype: int64
```

In [27]:
```
X = df['review_description']
y = df['sentiment']

X_train, X_test, y_train, y_test = train_test_split(
    X, y, test_size=0.2, random_state=42
)
```

In [29]:
```
vectorizer = TfidfVectorizer(stop_words='english', max_features=5000)

X_train_tfidf = vectorizer.fit_transform(X_train)
X_test_tfidf = vectorizer.transform(X_test)
```

In [35]:
```
model = LogisticRegression(max_iter=1000)

model.fit(X_train_tfidf, y_train)
```

Out[35]:

LogisticRegression

LogisticRegression(max_iter=1000)

In [39]:
```
df = pd.read_csv("Uber Customer Reviews.csv", low_memory=False)

print(df[['review_description', 'rating']].head())
```

```
                                    review_description  rating
0  Suddenly,thedriver  can't  havemy  location an...       1
1  Very  cordial.. Andhelpedwith aquickturnaro...       5
2                                Very goodexperience       5
3                                        All I use         5
4  I  haveenjoyed  travelingbyUbermy  drivers ha...       5
```

In [41]:
```
y_pred = model.predict(X_test_tfidf)
```

In [43]:
```
y_pred
```

Out[43]:
```
array([1, 1, 1, ..., 0, 0, 0])
```

In [45]:
```
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred)
recall = recall_score(y_test, y_pred)

print("Accuracy :", accuracy)
print("Precision:", precision)
print("Recall    :", recall)

print("\nClassification Report:\n")
print(classification_report(y_test, y_pred))
```

```
Accuracy  :  0.9288888017027453
Precision:  0.9462504392475779
Recall : 0.9495872705159298

Classification Report:

              precision    recall  f1-score   support

           0       0.89      0.88      0.89     64953
           1       0.95      0.95      0.95    138953

    accuracy                           0.93    203906
   macro avg       0.92      0.92      0.92    203906
weighted avg       0.93      0.93      0.93    203906
```

In [47]:
```
sample_review = ["Driver was very rude and cancelled the ride"]

sample_tfidf = vectorizer.transform(sample_review)

prediction = model.predict(sample_tfidf)

if prediction[0] == 1:
    print("Positive Review 😊")
else:
    print("Negative Review 😡")
```

```
Negative Review 😡
```

In [ ]: