Report Of Tweet

1. Introduction

Here i get the problem from the dataset where a tweet is extracted from twitter . here i got the dataset of tweet_id,id,hastage_generation_time,tweet_created,screen_name,name,user_description,entities,tweet_source,verifed. Where i had analyze the data in starting and i found the problem in the dataset of user_description found the stopword using id is more , i also find the problem why it has many false in verified for that i had gone for the analysis . I got the motivation of tweet analyses by the twitter as many on use this app and it has been very interesting for me to give the analysis on this.

2. Library Used

The library which i used in this analysis is numpy, pandas for utilities , seaborn,worldclod and matplotib for plottiing , wordnetlemmatizer,porter stemma and stopword from nltk for natural language processing ,and some from sklearn.

```python
# utilities
import re
import numpy as np
import pandas as pd
# plotting
import seaborn as sns
from wordcloud import WordCloud
import matplotlib.pyplot as plt
# nltk
from nltk.stem import WordNetLemmatizer
# sklearn
from sklearn.svm import LinearSVC
from sklearn.naive_bayes import BernoulliNB
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import train_test_split
from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.metrics import confusion_matrix, classification_report
```

3. Experiment output

After all the analyses i feel that there many id who iphone and android user for the tweet and mostly they are from india and canada . finally it is showing all false verified whos tweet is not understandable . Almost all Indian who is supporting sushant shing rajput in their tweet and telling about them . that what i found and yes i also found the Accuracy: As far as the accuracy of the model is concerned Logistic Regression performs better than SVM which in turn performs better than Bernoulli Naive Bayes.

In our problem statement, Logistic Regression is following the principle of Occam's Razor which defines that for a particular problem statement if the data has no assumption, then the simplest model works the best. Since our dataset does not have any assumptions and Logistic Regression is a simple model, therefore the concept holds true for the above-mentioned dataset.

We, therefore, conclude that the Logistic Regression is the best model for the above-given dataset.