Toxicity Analysis Using NLP

Subrat Kumar Samal

Dataset(s)

The data set used is “train.csv” as in Toxicity Project Assessment

Motivation

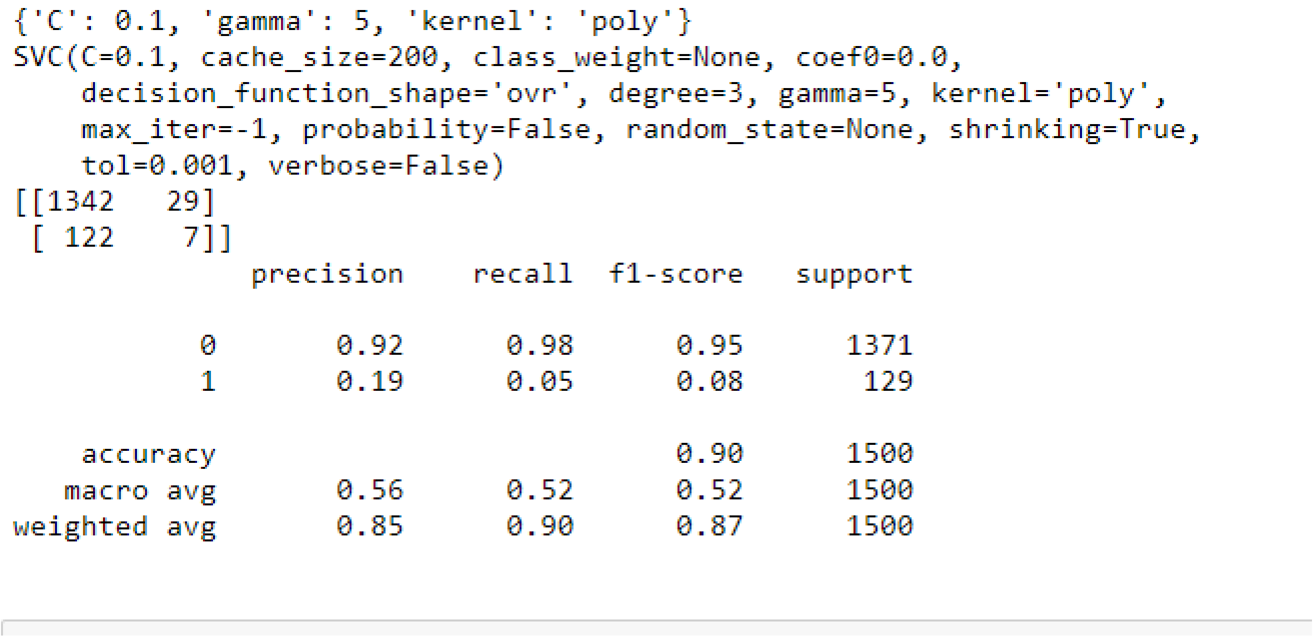
- This project is an Exploratory Data analysis on the toxic comments using NLP and Machine learning Techniques.
- The unstructured data was clea.ned using NLP techniques so as to be made available for analysis using Machine learning algorithms

Findings

- The Toxic: Non-toxic comments as in the input file is approximately 1:10
- Upon tokenization and finding top 5 contextual words as

  ['please', 'talk', 'wikipedia', 'page', 'article’]
- The values for Recall and Precision for class 1 are always 0 on using SVC.The result is same when model is used for Train data as well. This implies that the model seems to focus on the 0s only. So we need to tune the hyperparameters
- Upon using GridSearch and StratifiedKFold we see improvement in Recall values along with optimized accuracy of 91% for the optimized SVM model

- Test

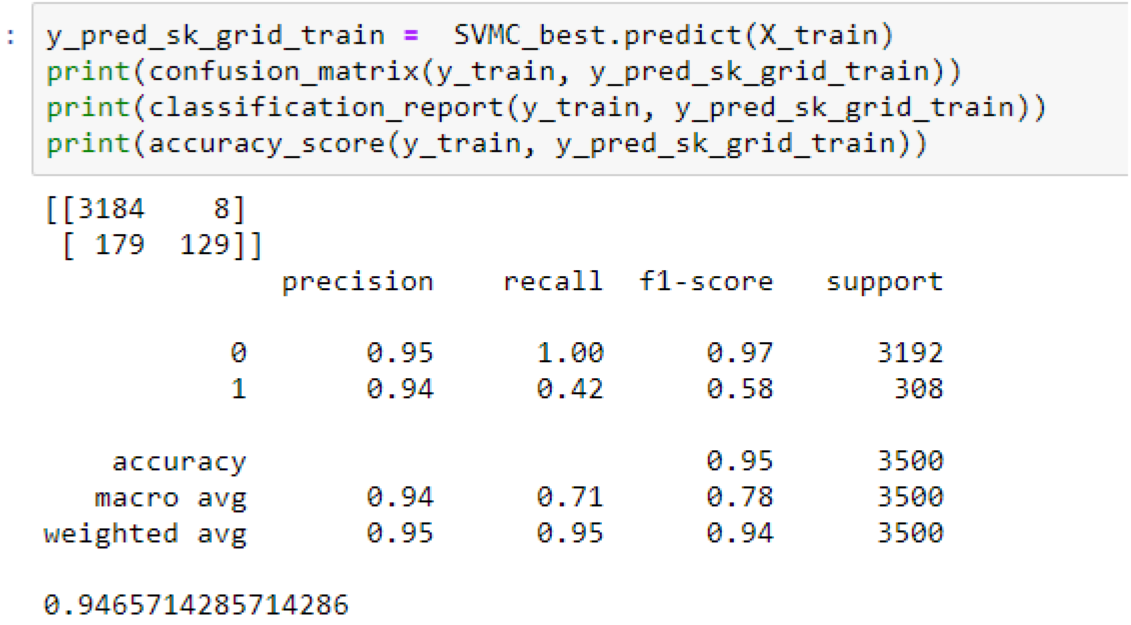


```
{'C': 0.1, 'gamma': 5, 'kernel': 'poly'}
SVC(C=0.1, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape='ovr', degree=3, gamma=5, kernel='poly',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
[[1342   29]
 [ 122    7]]
              precision    recall  f1-score   support

           0       0.92      0.98      0.95      1371
           1       0.19      0.05      0.08       129

    accuracy                           0.90      1500
   macro avg       0.56      0.52      0.52      1500
weighted avg       0.85      0.90      0.87      1500
```

- Train



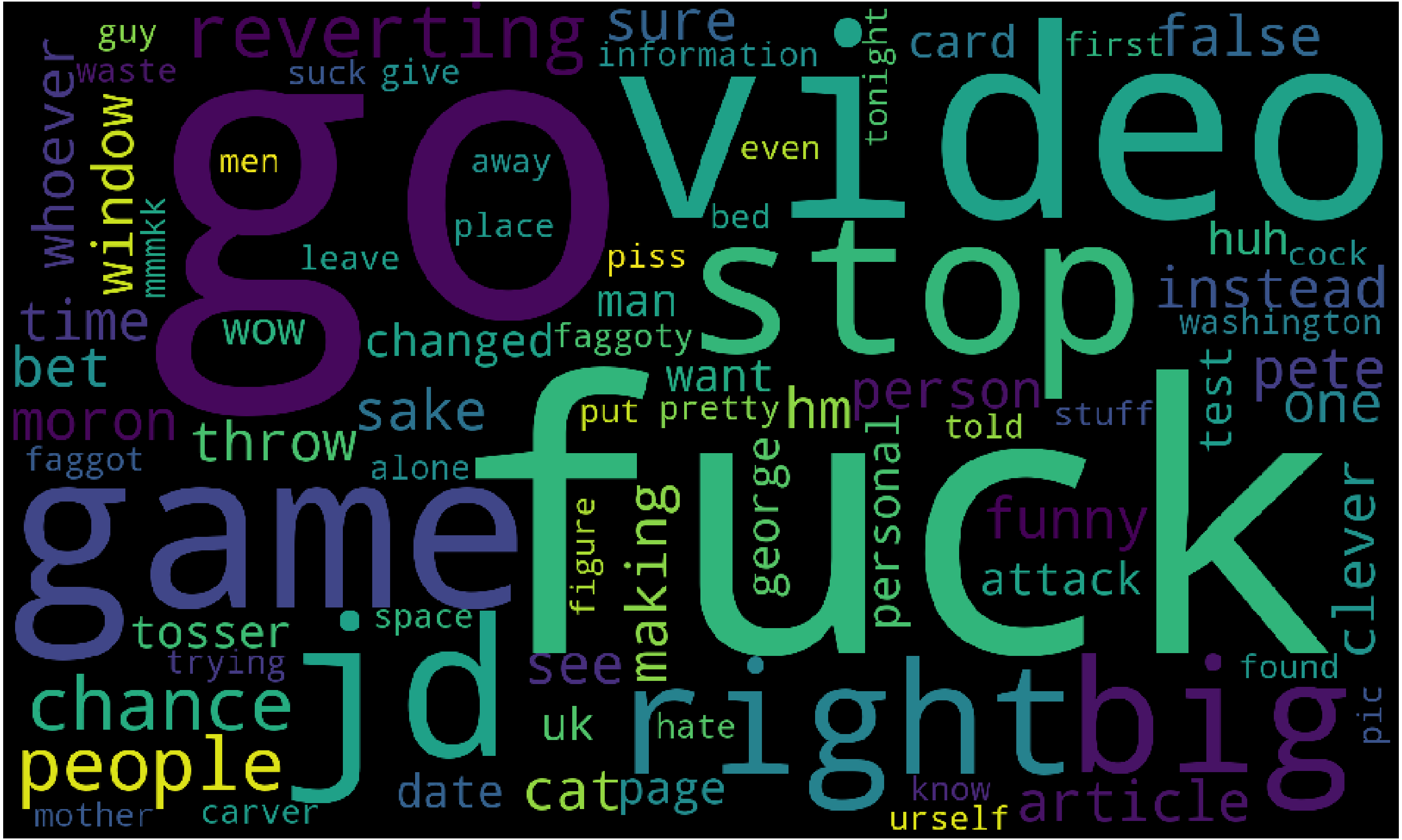
```
y_pred_sk_grid_train =  SVMC_best.predict(X_train)
print(confusion_matrix(y_train, y_pred_sk_grid_train))
print(classification_report(y_train, y_pred_sk_grid_train))
print(accuracy_score(y_train, y_pred_sk_grid_train))
```

```
[[3184    8]
 [ 179  129]]
              precision    recall  f1-score   support

           0       0.95      1.00      0.97      3192
           1       0.94      0.42      0.58       308

    accuracy                           0.95      3500
   macro avg       0.94      0.71      0.78      3500
weighted avg       0.95      0.95      0.94      3500

0.9465714285714286
```


go
video
fuck
stop
game
jd
right
big
reverting
sure
guy
waste
whoever
window
mmmkk
suck
give
information
tonight
card
first
false
men
away
even
bed
place
piss
leave
huh
cock
instead
washington
time
wow
changed
faggoty
man
bet
want
person
pete
test
one
moron
sake
hm
put
pretty
stuff
told
throw
faggot
alone
george
personal
funny
clever
making
figure
attack
tosser
space
see
trying
found
chance
uk
hate
pic
people
date
cat
page
know
article
mother
carver
urself