SVM on Amazon Reviews

June 25, 2018

```
In [ ]: import sqlite3
        import pandas as pd
        import numpy as np
        import matplotlib.pyplot as plt
        import plotly.plotly as py
        import plotly.graph_objs as go
        from sklearn.cross_validation import train_test_split
        from sklearn.neighbors import KNeighborsClassifier
        from sklearn.metrics import accuracy_score
        from sklearn.cross_validation import cross_val_score
        from collections import Counter
        from sklearn.metrics import accuracy_score
        from sklearn import cross_validation
        from sklearn.feature_extraction.text import TfidfTransformer
        from sklearn.feature_extraction.text import TfidfVectorizer
        from sklearn.feature_extraction.text import CountVectorizer
        from sklearn.metrics import confusion_matrix
        from sklearn import metrics
        from sklearn.metrics import roc_curve, auc
        from nltk.stem.porter import PorterStemmer
        from sklearn.decomposition import TruncatedSVD
        from sklearn.model_selection import TimeSeriesSplit
        from sklearn.model_selection import GridSearchCV
        from sklearn.linear_model import LogisticRegression
```

---

1 Loading and Sampling the dataset(10k data-points)

```
In [0]: final = pd.read_csv("final.csv")


        final_data = final.sample(n = 10000)

        final_data = final_data.drop(["Text"], axis = 1)
        final_data = final_data.drop(final_data.columns[0], axis = 1)
```

```
In [5]: labels = final_data.Score
        final_data = final_data.sort_values("Time")
        final_data.shape
```

```
Out[5]: (10000, 10)
```

2 Train/Test Split

```
In [0]: n = final_data.shape[0]
        train_size = 0.7

        train_set = final_data.iloc[:int(n*train_size)]
        test_set =final_data.iloc[int(n*train_size):]

        X_train = train_set.CleanedText
        y_train = train_set.Score

        X_test = test_set.CleanedText
        y_test= test_set.Score
```

---

3 Bag of words Vectorization

```
In [28]: count_vect = CountVectorizer() #in scikit-learn
         X_train1 = count_vect.fit_transform(X_train)
         X_test1 = count_vect.transform(X_test)
```

```
In [29]: #Normalize
         from sklearn.preprocessing import normalize
         X_train1 = normalize(X_train1)
         X_test1  = normalize(X_test1)
```

3.1 GridSearch Cross-Validation

```
In [30]: from sklearn.model_selection import GridSearchCV
         from sklearn.svm import SVC


         Cs = [0.001, 0.01, 0.1, 1, 10]
         gammas = [0.01, 0.1, 1, 10,100]
         param_grid = {'C': Cs, 'gamma' : gammas}
         rbf_svm = SVC(kernel='rbf')
         rbf_svm_grid_cv = GridSearchCV(rbf_svm, param_grid=param_grid, n_jobs=-1, verbose=30)
         rbf_svm_grid_cv
```

```
Out[30]: GridSearchCV(cv=None, error_score='raise',
                estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
```

```
            decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
            max_iter=-1, probability=False, random_state=None, shrinking=True,
            tol=0.001, verbose=False),
                 fit_params=None, iid=True, n_jobs=-1,
                 param_grid={'C': [0.001, 0.01, 0.1, 1, 10], 'gamma': [0.01, 0.1, 1, 10, 100]},
                 pre_dispatch='2*n_jobs', refit=True, return_train_score='warn',
                 scoring=None, verbose=30)
```

In [31]:
```
#perform gridsearch
rbf_svm_grid_cv.fit(X_train1, y_train)
```

```
Fitting 3 folds for each of 25 candidates, totalling 75 fits
[CV] C=0.001, gamma=0.01 ...
[CV] C=0.001, gamma=0.01 ...
[CV] ... C=0.001, gamma=0.01, score=0.8474721508140531, total=   8.6s
[CV] C=0.001, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done   1 tasks      | elapsed:   15.0s


[CV] ... C=0.001, gamma=0.01, score=0.8474721508140531, total=   9.2s
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:   15.5s


[CV] ... C=0.001, gamma=0.01, score=0.847770154373928, total=   9.7s
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   3 tasks      | elapsed:   30.9s


[CV] ... C=0.001, gamma=0.1, score=0.8474721508140531, total=   9.6s
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   4 tasks      | elapsed:   31.3s


[CV] ... C=0.001, gamma=0.1, score=0.8474721508140531, total=   9.7s
[CV] C=0.001, gamma=1 ...


[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:   46.8s
```

```
[CV] ... C=0.001, gamma=0.1, score=0.847770154373928, total=   9.8s
[CV] C=0.001, gamma=1 ...


[Parallel(n_jobs=-1)]: Done   6 tasks      | elapsed:   47.4s


[CV] ... C=0.001, gamma=1, score=0.8474721508140531, total=  11.3s
[CV] C=0.001, gamma=1 ...


[Parallel(n_jobs=-1)]: Done   7 tasks      | elapsed:  1.1min


[CV] ... C=0.001, gamma=1, score=0.8474721508140531, total=  11.2s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done   8 tasks      | elapsed:  1.1min


[CV] ... C=0.001, gamma=1, score=0.847770154373928, total=  11.3s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:  1.4min


[CV] ... C=0.001, gamma=10, score=0.8474721508140531, total=  12.3s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:  1.4min


[CV] ... C=0.001, gamma=10, score=0.8474721508140531, total=  12.1s
[CV] C=0.001, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  11 tasks      | elapsed:  1.7min


[CV] ... C=0.001, gamma=10, score=0.847770154373928, total=  12.2s
[CV] C=0.001, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  12 tasks      | elapsed:  1.8min


[CV] ... C=0.001, gamma=100, score=0.8474721508140531, total=  13.5s
[CV] C=0.001, gamma=100 ...
```

```
[Parallel(n_jobs=-1)]: Done  13 tasks      | elapsed:  2.1min


[CV] ... C=0.001, gamma=100, score=0.8474721508140531, total=  13.4s
[CV] C=0.01, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  14 tasks      | elapsed:  2.1min


[CV] ... C=0.01, gamma=0.01, score=0.8474721508140531, total=   9.9s
[CV] C=0.01, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  15 tasks      | elapsed:  2.4min


[CV] ... C=0.001, gamma=100, score=0.847770154373928, total=  13.7s
[CV] C=0.01, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:  2.5min


[CV] ... C=0.01, gamma=0.01, score=0.8474721508140531, total=   9.7s
[CV] C=0.01, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:  2.7min


[CV] ... C=0.01, gamma=0.01, score=0.847770154373928, total=   9.9s
[CV] C=0.01, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  18 tasks      | elapsed:  2.8min


[CV] ... C=0.01, gamma=0.1, score=0.8474721508140531, total=  12.1s
[CV] C=0.01, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  19 tasks      | elapsed:  3.0min


[CV] ... C=0.01, gamma=0.1, score=0.8474721508140531, total=  12.0s
[CV] C=0.01, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  20 tasks      | elapsed:  3.1min
```

```
[CV] ... C=0.01, gamma=0.1, score=0.847770154373928, total=  11.6s
[CV] C=0.01, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  21 tasks      | elapsed:  3.3min

[CV] ... C=0.01, gamma=1, score=0.8474721508140531, total=  17.1s
[CV] C=0.01, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  22 tasks      | elapsed:  3.5min

[CV] ... C=0.01, gamma=1, score=0.8474721508140531, total=  15.7s
[CV] C=0.01, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  23 tasks      | elapsed:  3.7min

[CV] ... C=0.01, gamma=1, score=0.847770154373928, total=  17.9s
[CV] C=0.01, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:  4.0min

[CV] ... C=0.01, gamma=10, score=0.8474721508140531, total=  32.2s
[CV] C=0.01, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:  4.6min

[CV] ... C=0.01, gamma=10, score=0.8474721508140531, total=  31.7s
[CV] C=0.01, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  26 tasks      | elapsed:  4.9min

[CV] ... C=0.01, gamma=10, score=0.847770154373928, total=  31.8s
[CV] C=0.01, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  27 tasks      | elapsed:  5.5min

[CV] ... C=0.01, gamma=100, score=0.8474721508140531, total=  29.5s
[CV] C=0.01, gamma=100 ...
```

```
[Parallel(n_jobs=-1)]: Done  28 tasks      | elapsed:  5.7min


[CV] ... C=0.01, gamma=100, score=0.8474721508140531, total=  29.4s
[CV] C=0.1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  29 tasks      | elapsed:  6.3min


[CV] ... C=0.01, gamma=100, score=0.847770154373928, total=  30.2s
[CV] C=0.1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  30 tasks      | elapsed:  6.6min


[CV] ... C=0.1, gamma=0.01, score=0.8474721508140531, total=  13.3s
[CV] C=0.1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  31 tasks      | elapsed:  6.7min


[CV] ... C=0.1, gamma=0.01, score=0.8474721508140531, total=  13.1s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  32 tasks      | elapsed:  6.9min


[CV] ... C=0.1, gamma=0.01, score=0.847770154373928, total=  13.5s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  33 tasks      | elapsed:  7.0min


[CV] ... C=0.1, gamma=0.1, score=0.8474721508140531, total=  13.2s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  34 tasks      | elapsed:  7.2min


[CV] ... C=0.1, gamma=0.1, score=0.8474721508140531, total=  12.6s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  35 tasks      | elapsed:  7.3min
```

```
[CV] ... C=0.1, gamma=0.1, score=0.847770154373928, total=  13.6s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  36 tasks      | elapsed:  7.6min

[CV] ... C=0.1, gamma=1, score=0.8474721508140531, total=  17.8s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  37 tasks      | elapsed:  7.8min

[CV] ... C=0.1, gamma=1, score=0.8474721508140531, total=  16.2s
[CV] C=0.1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  38 tasks      | elapsed:  8.0min

[CV] ... C=0.1, gamma=1, score=0.847770154373928, total=  16.9s
[CV] C=0.1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  39 tasks      | elapsed:  8.3min

[CV] ... C=0.1, gamma=10, score=0.8474721508140531, total=  28.4s
[CV] C=0.1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  40 tasks      | elapsed:  8.8min

[CV] ... C=0.1, gamma=10, score=0.8474721508140531, total=  28.0s
[CV] C=0.1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  41 tasks      | elapsed:  9.1min

[CV] ... C=0.1, gamma=10, score=0.847770154373928, total=  28.2s
[CV] C=0.1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed:  9.6min

[CV] ... C=0.1, gamma=100, score=0.8474721508140531, total=  28.3s
[CV] C=0.1, gamma=100 ...
```

```
[Parallel(n_jobs=-1)]: Done  43 tasks      | elapsed:  9.9min


[CV] ... C=0.1, gamma=100, score=0.8474721508140531, total=  30.5s
[CV] C=1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  44 tasks      | elapsed: 10.5min


[CV] ... C=0.1, gamma=100, score=0.847770154373928, total=  29.9s
[CV] C=1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  45 tasks      | elapsed: 10.7min


[CV] ... C=1, gamma=0.01, score=0.8474721508140531, total=  13.5s
[CV] C=1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  46 tasks      | elapsed: 10.8min


[CV] ... C=1, gamma=0.01, score=0.8474721508140531, total=  13.0s
[CV] C=1, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  47 tasks      | elapsed: 11.1min


[CV] ... C=1, gamma=0.01, score=0.847770154373928, total=  13.7s
[CV] C=1, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  48 tasks      | elapsed: 11.2min


[CV] ... C=1, gamma=0.1, score=0.8474721508140531, total=  13.0s
[CV] C=1, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  49 tasks      | elapsed: 11.4min


[CV] ... C=1, gamma=0.1, score=0.8483290488431876, total=  12.4s
[CV] C=1, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  50 tasks      | elapsed: 11.5min
```

```
[CV] ... C=1, gamma=0.1, score=0.847770154373928, total=  13.1s
[CV] C=1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  51 tasks      | elapsed: 11.7min

[CV] ... C=1, gamma=1, score=0.8753213367609255, total=  18.0s
[CV] C=1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  52 tasks      | elapsed: 12.0min

[CV] ... C=1, gamma=1, score=0.87146529562982, total=  17.3s
[CV] C=1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  53 tasks      | elapsed: 12.2min

[CV] ... C=1, gamma=1, score=0.880360205831904, total=  17.0s
[CV] C=1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  54 tasks      | elapsed: 12.4min

[CV] ... C=1, gamma=10, score=0.8474721508140531, total=  30.5s
[CV] C=1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  55 tasks      | elapsed: 13.1min

[CV] ... C=1, gamma=10, score=0.8474721508140531, total=  31.9s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  56 tasks      | elapsed: 13.3min

[CV] ... C=1, gamma=10, score=0.847770154373928, total=  32.0s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  57 tasks      | elapsed: 14.0min

[CV] ... C=1, gamma=100, score=0.8474721508140531, total=  32.8s
[CV] C=1, gamma=100 ...
```

```
[Parallel(n_jobs=-1)]: Done  58 tasks      | elapsed: 14.2min


[CV] ... C=1, gamma=100, score=0.8474721508140531, total=  31.6s
[CV] C=10, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  59 tasks      | elapsed: 14.8min


[CV] ... C=1, gamma=100, score=0.847770154373928, total=  31.8s
[CV] C=10, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  60 tasks      | elapsed: 15.1min


[CV] ... C=10, gamma=0.01, score=0.8479005998286204, total=  13.7s
[CV] C=10, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  61 tasks      | elapsed: 15.2min


[CV] ... C=10, gamma=0.01, score=0.8483290488431876, total=  13.1s
[CV] C=10, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  62 tasks      | elapsed: 15.5min


[CV] ... C=10, gamma=0.01, score=0.8486277873070326, total=  13.9s
[CV] C=10, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  63 tasks      | elapsed: 15.6min


[CV] ... C=10, gamma=0.1, score=0.9005998286203942, total=  13.5s
[CV] C=10, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  64 tasks      | elapsed: 15.8min


[CV] ... C=10, gamma=0.1, score=0.8873179091688089, total=  12.9s
[CV] C=10, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  65 tasks      | elapsed: 15.9min
```

```
[CV] ... C=10, gamma=0.1, score=0.9005145797598628, total=  12.9s
[CV] C=10, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  66 tasks      | elapsed: 16.2min


[CV] ... C=10, gamma=1, score=0.9023136246786633, total=  26.3s
[CV] C=10, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  67 tasks      | elapsed: 16.6min


[CV] ... C=10, gamma=1, score=0.8881748071979434, total=  24.6s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  68 tasks      | elapsed: 16.8min


[CV] ... C=10, gamma=1, score=0.8945111492281304, total=  26.3s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  69 tasks      | elapsed: 17.2min


[CV] ... C=10, gamma=10, score=0.8474721508140531, total=  34.6s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  70 tasks      | elapsed: 17.7min


[CV] ... C=10, gamma=10, score=0.8474721508140531, total=  33.2s
[CV] C=10, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  71 tasks      | elapsed: 18.1min


[CV] ... C=10, gamma=10, score=0.847770154373928, total=  35.0s
[CV] C=10, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  72 tasks      | elapsed: 18.7min
```

```
[CV] ... C=10, gamma=100, score=0.8474721508140531, total=  35.4s
[CV] C=10, gamma=100 ...
[CV] ... C=10, gamma=100, score=0.8474721508140531, total=  34.5s
[CV] ... C=10, gamma=100, score=0.847770154373928, total=  34.5s


[Parallel(n_jobs=-1)]: Done  75 out of  75 | elapsed: 19.9min remaining:    0.0s
[Parallel(n_jobs=-1)]: Done  75 out of  75 | elapsed: 19.9min finished


Out[31]: GridSearchCV(cv=None, error_score='raise',
             estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
           decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
           max_iter=-1, probability=False, random_state=None, shrinking=True,
           tol=0.001, verbose=False),
             fit_params=None, iid=True, n_jobs=-1,
             param_grid={'C': [0.001, 0.01, 0.1, 1, 10], 'gamma': [0.01, 0.1, 1, 10, 100]},
             pre_dispatch='2*n_jobs', refit=True, return_train_score='warn',
             scoring=None, verbose=30)
```

```
In [33]: print("Best parameters: ", rbf_svm_grid_cv.best_params_)
         print("Best cross-validation score: {:.3f}".format(rbf_svm_grid_cv.best_score_))
```

```
Best parameters:  {'C': 10, 'gamma': 0.1}
Best cross-validation score: 0.896
```

```
In [35]: from sklearn.metrics import classification_report

         y_pred = rbf_svm_grid_cv.predict(X_test1)
         print(classification_report(y_test, y_pred))
```

```
             precision    recall  f1-score   support

   negative       0.83      0.54      0.65       540
   positive       0.91      0.98      0.94      2460

avg / total       0.89      0.90      0.89      3000
```

3.2 RandomSearch Cross-Validation

```
In [37]: from sklearn.model_selection import RandomizedSearchCV
         from scipy import stats

         param_grid  = {"C": stats.uniform(0.001, 10), "gamma": stats.uniform(0.1, 100)}

         rbf_svm = SVC(kernel='rbf')
         rbf_svm_grid_cv =  RandomizedSearchCV(rbf_svm, param_grid, n_jobs=-1, verbose=30)
         rbf_svm_grid_cv
```

```
Out[37]: RandomizedSearchCV(cv=None, error_score='raise',
                   estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
           decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
           max_iter=-1, probability=False, random_state=None, shrinking=True,
           tol=0.001, verbose=False),
                   fit_params=None, iid=True, n_iter=10, n_jobs=-1,
                   param_distributions={'C': <scipy.stats._distn_infrastructure.rv_frozen obje
                   pre_dispatch='2*n_jobs', random_state=None, refit=True,
                   return_train_score='warn', scoring=None, verbose=30)

In [38]: #perform gridsearch
         rbf_svm_grid_cv.fit(X_train1, y_train)

Fitting 3 folds for each of 10 candidates, totalling 30 fits
[CV] C=7.344875389757782, gamma=59.593754547849436 ...
[CV] C=7.344875389757782, gamma=59.593754547849436 ...
[CV]  C=7.344875389757782, gamma=59.593754547849436, score=0.8474721508140531, total=  15.2s
[CV] C=7.344875389757782, gamma=59.593754547849436 ...


[Parallel(n_jobs=-1)]: Done   1 tasks      | elapsed:   25.5s


[CV]  C=7.344875389757782, gamma=59.593754547849436, score=0.8474721508140531, total=  15.5s
[CV] C=1.4463249633789066, gamma=59.214645069168306 ...


[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:   26.2s


[CV]  C=7.344875389757782, gamma=59.593754547849436, score=0.847770154373928, total=  15.8s
[CV] C=1.4463249633789066, gamma=59.214645069168306 ...


[Parallel(n_jobs=-1)]: Done   3 tasks      | elapsed:   51.6s


[CV]  C=1.4463249633789066, gamma=59.214645069168306, score=0.8474721508140531, total=  15.7s
[CV] C=1.4463249633789066, gamma=59.214645069168306 ...


[Parallel(n_jobs=-1)]: Done   4 tasks      | elapsed:   52.5s


[CV]  C=1.4463249633789066, gamma=59.214645069168306, score=0.8474721508140531, total=  15.4s
[CV] C=2.631757193426322, gamma=86.40714456170895 ...


[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:  1.3min
```

```
[CV]  C=1.4463249633789066, gamma=59.214645069168306, score=0.847770154373928, total=  15.4s
[CV] C=2.631757193426322, gamma=86.40714456170895 ...

[Parallel(n_jobs=-1)]: Done   6 tasks      | elapsed:  1.3min

[CV]  C=2.631757193426322, gamma=86.40714456170895, score=0.8474721508140531, total=  15.9s
[CV] C=2.631757193426322, gamma=86.40714456170895 ...

[Parallel(n_jobs=-1)]: Done   7 tasks      | elapsed:  1.7min

[CV]  C=2.631757193426322, gamma=86.40714456170895, score=0.8474721508140531, total=  15.8s
[CV] C=1.6889919460345937, gamma=41.288893020266364 ...

[Parallel(n_jobs=-1)]: Done   8 tasks      | elapsed:  1.7min

[CV]  C=2.631757193426322, gamma=86.40714456170895, score=0.847770154373928, total=  15.7s
[CV] C=1.6889919460345937, gamma=41.288893020266364 ...

[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:  2.2min

[CV]  C=1.6889919460345937, gamma=41.288893020266364, score=0.8474721508140531, total=  16.1s
[CV] C=1.6889919460345937, gamma=41.288893020266364 ...

[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:  2.2min

[CV]  C=1.6889919460345937, gamma=41.288893020266364, score=0.8474721508140531, total=  16.0s
[CV] C=2.964901730766185, gamma=77.48678037969415 ...

[Parallel(n_jobs=-1)]: Done  11 tasks      | elapsed:  2.6min

[CV]  C=1.6889919460345937, gamma=41.288893020266364, score=0.847770154373928, total=  16.1s
[CV] C=2.964901730766185, gamma=77.48678037969415 ...

[Parallel(n_jobs=-1)]: Done  12 tasks      | elapsed:  2.6min

[CV]  C=2.964901730766185, gamma=77.48678037969415, score=0.8474721508140531, total=  15.9s
[CV] C=2.964901730766185, gamma=77.48678037969415 ...
```

```
[Parallel(n_jobs=-1)]: Done  13 tasks      | elapsed:  3.0min


[CV]  C=2.964901730766185, gamma=77.48678037969415, score=0.8474721508140531, total=  15.7s
[CV] C=1.520133118164242, gamma=0.42656776863930757 ...


[Parallel(n_jobs=-1)]: Done  14 tasks      | elapsed:  3.1min


[CV]  C=1.520133118164242, gamma=0.42656776863930757, score=0.8924592973436161, total=   7.0s
[CV] C=1.520133118164242, gamma=0.42656776863930757 ...


[Parallel(n_jobs=-1)]: Done  15 tasks      | elapsed:  3.3min


[CV]  C=1.520133118164242, gamma=0.42656776863930757, score=0.8860325621251071, total=   6.7s
[CV] C=1.520133118164242, gamma=0.42656776863930757 ...


[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:  3.4min


[CV]  C=2.964901730766185, gamma=77.48678037969415, score=0.847770154373928, total=  15.7s
[CV] C=9.999600306870823, gamma=90.11659298311699 ...


[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:  3.5min


[CV]  C=1.520133118164242, gamma=0.42656776863930757, score=0.8876500857632933, total=   7.0s
[CV] C=9.999600306870823, gamma=90.11659298311699 ...


[Parallel(n_jobs=-1)]: Done  18 tasks      | elapsed:  3.6min


[CV]  C=9.999600306870823, gamma=90.11659298311699, score=0.8474721508140531, total=  15.6s
[CV] C=9.999600306870823, gamma=90.11659298311699 ...


[Parallel(n_jobs=-1)]: Done  19 tasks      | elapsed:  3.9min


[CV]  C=9.999600306870823, gamma=90.11659298311699, score=0.8474721508140531, total=  15.7s
[CV] C=0.9105204517019218, gamma=70.05308085085566 ...


[Parallel(n_jobs=-1)]: Done  20 tasks      | elapsed:  4.1min
```

```
[CV]  C=9.999600306870823, gamma=90.11659298311699, score=0.847770154373928, total=  15.6s
[CV] C=0.9105204517019218, gamma=70.05308085085566 ...


[Parallel(n_jobs=-1)]: Done  21 tasks      | elapsed:  4.3min


[CV]  C=0.9105204517019218, gamma=70.05308085085566, score=0.8474721508140531, total=  14.6s
[CV] C=0.9105204517019218, gamma=70.05308085085566 ...


[Parallel(n_jobs=-1)]: Done  22 tasks      | elapsed:  4.5min


[CV]  C=0.9105204517019218, gamma=70.05308085085566, score=0.8474721508140531, total=  14.4s
[CV] C=3.5333876179156607, gamma=50.005666866144104 ...


[Parallel(n_jobs=-1)]: Done  23 tasks      | elapsed:  4.7min


[CV]  C=0.9105204517019218, gamma=70.05308085085566, score=0.847770154373928, total=  14.6s
[CV] C=3.5333876179156607, gamma=50.005666866144104 ...


[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:  4.9min


[CV]  C=3.5333876179156607, gamma=50.005666866144104, score=0.8474721508140531, total=  16.2s
[CV] C=3.5333876179156607, gamma=50.005666866144104 ...


[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:  5.2min


[CV]  C=3.5333876179156607, gamma=50.005666866144104, score=0.8474721508140531, total=  15.8s
[CV] C=1.347251919682676, gamma=60.17423447209802 ...


[Parallel(n_jobs=-1)]: Done  26 tasks      | elapsed:  5.3min


[CV]  C=3.5333876179156607, gamma=50.005666866144104, score=0.847770154373928, total=  15.9s
[CV] C=1.347251919682676, gamma=60.17423447209802 ...


[Parallel(n_jobs=-1)]: Done  27 tasks      | elapsed:  5.6min
```

```
[CV]  C=1.347251919682676, gamma=60.17423447209802, score=0.8474721508140531, total=  15.6s
[CV] C=1.347251919682676, gamma=60.17423447209802 ...
[CV]  C=1.347251919682676, gamma=60.17423447209802, score=0.8474721508140531, total=  15.1s
[CV]  C=1.347251919682676, gamma=60.17423447209802, score=0.847770154373928, total=  15.4s


[Parallel(n_jobs=-1)]: Done  30 out of  30 | elapsed:  6.1min finished
```

```
Out[38]: RandomizedSearchCV(cv=None, error_score='raise',
                   estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
           decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
           max_iter=-1, probability=False, random_state=None, shrinking=True,
           tol=0.001, verbose=False),
                   fit_params=None, iid=True, n_iter=10, n_jobs=-1,
                   param_distributions={'C': <scipy.stats._distn_infrastructure.rv_frozen obje
                   pre_dispatch='2*n_jobs', random_state=None, refit=True,
                   return_train_score='warn', scoring=None, verbose=30)
```

```
In [39]: print("Best parameters: ", rbf_svm_grid_cv.best_params_)
         print("Best cross-validation score: {:.3f}".format(rbf_svm_grid_cv.best_score_))
```

```
Best parameters:  {'C': 1.520133118164242, 'gamma': 0.42656776863930757}
Best cross-validation score: 0.889
```

```
In [40]: y_pred = rbf_svm_grid_cv.predict(X_test1)
         print(classification_report(y_test, y_pred))
```

```
             precision    recall  f1-score   support

   negative       0.90      0.41      0.56       540
   positive       0.88      0.99      0.93      2460

avg / total       0.89      0.89      0.87      3000
```

---

4 TF-idf Vectorization

```
In [0]: #TF-IDF
        tf_idf_vect = TfidfVectorizer(ngram_range=(1,2))
        X_train2 = tf_idf_vect.fit_transform(X_train)
        X_test2 = tf_idf_vect.transform(X_test)
```

```
In [0]: #Normalize
        from sklearn.preprocessing import normalize
        X_train2 = normalize(X_train2)
        X_test2  = normalize(X_test2)
```

4.1 GridSearch Cross-Validation

```
In [10]: from sklearn.model_selection import GridSearchCV
         from sklearn.svm import SVC


         Cs = [0.001, 0.01, 0.1, 1, 10]
         gammas = [0.01, 0.1, 1, 10,100]
         param_grid = {'C': Cs, 'gamma' : gammas}
         rbf_svm = SVC(kernel='rbf')
         rbf_svm_grid_cv = GridSearchCV(rbf_svm, param_grid=param_grid, n_jobs = -1, verbose=3
         rbf_svm_grid_cv

         #perform gridsearch
         rbf_svm_grid_cv.fit(X_train2, y_train)
```

```
Fitting 3 folds for each of 25 candidates, totalling 75 fits
[CV] C=0.001, gamma=0.01 ...
[CV] C=0.001, gamma=0.01 ...
[CV] ... C=0.001, gamma=0.01, score=0.8534076296613802, total=   5.5s
[CV] C=0.001, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done   1 tasks      | elapsed:    9.3s


[CV] ... C=0.001, gamma=0.01, score=0.8534704370179949, total=   5.6s
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:    9.5s


[CV] ... C=0.001, gamma=0.01, score=0.8534076296613802, total=   5.4s
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   3 tasks      | elapsed:   18.4s


[CV] ... C=0.001, gamma=0.1, score=0.8534704370179949, total=   5.5s
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   4 tasks      | elapsed:   19.0s


[CV] ... C=0.001, gamma=0.1, score=0.8534076296613802, total=   5.5s
[CV] C=0.001, gamma=1 ...
```

```
[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:   27.7s


[CV] ... C=0.001, gamma=0.1, score=0.8534076296613802, total=   5.4s
[CV] C=0.001, gamma=1 ...


[Parallel(n_jobs=-1)]: Done   6 tasks      | elapsed:   28.1s


[CV] ... C=0.001, gamma=1, score=0.8534076296613802, total=   6.4s
[CV] C=0.001, gamma=1 ...
[CV] ... C=0.001, gamma=1, score=0.8534704370179949, total=   6.7s


[Parallel(n_jobs=-1)]: Done   7 tasks      | elapsed:   38.9s
[Parallel(n_jobs=-1)]: Done   8 tasks      | elapsed:   39.1s


[CV] C=0.001, gamma=10 ...
[CV] ... C=0.001, gamma=1, score=0.8534076296613802, total=   6.2s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:   49.4s


[CV] ... C=0.001, gamma=10, score=0.8534704370179949, total=   7.2s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:   51.2s


[CV] ... C=0.001, gamma=10, score=0.8534076296613802, total=   7.0s
[CV] C=0.001, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  11 tasks      | elapsed:  1.0min


[CV] ... C=0.001, gamma=10, score=0.8534076296613802, total=   7.0s
[CV] C=0.001, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  12 tasks      | elapsed:  1.0min


[CV] ... C=0.001, gamma=100, score=0.8534704370179949, total=   7.8s
[CV] C=0.001, gamma=100 ...
```

```
[Parallel(n_jobs=-1)]: Done  13 tasks      | elapsed:  1.2min


[CV] ... C=0.001, gamma=100, score=0.8534076296613802, total=   7.9s
[CV] C=0.01, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  14 tasks      | elapsed:  1.3min


[CV] ... C=0.01, gamma=0.01, score=0.8534704370179949, total=   5.4s
[CV] C=0.01, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  15 tasks      | elapsed:  1.4min


[CV] ... C=0.001, gamma=100, score=0.8534076296613802, total=   7.9s
[CV] C=0.01, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:  1.5min


[CV] ... C=0.01, gamma=0.01, score=0.8534076296613802, total=   5.2s
[CV] C=0.01, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:  1.6min


[CV] ... C=0.01, gamma=0.01, score=0.8534076296613802, total=   5.3s
[CV] C=0.01, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  18 tasks      | elapsed:  1.6min


[CV] ... C=0.01, gamma=0.1, score=0.8534704370179949, total=   8.1s
[CV] C=0.01, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  19 tasks      | elapsed:  1.8min


[CV] ... C=0.01, gamma=0.1, score=0.8534076296613802, total=   8.0s
[CV] C=0.01, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  20 tasks      | elapsed:  1.8min
```

```
[CV] ... C=0.01, gamma=0.1, score=0.8534076296613802, total=   7.9s
[CV] C=0.01, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  21 tasks      | elapsed:  2.0min

[CV] ... C=0.01, gamma=1, score=0.8534704370179949, total=  14.0s
[CV] C=0.01, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  22 tasks      | elapsed:  2.2min

[CV] ... C=0.01, gamma=1, score=0.8534076296613802, total=  14.0s
[CV] C=0.01, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  23 tasks      | elapsed:  2.4min

[CV] ... C=0.01, gamma=1, score=0.8534076296613802, total=  13.6s
[CV] C=0.01, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:  2.6min

[CV] ... C=0.01, gamma=10, score=0.8534704370179949, total=  16.8s
[CV] C=0.01, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:  2.9min

[CV] ... C=0.01, gamma=10, score=0.8534076296613802, total=  16.3s
[CV] C=0.01, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  26 tasks      | elapsed:  3.1min

[CV] ... C=0.01, gamma=10, score=0.8534076296613802, total=  16.3s
[CV] C=0.01, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  27 tasks      | elapsed:  3.4min

[CV] ... C=0.01, gamma=100, score=0.8534704370179949, total=  16.4s
[CV] C=0.01, gamma=100 ...
```

```
[Parallel(n_jobs=-1)]: Done  28 tasks      | elapsed:  3.5min


[CV] ... C=0.01, gamma=100, score=0.8534076296613802, total=  16.3s
[CV] C=0.1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  29 tasks      | elapsed:  3.8min


[CV] ... C=0.01, gamma=100, score=0.8534076296613802, total=  16.2s
[CV] C=0.1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  30 tasks      | elapsed:  4.0min


[CV] ... C=0.1, gamma=0.01, score=0.8534704370179949, total=   8.2s
[CV] C=0.1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  31 tasks      | elapsed:  4.1min


[CV] ... C=0.1, gamma=0.01, score=0.8534076296613802, total=   7.9s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  32 tasks      | elapsed:  4.2min


[CV] ... C=0.1, gamma=0.01, score=0.8534076296613802, total=   8.1s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  33 tasks      | elapsed:  4.3min


[CV] ... C=0.1, gamma=0.1, score=0.8534704370179949, total=  11.6s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  34 tasks      | elapsed:  4.6min


[CV] ... C=0.1, gamma=0.1, score=0.8534076296613802, total=  12.2s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  35 tasks      | elapsed:  4.6min
```

```
[CV] ... C=0.1, gamma=0.1, score=0.8534076296613802, total=  12.2s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  36 tasks      | elapsed:  4.9min

[CV] ... C=0.1, gamma=1, score=0.8534704370179949, total=  15.4s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  37 tasks      | elapsed:  5.0min

[CV] ... C=0.1, gamma=1, score=0.8534076296613802, total=  15.2s
[CV] C=0.1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  38 tasks      | elapsed:  5.3min

[CV] ... C=0.1, gamma=1, score=0.8534076296613802, total=  15.2s
[CV] C=0.1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  39 tasks      | elapsed:  5.5min

[CV] ... C=0.1, gamma=10, score=0.8534704370179949, total=  16.7s
[CV] C=0.1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  40 tasks      | elapsed:  5.8min

[CV] ... C=0.1, gamma=10, score=0.8534076296613802, total=  16.8s
[CV] C=0.1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  41 tasks      | elapsed:  5.9min

[CV] ... C=0.1, gamma=10, score=0.8534076296613802, total=  16.1s
[CV] C=0.1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed:  6.3min

[CV] ... C=0.1, gamma=100, score=0.8534704370179949, total=  16.8s
[CV] C=0.1, gamma=100 ...
```

```
[Parallel(n_jobs=-1)]: Done  43 tasks      | elapsed:  6.4min


[CV] ... C=0.1, gamma=100, score=0.8534076296613802, total=  16.0s
[CV] C=1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  44 tasks      | elapsed:  6.7min


[CV] ... C=0.1, gamma=100, score=0.8534076296613802, total=  16.4s
[CV] C=1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  45 tasks      | elapsed:  6.9min


[CV] ... C=1, gamma=0.01, score=0.8534704370179949, total=  11.1s
[CV] C=1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  46 tasks      | elapsed:  7.0min


[CV] ... C=1, gamma=0.01, score=0.8534076296613802, total=  11.3s
[CV] C=1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  47 tasks      | elapsed:  7.2min


[CV] ... C=1, gamma=0.01, score=0.8534076296613802, total=  11.2s
[CV] C=1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  48 tasks      | elapsed:  7.3min


[CV] ... C=1, gamma=0.1, score=0.8534704370179949, total=  13.6s
[CV] C=1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  49 tasks      | elapsed:  7.6min


[CV] ... C=1, gamma=0.1, score=0.8534076296613802, total=  13.0s
[CV] C=1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  50 tasks      | elapsed:  7.7min
```

```
[CV] ... C=1, gamma=0.1, score=0.8534076296613802, total=  13.1s
[CV] C=1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  51 tasks      | elapsed:  7.9min

[CV] ... C=1, gamma=1, score=0.8534704370179949, total=  15.5s
[CV] C=1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  52 tasks      | elapsed:  8.1min

[CV] ... C=1, gamma=1, score=0.8542648949849978, total=  15.6s
[CV] C=1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  53 tasks      | elapsed:  8.3min

[CV] ... C=1, gamma=1, score=0.853836262323189, total=  15.3s
[CV] C=1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  54 tasks      | elapsed:  8.5min

[CV] ... C=1, gamma=10, score=0.8534704370179949, total=  16.6s
[CV] C=1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  55 tasks      | elapsed:  8.8min

[CV] ... C=1, gamma=10, score=0.8534076296613802, total=  16.2s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  56 tasks      | elapsed:  9.0min

[CV] ... C=1, gamma=10, score=0.8534076296613802, total=  16.4s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  57 tasks      | elapsed:  9.3min

[CV] ... C=1, gamma=100, score=0.8534704370179949, total=  16.8s
[CV] C=1, gamma=100 ...
```

```
[Parallel(n_jobs=-1)]: Done  58 tasks      | elapsed:  9.5min


[CV] ... C=1, gamma=100, score=0.8534076296613802, total=  16.5s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  59 tasks      | elapsed:  9.8min


[CV] ... C=1, gamma=100, score=0.8534076296613802, total=  16.4s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  60 tasks      | elapsed:  9.9min


[CV] ... C=10, gamma=0.01, score=0.8534704370179949, total=  11.3s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  61 tasks      | elapsed: 10.1min


[CV] ... C=10, gamma=0.01, score=0.8534076296613802, total=  10.8s
[CV] C=10, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  62 tasks      | elapsed: 10.2min


[CV] ... C=10, gamma=0.01, score=0.8534076296613802, total=  11.1s
[CV] C=10, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  63 tasks      | elapsed: 10.4min


[CV] ... C=10, gamma=0.1, score=0.8864610111396743, total=  14.7s
[CV] C=10, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  64 tasks      | elapsed: 10.6min


[CV] ... C=10, gamma=0.1, score=0.8945563651950279, total=  14.5s
[CV] C=10, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  65 tasks      | elapsed: 10.8min
```

```
[CV] ... C=10, gamma=0.1, score=0.9001285897985426, total=  14.2s
[CV] C=10, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  66 tasks      | elapsed: 11.0min


[CV] ... C=10, gamma=1, score=0.856898029134533, total=  17.0s
[CV] C=10, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  67 tasks      | elapsed: 11.2min


[CV] ... C=10, gamma=1, score=0.8602657522503214, total=  16.4s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  68 tasks      | elapsed: 11.4min


[CV] ... C=10, gamma=1, score=0.8675525075010716, total=  16.6s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  69 tasks      | elapsed: 11.7min


[CV] ... C=10, gamma=10, score=0.8534704370179949, total=  17.9s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  70 tasks      | elapsed: 11.9min


[CV] ... C=10, gamma=10, score=0.8534076296613802, total=  17.7s
[CV] C=10, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  71 tasks      | elapsed: 12.2min


[CV] ... C=10, gamma=10, score=0.8534076296613802, total=  17.5s
[CV] C=10, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  72 tasks      | elapsed: 12.4min
```

```
[CV] ... C=10, gamma=100, score=0.8534704370179949, total=  17.6s
[CV] C=10, gamma=100 ...
[CV] ... C=10, gamma=100, score=0.8534076296613802, total=  17.5s
[CV] ... C=10, gamma=100, score=0.8534076296613802, total=  16.7s


[Parallel(n_jobs=-1)]: Done  75 out of  75 | elapsed: 13.1min remaining:    0.0s
[Parallel(n_jobs=-1)]: Done  75 out of  75 | elapsed: 13.1min finished


Out[10]: GridSearchCV(cv=None, error_score='raise',
              estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
            decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
            max_iter=-1, probability=False, random_state=None, shrinking=True,
            tol=0.001, verbose=False),
              fit_params=None, iid=True, n_jobs=-1,
              param_grid={'C': [0.001, 0.01, 0.1, 1, 10], 'gamma': [0.01, 0.1, 1, 10, 100]},
              pre_dispatch='2*n_jobs', refit=True, return_train_score='warn',
              scoring=None, verbose=30)

In [11]: print("Best parameters: ", rbf_svm_grid_cv.best_params_)
         print("Best cross-validation score: {:.3f}".format(rbf_svm_grid_cv.best_score_))

Best parameters:  {'C': 10, 'gamma': 0.1}
Best cross-validation score: 0.894


In [12]: from sklearn.metrics import classification_report

         y_pred = rbf_svm_grid_cv.predict(X_test2)
         print(classification_report(y_test, y_pred))

             precision    recall  f1-score   support

   negative       0.83      0.52      0.64       569
   positive       0.90      0.98      0.93      2431

avg / total       0.88      0.89      0.88      3000
```

4.2 RandomSearch Cross-Validation

```
In [13]: from sklearn.model_selection import RandomizedSearchCV
         from scipy import stats

         param_grid  = {"C": stats.uniform(0.001, 10), "gamma": stats.uniform(0.1, 100)}

         rbf_svm = SVC(kernel='rbf')
```

```
        rbf_svm_grid_cv =  RandomizedSearchCV(rbf_svm, param_grid, n_jobs=-1, verbose=30)
        rbf_svm_grid_cv

        #perform gridsearch
        rbf_svm_grid_cv.fit(X_train2, y_train)
```

```
Fitting 3 folds for each of 10 candidates, totalling 30 fits
[CV] C=3.9642771832596115, gamma=29.78300445438972 ...
[CV] C=3.9642771832596115, gamma=29.78300445438972 ...
[CV]  C=3.9642771832596115, gamma=29.78300445438972, score=0.8534076296613802, total=  17.9s
[CV] C=3.9642771832596115, gamma=29.78300445438972 ...


[Parallel(n_jobs=-1)]: Done   1 tasks      | elapsed:   29.9s


[CV]  C=3.9642771832596115, gamma=29.78300445438972, score=0.8534704370179949, total=  18.2s
[CV] C=1.3456356932942815, gamma=38.7452364654463 ...


[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:   30.6s


[CV]  C=3.9642771832596115, gamma=29.78300445438972, score=0.8534076296613802, total=  18.0s
[CV] C=1.3456356932942815, gamma=38.7452364654463 ...


[Parallel(n_jobs=-1)]: Done   3 tasks      | elapsed:   59.8s


[CV]  C=1.3456356932942815, gamma=38.7452364654463, score=0.8534704370179949, total=  17.7s
[CV] C=1.3456356932942815, gamma=38.7452364654463 ...


[Parallel(n_jobs=-1)]: Done   4 tasks      | elapsed:  1.0min


[CV]  C=1.3456356932942815, gamma=38.7452364654463, score=0.8534076296613802, total=  17.2s
[CV] C=1.6007164983489452, gamma=83.30034422776316 ...
[CV]  C=1.3456356932942815, gamma=38.7452364654463, score=0.8534076296613802, total=  16.8s
[CV] C=1.6007164983489452, gamma=83.30034422776316 ...


[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:  1.5min
[Parallel(n_jobs=-1)]: Done   6 tasks      | elapsed:  1.5min


[CV]  C=1.6007164983489452, gamma=83.30034422776316, score=0.8534076296613802, total=  17.3s
[CV] C=1.6007164983489452, gamma=83.30034422776316 ...
```

```
[Parallel(n_jobs=-1)]: Done   7 tasks      | elapsed:  2.0min


[CV]  C=1.6007164983489452, gamma=83.30034422776316, score=0.8534704370179949, total=  18.0s
[CV] C=3.661944348896329, gamma=96.2328762407625 ...


[Parallel(n_jobs=-1)]: Done   8 tasks      | elapsed:  2.0min


[CV]  C=1.6007164983489452, gamma=83.30034422776316, score=0.8534076296613802, total=  17.2s
[CV] C=3.661944348896329, gamma=96.2328762407625 ...


[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:  2.4min


[CV]  C=3.661944348896329, gamma=96.2328762407625, score=0.8534704370179949, total=  18.0s
[CV] C=3.661944348896329, gamma=96.2328762407625 ...


[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:  2.5min


[CV]  C=3.661944348896329, gamma=96.2328762407625, score=0.8534076296613802, total=  17.6s
[CV] C=9.435122773335705, gamma=24.682583638011348 ...


[Parallel(n_jobs=-1)]: Done  11 tasks      | elapsed:  2.9min


[CV]  C=3.661944348896329, gamma=96.2328762407625, score=0.8534076296613802, total=  17.3s
[CV] C=9.435122773335705, gamma=24.682583638011348 ...


[Parallel(n_jobs=-1)]: Done  12 tasks      | elapsed:  3.0min


[CV]  C=9.435122773335705, gamma=24.682583638011348, score=0.8534704370179949, total=  17.8s
[CV] C=9.435122773335705, gamma=24.682583638011348 ...


[Parallel(n_jobs=-1)]: Done  13 tasks      | elapsed:  3.4min


[CV]  C=9.435122773335705, gamma=24.682583638011348, score=0.8534076296613802, total=  17.2s
[CV] C=0.7132712411699417, gamma=72.174960251962 ...


[Parallel(n_jobs=-1)]: Done  14 tasks      | elapsed:  3.5min
```

```
[CV]  C=9.435122773335705, gamma=24.682583638011348, score=0.8534076296613802, total=  17.7s
[CV] C=0.7132712411699417, gamma=72.174960251962 ...

[Parallel(n_jobs=-1)]: Done  15 tasks      | elapsed:  3.9min

[CV]  C=0.7132712411699417, gamma=72.174960251962, score=0.8534704370179949, total=  16.8s
[CV] C=0.7132712411699417, gamma=72.174960251962 ...

[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:  3.9min

[CV]  C=0.7132712411699417, gamma=72.174960251962, score=0.8534076296613802, total=  16.0s
[CV] C=2.942296154149444, gamma=51.35304520571159 ...

[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:  4.4min

[CV]  C=0.7132712411699417, gamma=72.174960251962, score=0.8534076296613802, total=  16.2s
[CV] C=2.942296154149444, gamma=51.35304520571159 ...

[Parallel(n_jobs=-1)]: Done  18 tasks      | elapsed:  4.4min

[CV]  C=2.942296154149444, gamma=51.35304520571159, score=0.8534704370179949, total=  22.1s
[CV] C=2.942296154149444, gamma=51.35304520571159 ...

[Parallel(n_jobs=-1)]: Done  19 tasks      | elapsed:  5.0min

[CV]  C=2.942296154149444, gamma=51.35304520571159, score=0.8534076296613802, total=  22.2s
[CV] C=4.678180892232197, gamma=9.787874398952345 ...

[Parallel(n_jobs=-1)]: Done  20 tasks      | elapsed:  5.0min

[CV]  C=4.678180892232197, gamma=9.787874398952345, score=0.8534704370179949, total=  17.1s
[CV] C=4.678180892232197, gamma=9.787874398952345 ...

[Parallel(n_jobs=-1)]: Done  21 tasks      | elapsed:  5.5min

[CV]  C=2.942296154149444, gamma=51.35304520571159, score=0.8534076296613802, total=  22.6s
[CV] C=4.678180892232197, gamma=9.787874398952345 ...
```

```
[Parallel(n_jobs=-1)]: Done  22 tasks      | elapsed:  5.5min


[CV]  C=4.678180892232197, gamma=9.787874398952345, score=0.8534076296613802, total=  17.5s
[CV] C=5.615261329734922, gamma=66.67437251896835 ...


[Parallel(n_jobs=-1)]: Done  23 tasks      | elapsed:  5.9min


[CV]  C=4.678180892232197, gamma=9.787874398952345, score=0.8534076296613802, total=  17.3s
[CV] C=5.615261329734922, gamma=66.67437251896835 ...


[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:  6.0min


[CV]  C=5.615261329734922, gamma=66.67437251896835, score=0.8534704370179949, total=  17.6s
[CV] C=5.615261329734922, gamma=66.67437251896835 ...


[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:  6.4min


[CV]  C=5.615261329734922, gamma=66.67437251896835, score=0.8534076296613802, total=  17.0s
[CV] C=9.710920645923661, gamma=65.1088261351427 ...


[Parallel(n_jobs=-1)]: Done  26 tasks      | elapsed:  6.5min


[CV]  C=5.615261329734922, gamma=66.67437251896835, score=0.8534076296613802, total=  17.4s
[CV] C=9.710920645923661, gamma=65.1088261351427 ...


[Parallel(n_jobs=-1)]: Done  27 tasks      | elapsed:  6.9min


[CV]  C=9.710920645923661, gamma=65.1088261351427, score=0.8534704370179949, total=  17.4s
[CV] C=9.710920645923661, gamma=65.1088261351427 ...
[CV]  C=9.710920645923661, gamma=65.1088261351427, score=0.8534076296613802, total=  17.3s
[CV]  C=9.710920645923661, gamma=65.1088261351427, score=0.8534076296613802, total=  16.9s


[Parallel(n_jobs=-1)]: Done  30 out of  30 | elapsed:  7.4min finished


Out[13]: RandomizedSearchCV(cv=None, error_score='raise',
                   estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
           decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
```

```
            max_iter=-1, probability=False, random_state=None, shrinking=True,
            tol=0.001, verbose=False),
                    fit_params=None, iid=True, n_iter=10, n_jobs=-1,
                    param_distributions={'C': <scipy.stats._distn_infrastructure.rv_frozen obje
                    pre_dispatch='2*n_jobs', random_state=None, refit=True,
                    return_train_score='warn', scoring=None, verbose=30)
```

```
In [14]: print("Best parameters: ", rbf_svm_grid_cv.best_params_)
         print("Best cross-validation score: {:.3f}".format(rbf_svm_grid_cv.best_score_))
```

```
Best parameters:  {'C': 3.9642771832596115, 'gamma': 29.78300445438972}
Best cross-validation score: 0.853
```

```
In [15]: y_pred = rbf_svm_grid_cv.predict(X_test2)
         print(classification_report(y_test, y_pred))
```

```
             precision    recall  f1-score   support

   negative       0.00      0.00      0.00       569
   positive       0.81      1.00      0.90      2431

avg / total       0.66      0.81      0.73      3000
```

```
/usr/local/lib/python3.6/dist-packages/sklearn/metrics/classification.py:1135: UndefinedMetricW

Precision and F-score are ill-defined and being set to 0.0 in labels with no predicted samples
```

---

5 Word2Vec Vectorization

```
In [16]: import nltk
         nltk.download('stopwords')
         import re
         # Tutorial about Python regular expressions: https://pymotw.com/2/re/
         import string
         from nltk.corpus import stopwords
         from nltk.stem import PorterStemmer
         from nltk.stem.wordnet import WordNetLemmatizer

         stop = set(stopwords.words('english')) #set of stopwords
         sno = nltk.stem.SnowballStemmer('english') #initialising the snowball stemmer
```

```
        def cleanhtml(sentence): #function to clean the word of any html-tags
            cleanr = re.compile('<.*?>')
            cleantext = re.sub(cleanr, ' ', sentence)
            return cleantext
        def cleanpunc(sentence): #function to clean the word of any punctuation or special ch
            cleaned = re.sub(r'[?|!|\'|"|#]',r'',sentence)
            cleaned = re.sub(r'[.|,|)|(|\|/]',r' ',cleaned)
            return  cleaned
        print(stop)
        print('************************************')
        print(sno.stem('tasty'))
```

```
[nltk_data] Downloading package stopwords to /content/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.
{"it's", 'but', 'how', 'both', 'such', 'yourself', 'before', 'hasn', 'their', 'be', 'over', 'ju
**********************************
tasti
```

5.1 Training Word2Vec model using own text corpus

```
In [0]: import gensim
        i=0
        train_sent=[]
        for sent in X_train:
            filtered_sentence=[]
            sent=cleanhtml(sent)
            for w in sent.split():
                for cleaned_words in cleanpunc(w).split():
                    if(cleaned_words.isalpha()):
                        filtered_sentence.append(cleaned_words.lower())
                    else:
                        continue
            train_sent.append(filtered_sentence)
```

```
In [0]: test_sent=[]
        for sent in X_test:
            filtered_sentence=[]
            sent=cleanhtml(sent)
            for w in sent.split():
                for cleaned_words in cleanpunc(w).split():
                    if(cleaned_words.isalpha()):
                        filtered_sentence.append(cleaned_words.lower())
                    else:
                        continue
            test_sent.append(filtered_sentence)
```

```
In [0]: from gensim.models import Word2Vec
        from gensim.models import KeyedVectors

        w2v_model=gensim.models.Word2Vec(train_sent,min_count=5,size=50, workers=4)
```

5.2 Applying Average Word2vec

```
In [23]: #AVG-W2V
         sent_vectors = []; # the avg-w2v for each sentence/review is stored in this list
         for sent in train_sent: # for each review/sentence
             sent_vec = np.zeros(50) # as word vectors are of zero length
             cnt_words =0; # num of words with a valid vector in the sentence/review
             for word in sent: # for each word in a review/sentence
                 try:
                     vec = w2v_model.wv[word]
                     sent_vec += vec
                     cnt_words += 1
                 except:
                     pass
             sent_vec /= cnt_words
             sent_vectors.append(sent_vec)
         print(len(sent_vectors))
         print(len(sent_vectors[0]))
```

```
7000
50
```

```
In [24]: #AVG-W2V
         sent_vectors2 = []; # the avg-w2v for each sentence/review is stored in this list
         for sent in test_sent: # for each review/sentence
             sent_vec = np.zeros(50) # as word vectors are of zero length
             cnt_words =0; # num of words with a valid vector in the sentence/review
             for word in sent: # for each word in a review/sentence
                 try:
                     vec = w2v_model.wv[word]
                     sent_vec += vec
                     cnt_words += 1
                 except:
                     pass
             sent_vec /= cnt_words
             sent_vectors2.append(sent_vec)
         print(len(sent_vectors2))
         print(len(sent_vectors2[0]))
```

```
3000
50
```

```
In [0]: #Normalize
        from sklearn.preprocessing import normalize
        X_train3 = normalize(sent_vectors)
        X_test3 = normalize(sent_vectors2)
```

5.3 Applying GridSearchCV

```
In [26]: from sklearn.model_selection import GridSearchCV
         from sklearn.svm import SVC


         Cs = [0.001, 0.01, 0.1, 1, 10]
         gammas = [0.01, 0.1, 1, 10,100]
         param_grid = {'C': Cs, 'gamma' : gammas}
         rbf_svm = SVC(kernel='rbf')
         rbf_svm_grid_cv = GridSearchCV(rbf_svm, param_grid=param_grid, n_jobs=-1, verbose=30)
         rbf_svm_grid_cv

         #perform gridsearch
         rbf_svm_grid_cv.fit(X_train3, y_train)
```

```
Fitting 3 folds for each of 25 candidates, totalling 75 fits
[CV] C=0.001, gamma=0.01 ...
[CV] C=0.001, gamma=0.01 ...
[CV] ... C=0.001, gamma=0.01, score=0.8534076296613802, total=   0.8s
[CV] C=0.001, gamma=0.01 ...
[CV] ... C=0.001, gamma=0.01, score=0.8534704370179949, total=   0.9s
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   1 tasks      | elapsed:    1.3s
[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:    1.4s


[CV] ... C=0.001, gamma=0.01, score=0.8534076296613802, total=   1.1s
[CV] C=0.001, gamma=0.1 ...
[CV] ... C=0.001, gamma=0.1, score=0.8534704370179949, total=   1.2s
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   3 tasks      | elapsed:    3.3s
[Parallel(n_jobs=-1)]: Done   4 tasks      | elapsed:    3.4s


[CV] ... C=0.001, gamma=0.1, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=1 ...
[CV] ... C=0.001, gamma=0.1, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=1 ...
```

```
[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:    5.2s
[Parallel(n_jobs=-1)]: Done   6 tasks      | elapsed:    5.3s


[CV] ... C=0.001, gamma=1, score=0.8534704370179949, total=   1.2s
[CV] C=0.001, gamma=1 ...
[CV] ... C=0.001, gamma=1, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done   7 tasks      | elapsed:    7.2s
[Parallel(n_jobs=-1)]: Done   8 tasks      | elapsed:    7.3s


[CV] ... C=0.001, gamma=1, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:    9.1s


[CV] ... C=0.001, gamma=10, score=0.8534704370179949, total=   1.3s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:    9.3s


[CV] ... C=0.001, gamma=10, score=0.8534076296613802, total=   1.3s
[CV] C=0.001, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  11 tasks      | elapsed:   11.1s


[CV] ... C=0.001, gamma=10, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  12 tasks      | elapsed:   11.4s


[CV] ... C=0.001, gamma=100, score=0.8534704370179949, total=   1.3s
[CV] C=0.001, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  13 tasks      | elapsed:   13.2s
```

```
[CV] ... C=0.001, gamma=100, score=0.8534076296613802, total=   1.3s
[CV] C=0.01, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  14 tasks      | elapsed:   13.5s

[CV] ... C=0.001, gamma=100, score=0.8534076296613802, total=   1.3s
[CV] C=0.01, gamma=0.01 ...
[CV] ... C=0.01, gamma=0.01, score=0.8534704370179949, total=   1.2s
[CV] C=0.01, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  15 tasks      | elapsed:   15.4s
[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:   15.4s

[CV] ... C=0.01, gamma=0.01, score=0.8534076296613802, total=   1.2s
[CV] C=0.01, gamma=0.1 ...
[CV] ... C=0.01, gamma=0.01, score=0.8534076296613802, total=   1.2s
[CV] C=0.01, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:   17.3s
[Parallel(n_jobs=-1)]: Done  18 tasks      | elapsed:   17.3s

[CV] ... C=0.01, gamma=0.1, score=0.8534704370179949, total=   1.2s
[CV] ... C=0.01, gamma=0.1, score=0.8534076296613802, total=   1.2s
[CV] C=0.01, gamma=0.1 ...
[CV] C=0.01, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  19 tasks      | elapsed:   19.3s
[Parallel(n_jobs=-1)]: Done  20 tasks      | elapsed:   19.3s

[CV] ... C=0.01, gamma=0.1, score=0.8534076296613802, total=   1.2s
[CV] C=0.01, gamma=1 ...
[CV] ... C=0.01, gamma=1, score=0.8534704370179949, total=   1.3s
[CV] C=0.01, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  21 tasks      | elapsed:   21.3s
[Parallel(n_jobs=-1)]: Done  22 tasks      | elapsed:   21.4s

[CV] ... C=0.01, gamma=1, score=0.8534076296613802, total=   1.3s
[CV] C=0.01, gamma=10 ...
[CV] ... C=0.01, gamma=1, score=0.8534076296613802, total=   1.3s
[CV] C=0.01, gamma=10 ...
```

```
[Parallel(n_jobs=-1)]: Done  23 tasks      | elapsed:   23.4s
[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:   23.4s


[CV] ... C=0.01, gamma=10, score=0.8534704370179949, total=   1.6s
[CV] C=0.01, gamma=10 ...
[CV] ... C=0.01, gamma=10, score=0.8534076296613802, total=   1.6s
[CV] C=0.01, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   25.9s
[Parallel(n_jobs=-1)]: Done  26 tasks      | elapsed:   25.9s


[CV] ... C=0.01, gamma=10, score=0.8534076296613802, total=   1.6s
[CV] C=0.01, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  27 tasks      | elapsed:   28.3s


[CV] ... C=0.01, gamma=100, score=0.8534704370179949, total=   2.4s
[CV] C=0.01, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  28 tasks      | elapsed:   29.8s


[CV] ... C=0.01, gamma=100, score=0.8534076296613802, total=   2.4s
[CV] C=0.1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  29 tasks      | elapsed:   32.1s


[CV] ... C=0.01, gamma=100, score=0.8534076296613802, total=   2.4s
[CV] C=0.1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  30 tasks      | elapsed:   33.5s


[CV] ... C=0.1, gamma=0.01, score=0.8534704370179949, total=   1.1s
[CV] C=0.1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  31 tasks      | elapsed:   34.0s
```

```
[CV] ... C=0.1, gamma=0.01, score=0.8534076296613802, total=   1.1s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  32 tasks      | elapsed:   35.4s

[CV] ... C=0.1, gamma=0.01, score=0.8534076296613802, total=   1.2s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  33 tasks      | elapsed:   36.0s

[CV] ... C=0.1, gamma=0.1, score=0.8534704370179949, total=   1.3s
[CV] C=0.1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  34 tasks      | elapsed:   37.5s

[CV] ... C=0.1, gamma=0.1, score=0.8534076296613802, total=   1.3s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  35 tasks      | elapsed:   38.0s

[CV] ... C=0.1, gamma=0.1, score=0.8534076296613802, total=   1.3s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  36 tasks      | elapsed:   39.5s

[CV] ... C=0.1, gamma=1, score=0.8534704370179949, total=   1.6s
[CV] C=0.1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  37 tasks      | elapsed:   40.4s

[CV] ... C=0.1, gamma=1, score=0.8534076296613802, total=   1.5s
[CV] C=0.1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  38 tasks      | elapsed:   41.9s

[CV] ... C=0.1, gamma=1, score=0.8534076296613802, total=   1.5s
[CV] C=0.1, gamma=10 ...
```

```
[Parallel(n_jobs=-1)]: Done  39 tasks      | elapsed:   42.8s


[CV] ... C=0.1, gamma=10, score=0.8534704370179949, total=   2.0s
[CV] C=0.1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  40 tasks      | elapsed:   44.8s


[CV] ... C=0.1, gamma=10, score=0.8534076296613802, total=   2.0s
[CV] C=0.1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  41 tasks      | elapsed:   45.7s


[CV] ... C=0.1, gamma=10, score=0.8534076296613802, total=   2.0s
[CV] C=0.1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed:   47.7s


[CV] ... C=0.1, gamma=100, score=0.8534704370179949, total=   2.9s
[CV] C=0.1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  43 tasks      | elapsed:   50.2s


[CV] ... C=0.1, gamma=100, score=0.8534076296613802, total=   2.9s
[CV] C=1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  44 tasks      | elapsed:   52.1s


[CV] ... C=1, gamma=0.01, score=0.8534704370179949, total=   1.3s
[CV] C=1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  45 tasks      | elapsed:   54.1s


[CV] ... C=0.1, gamma=100, score=0.8534076296613802, total=   2.9s
[CV] C=1, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  46 tasks      | elapsed:   54.8s
```

```
[CV] ... C=1, gamma=0.01, score=0.8534076296613802, total=   1.3s
[CV] C=1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  47 tasks      | elapsed:   56.2s

[CV] ... C=1, gamma=0.01, score=0.8534076296613802, total=   1.3s
[CV] C=1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  48 tasks      | elapsed:   56.9s

[CV] ... C=1, gamma=0.1, score=0.8534704370179949, total=   1.5s
[CV] C=1, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  49 tasks      | elapsed:   58.6s

[CV] ... C=1, gamma=0.1, score=0.8534076296613802, total=   1.5s
[CV] C=1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  50 tasks      | elapsed:   59.2s

[CV] ... C=1, gamma=0.1, score=0.8534076296613802, total=   1.5s
[CV] C=1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  51 tasks      | elapsed:  1.0min

[CV] ... C=1, gamma=1, score=0.8534704370179949, total=   2.0s
[CV] C=1, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  52 tasks      | elapsed:  1.0min

[CV] ... C=1, gamma=1, score=0.8534076296613802, total=   2.0s
[CV] C=1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  53 tasks      | elapsed:  1.1min

[CV] ... C=1, gamma=1, score=0.8534076296613802, total=   2.1s
[CV] C=1, gamma=10 ...
```

```
[Parallel(n_jobs=-1)]: Done  54 tasks      | elapsed:  1.1min


[CV] ... C=1, gamma=10, score=0.8534704370179949, total=   2.0s
[CV] C=1, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  55 tasks      | elapsed:  1.1min


[CV] ... C=1, gamma=10, score=0.8534076296613802, total=   2.0s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  56 tasks      | elapsed:  1.1min


[CV] ... C=1, gamma=10, score=0.8534076296613802, total=   2.0s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  57 tasks      | elapsed:  1.2min


[CV] ... C=1, gamma=100, score=0.8538988860325621, total=   2.9s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  58 tasks      | elapsed:  1.2min


[CV] ... C=1, gamma=100, score=0.8534076296613802, total=   2.8s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  59 tasks      | elapsed:  1.2min


[CV] ... C=10, gamma=0.01, score=0.8534704370179949, total=   1.5s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  60 tasks      | elapsed:  1.3min


[CV] ... C=1, gamma=100, score=0.853836262323189, total=   3.0s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  61 tasks      | elapsed:  1.3min
```

```
[CV] ... C=10, gamma=0.01, score=0.8534076296613802, total=   1.5s
[CV] C=10, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  62 tasks      | elapsed:  1.3min

[CV] ... C=10, gamma=0.01, score=0.8534076296613802, total=   1.5s
[CV] C=10, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  63 tasks      | elapsed:  1.3min

[CV] ... C=10, gamma=0.1, score=0.8534704370179949, total=   2.0s
[CV] C=10, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  64 tasks      | elapsed:  1.4min

[CV] ... C=10, gamma=0.1, score=0.8534076296613802, total=   1.9s
[CV] C=10, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  65 tasks      | elapsed:  1.4min

[CV] ... C=10, gamma=0.1, score=0.8534076296613802, total=   2.0s
[CV] C=10, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  66 tasks      | elapsed:  1.4min

[CV] ... C=10, gamma=1, score=0.8534704370179949, total=   2.0s
[CV] C=10, gamma=1 ...

[Parallel(n_jobs=-1)]: Done  67 tasks      | elapsed:  1.4min

[CV] ... C=10, gamma=1, score=0.8534076296613802, total=   2.3s
[CV] C=10, gamma=10 ...

[Parallel(n_jobs=-1)]: Done  68 tasks      | elapsed:  1.5min

[CV] ... C=10, gamma=1, score=0.8534076296613802, total=   2.3s
[CV] C=10, gamma=10 ...
```

```
[Parallel(n_jobs=-1)]: Done  69 tasks      | elapsed:  1.5min


[CV] ... C=10, gamma=10, score=0.8556126820908312, total=   2.0s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  70 tasks      | elapsed:  1.5min


[CV] ... C=10, gamma=10, score=0.8581225889412774, total=   2.0s
[CV] C=10, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  71 tasks      | elapsed:  1.5min


[CV] ... C=10, gamma=10, score=0.853836262323189, total=   2.1s
[CV] C=10, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  72 tasks      | elapsed:  1.6min


[CV] ... C=10, gamma=100, score=0.8273350471293917, total=   3.4s
[CV] C=10, gamma=100 ...
[CV] ... C=10, gamma=100, score=0.8251178739819974, total=   3.4s
[CV] ... C=10, gamma=100, score=0.8251178739819974, total=   3.1s


[Parallel(n_jobs=-1)]: Done  75 out of  75 | elapsed:  1.7min remaining:    0.0s
[Parallel(n_jobs=-1)]: Done  75 out of  75 | elapsed:  1.7min finished


Out[26]: GridSearchCV(cv=None, error_score='raise',
              estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
            decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
            max_iter=-1, probability=False, random_state=None, shrinking=True,
            tol=0.001, verbose=False),
              fit_params=None, iid=True, n_jobs=-1,
              param_grid={'C': [0.001, 0.01, 0.1, 1, 10], 'gamma': [0.01, 0.1, 1, 10, 100]},
              pre_dispatch='2*n_jobs', refit=True, return_train_score='warn',
              scoring=None, verbose=30)

In [27]: print("Best parameters: ", rbf_svm_grid_cv.best_params_)
         print("Best cross-validation score: {:.3f}".format(rbf_svm_grid_cv.best_score_))

Best parameters:  {'C': 10, 'gamma': 10}
Best cross-validation score: 0.856
```

```
In [28]: from sklearn.metrics import classification_report

         y_pred = rbf_svm_grid_cv.predict(X_test3)
         print(classification_report(y_test, y_pred))
```

```
             precision    recall  f1-score   support

   negative       0.60      0.04      0.08       569
   positive       0.82      0.99      0.90      2431

avg / total       0.77      0.81      0.74      3000
```

5.4 RandomSearch Cross-Validation

```
In [29]: from sklearn.model_selection import RandomizedSearchCV
         from scipy import stats

         param_grid  = {"C": stats.uniform(0.001, 10), "gamma": stats.uniform(0.1, 100)}

         rbf_svm = SVC(kernel='rbf')
         rbf_svm_grid_cv =  RandomizedSearchCV(rbf_svm, param_grid, n_jobs=-1, verbose=30)
         rbf_svm_grid_cv

         #perform gridsearch
         rbf_svm_grid_cv.fit(X_train3, y_train)
```

```
Fitting 3 folds for each of 10 candidates, totalling 30 fits
[CV] C=4.831589679522305, gamma=17.36777007238206 ...
[CV] C=4.831589679522305, gamma=17.36777007238206 ...
[CV]  C=4.831589679522305, gamma=17.36777007238206, score=0.8572653236176596, total=   1.7s
[CV] C=4.831589679522305, gamma=17.36777007238206 ...
[CV]  C=4.831589679522305, gamma=17.36777007238206, score=0.8551842330762639, total=   1.8s
[CV] C=3.555084986226801, gamma=57.56056193440211 ...


[Parallel(n_jobs=-1)]: Done   1 tasks      | elapsed:    2.7s
[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:    2.8s


[CV]  C=4.831589679522305, gamma=17.36777007238206, score=0.8534076296613802, total=   2.4s
[CV] C=3.555084986226801, gamma=57.56056193440211 ...


[Parallel(n_jobs=-1)]: Done   3 tasks      | elapsed:    6.1s
```

```
[CV]  C=3.555084986226801, gamma=57.56056193440211, score=0.8526135389888603, total=   2.8s
[CV] C=3.555084986226801, gamma=57.56056193440211 ...

[Parallel(n_jobs=-1)]: Done   4 tasks      | elapsed:    7.0s

[CV]  C=3.555084986226801, gamma=57.56056193440211, score=0.8474067723960566, total=   2.7s
[CV] C=7.25782600865332, gamma=70.9056651024806 ...

[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:   10.0s

[CV]  C=3.555084986226801, gamma=57.56056193440211, score=0.8444063437633947, total=   2.6s
[CV] C=7.25782600865332, gamma=70.9056651024806 ...

[Parallel(n_jobs=-1)]: Done   6 tasks      | elapsed:   10.9s

[CV]  C=7.25782600865332, gamma=70.9056651024806, score=0.8363324764353042, total=   2.8s
[CV] C=7.25782600865332, gamma=70.9056651024806 ...

[Parallel(n_jobs=-1)]: Done   7 tasks      | elapsed:   14.1s

[CV]  C=7.25782600865332, gamma=70.9056651024806, score=0.8289755679382769, total=   2.8s
[CV] C=7.825453560410199, gamma=82.15317055737387 ...

[Parallel(n_jobs=-1)]: Done   8 tasks      | elapsed:   15.0s

[CV]  C=7.25782600865332, gamma=70.9056651024806, score=0.831118731247321, total=   2.8s
[CV] C=7.825453560410199, gamma=82.15317055737387 ...

[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:   18.2s

[CV]  C=7.825453560410199, gamma=82.15317055737387, score=0.8341902313624678, total=   2.8s
[CV] C=7.825453560410199, gamma=82.15317055737387 ...

[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:   19.3s

[CV]  C=7.825453560410199, gamma=82.15317055737387, score=0.8272610372910416, total=   2.8s
[CV] C=0.6299281378974121, gamma=90.18127355653424 ...
```

```
[Parallel(n_jobs=-1)]: Done  11 tasks      | elapsed:   22.4s


[CV]  C=7.825453560410199, gamma=82.15317055737387, score=0.8289755679382769, total=   2.9s
[CV] C=0.6299281378974121, gamma=90.18127355653424 ...


[Parallel(n_jobs=-1)]: Done  12 tasks      | elapsed:   23.6s


[CV]  C=0.6299281378974121, gamma=90.18127355653424, score=0.8534704370179949, total=   2.8s
[CV] C=0.6299281378974121, gamma=90.18127355653424 ...


[Parallel(n_jobs=-1)]: Done  13 tasks      | elapsed:   26.7s


[CV]  C=0.6299281378974121, gamma=90.18127355653424, score=0.8525503643377625, total=   2.8s
[CV] C=8.778662125624574, gamma=32.38467885995086 ...


[Parallel(n_jobs=-1)]: Done  14 tasks      | elapsed:   27.9s


[CV]  C=0.6299281378974121, gamma=90.18127355653424, score=0.8534076296613802, total=   2.9s
[CV] C=8.778662125624574, gamma=32.38467885995086 ...
[CV]  C=8.778662125624574, gamma=32.38467885995086, score=0.8444730077120822, total=   2.3s
[CV] C=8.778662125624574, gamma=32.38467885995086 ...


[Parallel(n_jobs=-1)]: Done  15 tasks      | elapsed:   31.1s
[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:   31.2s


[CV]  C=8.778662125624574, gamma=32.38467885995086, score=0.8461208744106301, total=   2.4s
[CV] C=1.3538159546608397, gamma=99.63626932317803 ...
[CV]  C=8.778662125624574, gamma=32.38467885995086, score=0.8426918131161595, total=   2.4s
[CV] C=1.3538159546608397, gamma=99.63626932317803 ...


[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:   34.6s
[Parallel(n_jobs=-1)]: Done  18 tasks      | elapsed:   34.7s


[CV]  C=1.3538159546608397, gamma=99.63626932317803, score=0.8564695801199658, total=   3.0s
[CV] C=1.3538159546608397, gamma=99.63626932317803 ...
[CV]  C=1.3538159546608397, gamma=99.63626932317803, score=0.853836262323189, total=   3.0s
[CV] C=7.729950108997679, gamma=7.50559153289041 ...
```

```
[Parallel(n_jobs=-1)]: Done  19 tasks      | elapsed:   39.2s
[Parallel(n_jobs=-1)]: Done  20 tasks      | elapsed:   39.3s


[CV]  C=7.729950108997679, gamma=7.50559153289041, score=0.8543273350471294, total=   2.0s
[CV] C=7.729950108997679, gamma=7.50559153289041 ...


[Parallel(n_jobs=-1)]: Done  21 tasks      | elapsed:   42.1s


[CV]  C=1.3538159546608397, gamma=99.63626932317803, score=0.8516930990141449, total=   3.1s
[CV] C=7.729950108997679, gamma=7.50559153289041 ...


[Parallel(n_jobs=-1)]: Done  22 tasks      | elapsed:   43.9s


[CV]  C=7.729950108997679, gamma=7.50559153289041, score=0.853836262323189, total=   2.0s
[CV] C=6.8506275628906215, gamma=38.017778693382155 ...


[Parallel(n_jobs=-1)]: Done  23 tasks      | elapsed:   45.0s


[CV]  C=7.729950108997679, gamma=7.50559153289041, score=0.8534076296613802, total=   2.0s
[CV] C=6.8506275628906215, gamma=38.017778693382155 ...


[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:   46.7s


[CV]  C=6.8506275628906215, gamma=38.017778693382155, score=0.8479005998286204, total=   2.3s
[CV] C=6.8506275628906215, gamma=38.017778693382155 ...


[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   48.4s


[CV]  C=6.8506275628906215, gamma=38.017778693382155, score=0.8439777111015859, total=   2.3s
[CV] C=2.8135405635998967, gamma=83.68330581001806 ...


[Parallel(n_jobs=-1)]: Done  26 tasks      | elapsed:   50.1s


[CV]  C=6.8506275628906215, gamma=38.017778693382155, score=0.8439777111015859, total=   2.4s
[CV] C=2.8135405635998967, gamma=83.68330581001806 ...
```

```
[Parallel(n_jobs=-1)]: Done  27 tasks      | elapsed:   51.9s


[CV]  C=2.8135405635998967, gamma=83.68330581001806, score=0.8496143958868895, total=   2.8s
[CV] C=2.8135405635998967, gamma=83.68330581001806 ...
[CV]  C=2.8135405635998967, gamma=83.68330581001806, score=0.8431204457779683, total=   2.7s
[CV]  C=2.8135405635998967, gamma=83.68330581001806, score=0.8474067723960566, total=   2.4s


[Parallel(n_jobs=-1)]: Done  30 out of  30 | elapsed:   57.7s finished


Out[29]: RandomizedSearchCV(cv=None, error_score='raise',
                    estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
           decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
           max_iter=-1, probability=False, random_state=None, shrinking=True,
           tol=0.001, verbose=False),
                    fit_params=None, iid=True, n_iter=10, n_jobs=-1,
                    param_distributions={'C': <scipy.stats._distn_infrastructure.rv_frozen obje
                    pre_dispatch='2*n_jobs', random_state=None, refit=True,
                    return_train_score='warn', scoring=None, verbose=30)

In [30]: print("Best parameters: ", rbf_svm_grid_cv.best_params_)
         print("Best cross-validation score: {:.3f}".format(rbf_svm_grid_cv.best_score_))

Best parameters:  {'C': 4.831589679522305, 'gamma': 17.36777007238206}
Best cross-validation score: 0.855


In [31]: y_pred = rbf_svm_grid_cv.predict(X_test3)
         print(classification_report(y_test, y_pred))

             precision    recall  f1-score   support

   negative       0.58      0.04      0.07       569
   positive       0.82      0.99      0.90      2431

avg / total       0.77      0.81      0.74      3000
```

---

6 Word2Vec Tf-idf Vectorization

```
In [0]: #TF-IDF

        tf_idf_vect = TfidfVectorizer()
        final_tf_idf = tf_idf_vect.fit_transform(X_train)
```

```
tfidf_feat = tf_idf_vect.get_feature_names()
 # tfidf words/col-names
# final_tf_idf is the sparse matrix with row= sentence, col=word and cell_val = tfidf

train_vectors = []; # the tfidf-w2v for each sentence/review is stored in this list
row=0;
for sent in train_sent: # for each review/sentence
    sent_vec = np.zeros(50) # as word vectors are of zero length
    weight_sum = 0; # num of words with a valid vector in the sentence/review
    for word in sent: # for each word in a review/sentence
        try:
            vec = w2v_model.wv[word]
            # obtain the tf_idf of a word in a sentence/review
            tfidf = final_tf_idf[row, tfidf_feat.index(word)]
            sent_vec += (vec * tfidf)
            weight_sum += tfidf


        except:
            pass

    sent_vec /= weight_sum
    #print(np.isnan(np.sum(sent_vec)))

    train_vectors.append(sent_vec)
    row += 1
```

In [0]:
```
#TF-IDF

tf_idf_vect = TfidfVectorizer()
final_tf_idf = tf_idf_vect.fit_transform(X_test)
tfidf_feat = tf_idf_vect.get_feature_names()
 # tfidf words/col-names
# final_tf_idf is the sparse matrix with row= sentence, col=word and cell_val = tfidf

test_vectors = []; # the tfidf-w2v for each sentence/review is stored in this list
row=0;
for sent in test_sent: # for each review/sentence
    sent_vec = np.zeros(50) # as word vectors are of zero length
    weight_sum = 0; # num of words with a valid vector in the sentence/review
    for word in sent: # for each word in a review/sentence
        try:
            vec = w2v_model.wv[word]
            # obtain the tf_idf of a word in a sentence/review
            tfidf = final_tf_idf[row, tfidf_feat.index(word)]
            sent_vec += (vec * tfidf)
            weight_sum += tfidf
```

```
            except:
                pass

        sent_vec /= weight_sum
        #print(np.isnan(np.sum(sent_vec)))

        test_vectors.append(sent_vec)
        row += 1
```

```
In [0]: #Normalize
        from sklearn.preprocessing import normalize
        X_train4 = normalize(train_vectors)
        X_test4 = normalize(test_vectors)
```

6.1 Applying GridSearchCV

```
In [37]: from sklearn.model_selection import GridSearchCV
         from sklearn.svm import SVC


         Cs = [0.001, 0.01, 0.1, 1, 10]
         gammas = [0.01, 0.1, 1, 10,100]
         param_grid = {'C': Cs, 'gamma' : gammas}
         rbf_svm = SVC(kernel='rbf')
         rbf_svm_grid_cv = GridSearchCV(rbf_svm, param_grid=param_grid, n_jobs=-1, verbose=30)
         rbf_svm_grid_cv

         #perform gridsearch
         rbf_svm_grid_cv.fit(X_train4, y_train)
```

```
Fitting 3 folds for each of 25 candidates, totalling 75 fits
[CV] C=0.001, gamma=0.01 ...
[CV] C=0.001, gamma=0.01 ...
[CV] ... C=0.001, gamma=0.01, score=0.8534704370179949, total=   0.9s
[CV] ... C=0.001, gamma=0.01, score=0.8534076296613802, total=   0.9s
[CV] C=0.001, gamma=0.01 ...
[CV] C=0.001, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done   1 tasks      | elapsed:    1.4s
[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:    1.4s


[CV] ... C=0.001, gamma=0.01, score=0.8534076296613802, total=   1.0s
[CV] ... C=0.001, gamma=0.1, score=0.8534704370179949, total=   1.0s
[CV] C=0.001, gamma=0.1 ...
[CV] C=0.001, gamma=0.1 ...
```

```
[Parallel(n_jobs=-1)]: Done   3 tasks      | elapsed:    3.1s
[Parallel(n_jobs=-1)]: Done   4 tasks      | elapsed:    3.1s


[CV] ... C=0.001, gamma=0.1, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=1 ...
[CV] ... C=0.001, gamma=0.1, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=1 ...


[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:    5.1s
[Parallel(n_jobs=-1)]: Done   6 tasks      | elapsed:    5.1s


[CV] ... C=0.001, gamma=1, score=0.8534704370179949, total=   1.2s
[CV] C=0.001, gamma=1 ...
[CV] ... C=0.001, gamma=1, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done   7 tasks      | elapsed:    7.0s
[Parallel(n_jobs=-1)]: Done   8 tasks      | elapsed:    7.1s


[CV] ... C=0.001, gamma=1, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=10 ...
[CV] ... C=0.001, gamma=10, score=0.8534704370179949, total=   1.2s
[CV] C=0.001, gamma=10 ...


[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:    9.0s
[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:    9.1s


[CV] ... C=0.001, gamma=10, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=100 ...
[CV] ... C=0.001, gamma=10, score=0.8534076296613802, total=   1.2s
[CV] C=0.001, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  11 tasks      | elapsed:   11.0s
[Parallel(n_jobs=-1)]: Done  12 tasks      | elapsed:   11.1s


[CV] ... C=0.001, gamma=100, score=0.8534704370179949, total=   1.3s
[CV] C=0.001, gamma=100 ...
[CV] ... C=0.001, gamma=100, score=0.8534076296613802, total=   1.3s
[CV] C=0.01, gamma=0.01 ...
```

```
[Parallel(n_jobs=-1)]: Done  13 tasks      | elapsed:   13.1s
[Parallel(n_jobs=-1)]: Done  14 tasks      | elapsed:   13.2s


[CV] ... C=0.01, gamma=0.01, score=0.8534704370179949, total=   1.1s
[CV] C=0.01, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  15 tasks      | elapsed:   15.0s


[CV] ... C=0.001, gamma=100, score=0.8534076296613802, total=   1.3s
[CV] C=0.01, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:   15.3s


[CV] ... C=0.01, gamma=0.01, score=0.8534076296613802, total=   1.2s
[CV] C=0.01, gamma=0.1 ...
[CV] ... C=0.01, gamma=0.01, score=0.8534076296613802, total=   1.2s
[CV] C=0.01, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:   17.0s
[Parallel(n_jobs=-1)]: Done  18 tasks      | elapsed:   17.2s


[CV] ... C=0.01, gamma=0.1, score=0.8534704370179949, total=   1.2s
[CV] C=0.01, gamma=0.1 ...
[CV] ... C=0.01, gamma=0.1, score=0.8534076296613802, total=   1.2s
[CV] C=0.01, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  19 tasks      | elapsed:   18.9s
[Parallel(n_jobs=-1)]: Done  20 tasks      | elapsed:   19.1s


[CV] ... C=0.01, gamma=0.1, score=0.8534076296613802, total=   1.2s
[CV] C=0.01, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  21 tasks      | elapsed:   20.9s


[CV] ... C=0.01, gamma=1, score=0.8534704370179949, total=   1.3s
[CV] C=0.01, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  22 tasks      | elapsed:   21.2s
```

```
[CV] ... C=0.01, gamma=1, score=0.8534076296613802, total=   1.3s
[CV] C=0.01, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  23 tasks      | elapsed:   22.9s


[CV] ... C=0.01, gamma=1, score=0.8534076296613802, total=   1.3s
[CV] C=0.01, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:   23.2s


[CV] ... C=0.01, gamma=10, score=0.8534704370179949, total=   1.6s
[CV] C=0.01, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   25.5s


[CV] ... C=0.01, gamma=10, score=0.8534076296613802, total=   1.7s
[CV] C=0.01, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  26 tasks      | elapsed:   25.8s


[CV] ... C=0.01, gamma=10, score=0.8534076296613802, total=   1.6s
[CV] C=0.01, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  27 tasks      | elapsed:   27.9s


[CV] ... C=0.01, gamma=100, score=0.8534704370179949, total=   2.5s
[CV] C=0.01, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  28 tasks      | elapsed:   29.7s


[CV] ... C=0.01, gamma=100, score=0.8534076296613802, total=   2.5s
[CV] C=0.1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  29 tasks      | elapsed:   31.8s
```

```
[CV] ... C=0.01, gamma=100, score=0.8534076296613802, total=   2.5s
[CV] C=0.1, gamma=0.01 ...
[CV] ... C=0.1, gamma=0.01, score=0.8534704370179949, total=   1.2s
[CV] C=0.1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  30 tasks      | elapsed:   33.6s
[Parallel(n_jobs=-1)]: Done  31 tasks      | elapsed:   33.7s


[CV] ... C=0.1, gamma=0.01, score=0.8534076296613802, total=   1.2s
[CV] C=0.1, gamma=0.1 ...
[CV] ... C=0.1, gamma=0.01, score=0.8534076296613802, total=   1.2s
[CV] C=0.1, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  32 tasks      | elapsed:   35.6s
[Parallel(n_jobs=-1)]: Done  33 tasks      | elapsed:   35.6s


[CV] ... C=0.1, gamma=0.1, score=0.8534704370179949, total=   1.3s
[CV] C=0.1, gamma=0.1 ...
[CV] ... C=0.1, gamma=0.1, score=0.8534076296613802, total=   1.3s
[CV] C=0.1, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  34 tasks      | elapsed:   37.6s
[Parallel(n_jobs=-1)]: Done  35 tasks      | elapsed:   37.7s


[CV] ... C=0.1, gamma=0.1, score=0.8534076296613802, total=   1.3s
[CV] C=0.1, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  36 tasks      | elapsed:   39.6s


[CV] ... C=0.1, gamma=1, score=0.8534704370179949, total=   1.6s
[CV] C=0.1, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  37 tasks      | elapsed:   40.1s


[CV] ... C=0.1, gamma=1, score=0.8534076296613802, total=   1.6s
[CV] C=0.1, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  38 tasks      | elapsed:   42.1s
```

```
[CV] ... C=0.1, gamma=1, score=0.8534076296613802, total=   1.5s
[CV] C=0.1, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  39 tasks      | elapsed:   42.5s


[CV] ... C=0.1, gamma=10, score=0.8534704370179949, total=   2.2s
[CV] C=0.1, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  40 tasks      | elapsed:   45.2s


[CV] ... C=0.1, gamma=10, score=0.8534076296613802, total=   2.1s
[CV] C=0.1, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  41 tasks      | elapsed:   45.6s


[CV] ... C=0.1, gamma=10, score=0.8534076296613802, total=   2.1s
[CV] C=0.1, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  42 tasks      | elapsed:   48.3s


[CV] ... C=0.1, gamma=100, score=0.8534704370179949, total=   3.1s
[CV] C=0.1, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  43 tasks      | elapsed:   50.3s


[CV] ... C=0.1, gamma=100, score=0.8534076296613802, total=   3.0s
[CV] C=1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  44 tasks      | elapsed:   53.0s


[CV] ... C=0.1, gamma=100, score=0.8534076296613802, total=   3.0s
[CV] C=1, gamma=0.01 ...
[CV] ... C=1, gamma=0.01, score=0.8534704370179949, total=   1.3s
[CV] C=1, gamma=0.01 ...


[Parallel(n_jobs=-1)]: Done  45 tasks      | elapsed:   54.9s
[Parallel(n_jobs=-1)]: Done  46 tasks      | elapsed:   55.0s
```

```
[CV] ... C=1, gamma=0.01, score=0.8534076296613802, total=   1.2s
[CV] C=1, gamma=0.1 ...
[CV] ... C=1, gamma=0.01, score=0.8534076296613802, total=   1.3s
[CV] C=1, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  47 tasks      | elapsed:   56.9s
[Parallel(n_jobs=-1)]: Done  48 tasks      | elapsed:   57.0s


[CV] ... C=1, gamma=0.1, score=0.8534704370179949, total=   1.5s
[CV] C=1, gamma=0.1 ...
[CV] ... C=1, gamma=0.1, score=0.8534076296613802, total=   1.5s
[CV] C=1, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  49 tasks      | elapsed:   59.3s
[Parallel(n_jobs=-1)]: Done  50 tasks      | elapsed:   59.4s


[CV] ... C=1, gamma=0.1, score=0.8534076296613802, total=   1.5s
[CV] C=1, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  51 tasks      | elapsed:  1.0min


[CV] ... C=1, gamma=1, score=0.8534704370179949, total=   2.2s
[CV] C=1, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  52 tasks      | elapsed:  1.0min


[CV] ... C=1, gamma=1, score=0.8534076296613802, total=   2.3s
[CV] C=1, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  53 tasks      | elapsed:  1.1min


[CV] ... C=1, gamma=1, score=0.8534076296613802, total=   2.2s
[CV] C=1, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  54 tasks      | elapsed:  1.1min


[CV] ... C=1, gamma=10, score=0.8534704370179949, total=   2.1s
[CV] C=1, gamma=10 ...
```

```
[Parallel(n_jobs=-1)]: Done  55 tasks      | elapsed:  1.1min


[CV] ... C=1, gamma=10, score=0.8534076296613802, total=   2.2s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  56 tasks      | elapsed:  1.1min


[CV] ... C=1, gamma=10, score=0.8534076296613802, total=   2.1s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  57 tasks      | elapsed:  1.2min


[CV] ... C=1, gamma=100, score=0.8534704370179949, total=   3.1s
[CV] C=1, gamma=100 ...

[Parallel(n_jobs=-1)]: Done  58 tasks      | elapsed:  1.2min


[CV] ... C=1, gamma=100, score=0.8542648949849978, total=   3.0s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  59 tasks      | elapsed:  1.3min


[CV] ... C=10, gamma=0.01, score=0.8534704370179949, total=   1.5s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  60 tasks      | elapsed:  1.3min


[CV] ... C=1, gamma=100, score=0.853836262323189, total=   3.1s
[CV] C=10, gamma=0.01 ...

[Parallel(n_jobs=-1)]: Done  61 tasks      | elapsed:  1.3min


[CV] ... C=10, gamma=0.01, score=0.8534076296613802, total=   1.5s
[CV] C=10, gamma=0.1 ...

[Parallel(n_jobs=-1)]: Done  62 tasks      | elapsed:  1.3min
```

```
[CV] ... C=10, gamma=0.01, score=0.8534076296613802, total=   1.5s
[CV] C=10, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  63 tasks      | elapsed:  1.3min


[CV] ... C=10, gamma=0.1, score=0.8534704370179949, total=   2.2s
[CV] C=10, gamma=0.1 ...


[Parallel(n_jobs=-1)]: Done  64 tasks      | elapsed:  1.4min


[CV] ... C=10, gamma=0.1, score=0.8534076296613802, total=   2.0s
[CV] C=10, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  65 tasks      | elapsed:  1.4min


[CV] ... C=10, gamma=0.1, score=0.8534076296613802, total=   2.0s
[CV] C=10, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  66 tasks      | elapsed:  1.4min


[CV] ... C=10, gamma=1, score=0.8534704370179949, total=   2.3s
[CV] C=10, gamma=1 ...


[Parallel(n_jobs=-1)]: Done  67 tasks      | elapsed:  1.4min


[CV] ... C=10, gamma=1, score=0.8534076296613802, total=   3.0s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  68 tasks      | elapsed:  1.5min


[CV] ... C=10, gamma=1, score=0.8534076296613802, total=   2.9s
[CV] C=10, gamma=10 ...


[Parallel(n_jobs=-1)]: Done  69 tasks      | elapsed:  1.5min


[CV] ... C=10, gamma=10, score=0.8513281919451585, total=   2.1s
[CV] C=10, gamma=10 ...
```

```
[Parallel(n_jobs=-1)]: Done  70 tasks      | elapsed:  1.5min


[CV] ... C=10, gamma=10, score=0.853836262323189, total=   2.3s
[CV] C=10, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  71 tasks      | elapsed:  1.6min


[CV] ... C=10, gamma=10, score=0.8542648949849978, total=   2.3s
[CV] C=10, gamma=100 ...


[Parallel(n_jobs=-1)]: Done  72 tasks      | elapsed:  1.6min


[CV] ... C=10, gamma=100, score=0.8234790059982862, total=   3.7s
[CV] C=10, gamma=100 ...
[CV] ... C=10, gamma=100, score=0.821260180025718, total=   3.6s
[CV] ... C=10, gamma=100, score=0.8169738534076296, total=   3.2s


[Parallel(n_jobs=-1)]: Done  75 out of  75 | elapsed:  1.7min remaining:    0.0s
[Parallel(n_jobs=-1)]: Done  75 out of  75 | elapsed:  1.7min finished


Out[37]: GridSearchCV(cv=None, error_score='raise',
             estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
           decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
           max_iter=-1, probability=False, random_state=None, shrinking=True,
           tol=0.001, verbose=False),
             fit_params=None, iid=True, n_jobs=-1,
             param_grid={'C': [0.001, 0.01, 0.1, 1, 10], 'gamma': [0.01, 0.1, 1, 10, 100]},
             pre_dispatch='2*n_jobs', refit=True, return_train_score='warn',
             scoring=None, verbose=30)

In [38]: print("Best parameters: ", rbf_svm_grid_cv.best_params_)
         print("Best cross-validation score: {:.3f}".format(rbf_svm_grid_cv.best_score_))

Best parameters:  {'C': 1, 'gamma': 100}
Best cross-validation score: 0.854


In [39]: from sklearn.metrics import classification_report

         y_pred = rbf_svm_grid_cv.predict(X_test4)
         print(classification_report(y_test, y_pred))
```

```
             precision    recall  f1-score   support

   negative       0.73      0.01      0.03       569
   positive       0.81      1.00      0.90      2431

avg / total       0.80      0.81      0.73      3000
```

6.2 Applying RandomSearch CV

```
In [40]: from sklearn.model_selection import RandomizedSearchCV
         from scipy import stats

         param_grid  = {"C": stats.uniform(0.001, 10), "gamma": stats.uniform(0.1, 100)}

         rbf_svm = SVC(kernel='rbf')
         rbf_svm_grid_cv =  RandomizedSearchCV(rbf_svm, param_grid, n_jobs=-1, verbose=30)
         rbf_svm_grid_cv

         #perform gridsearch
         rbf_svm_grid_cv.fit(X_train4, y_train)
```

```
Fitting 3 folds for each of 10 candidates, totalling 30 fits
[CV] C=3.826310675830332, gamma=38.987418035791514 ...
[CV] C=3.826310675830332, gamma=38.987418035791514 ...
[CV]  C=3.826310675830332, gamma=38.987418035791514, score=0.8495499357051007, total=   1.8s
[CV] C=3.826310675830332, gamma=38.987418035791514 ...
[CV]  C=3.826310675830332, gamma=38.987418035791514, score=0.8419023136246787, total=   1.8s
[CV] C=2.037069535239874, gamma=55.27804816327809 ...


[Parallel(n_jobs=-1)]: Done   1 tasks      | elapsed:    2.8s
[Parallel(n_jobs=-1)]: Done   2 tasks      | elapsed:    2.8s


[CV]  C=3.826310675830332, gamma=38.987418035791514, score=0.8448349764252036, total=   2.7s
[CV] C=2.037069535239874, gamma=55.27804816327809 ...


[Parallel(n_jobs=-1)]: Done   3 tasks      | elapsed:    6.7s


[CV]  C=2.037069535239874, gamma=55.27804816327809, score=0.8470437017994858, total=   2.7s
[CV] C=2.037069535239874, gamma=55.27804816327809 ...
```

```
[Parallel(n_jobs=-1)]: Done   4 tasks      | elapsed:    6.9s


[CV]  C=2.037069535239874, gamma=55.27804816327809, score=0.8521217316759537, total=   2.8s
[CV] C=4.329336830073098, gamma=40.16790776891425 ...
[CV]  C=2.037069535239874, gamma=55.27804816327809, score=0.8504072010287184, total=   2.7s
[CV] C=4.329336830073098, gamma=40.16790776891425 ...


[Parallel(n_jobs=-1)]: Done   5 tasks      | elapsed:   10.8s
[Parallel(n_jobs=-1)]: Done   6 tasks      | elapsed:   10.8s


[CV]  C=4.329336830073098, gamma=40.16790776891425, score=0.8397600685518424, total=   2.7s
[CV]  C=4.329336830073098, gamma=40.16790776891425, score=0.8465495070724389, total=   2.7s
[CV] C=4.329336830073098, gamma=40.16790776891425 ...
[CV] C=7.696926304605969, gamma=1.9372036764791933 ...


[Parallel(n_jobs=-1)]: Done   7 tasks      | elapsed:   14.6s
[Parallel(n_jobs=-1)]: Done   8 tasks      | elapsed:   14.6s


[CV]  C=7.696926304605969, gamma=1.9372036764791933, score=0.8534704370179949, total=   2.4s
[CV] C=7.696926304605969, gamma=1.9372036764791933 ...


[Parallel(n_jobs=-1)]: Done   9 tasks      | elapsed:   17.9s


[CV]  C=4.329336830073098, gamma=40.16790776891425, score=0.8452636090870125, total=   2.7s
[CV] C=7.696926304605969, gamma=1.9372036764791933 ...


[Parallel(n_jobs=-1)]: Done  10 tasks      | elapsed:   18.5s


[CV]  C=7.696926304605969, gamma=1.9372036764791933, score=0.8534076296613802, total=   2.6s
[CV] C=0.20431319626664568, gamma=47.634857925529516 ...


[Parallel(n_jobs=-1)]: Done  11 tasks      | elapsed:   21.3s


[CV]  C=7.696926304605969, gamma=1.9372036764791933, score=0.8534076296613802, total=   2.6s
[CV] C=0.20431319626664568, gamma=47.634857925529516 ...


[Parallel(n_jobs=-1)]: Done  12 tasks      | elapsed:   21.9s
```

```
[CV]  C=0.20431319626664568, gamma=47.634857925529516, score=0.8534704370179949, total=   2.5s
[CV] C=0.20431319626664568, gamma=47.634857925529516 ...


[Parallel(n_jobs=-1)]: Done  13 tasks      | elapsed:   25.1s


[CV]  C=0.20431319626664568, gamma=47.634857925529516, score=0.8534076296613802, total=   2.4s
[CV] C=7.162626156152059, gamma=30.169646524951155 ...


[Parallel(n_jobs=-1)]: Done  14 tasks      | elapsed:   25.5s


[CV]  C=0.20431319626664568, gamma=47.634857925529516, score=0.8534076296613802, total=   2.6s
[CV] C=7.162626156152059, gamma=30.169646524951155 ...
[CV]  C=7.162626156152059, gamma=30.169646524951155, score=0.8393316195372751, total=   2.4s
[CV] C=7.162626156152059, gamma=30.169646524951155 ...


[Parallel(n_jobs=-1)]: Done  15 tasks      | elapsed:   28.8s
[Parallel(n_jobs=-1)]: Done  16 tasks      | elapsed:   29.0s


[CV]  C=7.162626156152059, gamma=30.169646524951155, score=0.8452636090870125, total=   2.5s
[CV] C=2.6433507719894913, gamma=21.51920890585667 ...
[CV]  C=7.162626156152059, gamma=30.169646524951155, score=0.8409772824689241, total=   2.4s
[CV] C=2.6433507719894913, gamma=21.51920890585667 ...


[Parallel(n_jobs=-1)]: Done  17 tasks      | elapsed:   32.4s
[Parallel(n_jobs=-1)]: Done  18 tasks      | elapsed:   32.5s


[CV]  C=2.6433507719894913, gamma=21.51920890585667, score=0.8513281919451585, total=   2.3s
[CV] C=2.6433507719894913, gamma=21.51920890585667 ...
[CV]  C=2.6433507719894913, gamma=21.51920890585667, score=0.8529789969995714, total=   2.3s
[CV] C=6.047013289740188, gamma=74.26334500056159 ...


[Parallel(n_jobs=-1)]: Done  19 tasks      | elapsed:   35.7s
[Parallel(n_jobs=-1)]: Done  20 tasks      | elapsed:   35.8s


[CV]  C=2.6433507719894913, gamma=21.51920890585667, score=0.8534076296613802, total=   2.2s
[CV] C=6.047013289740188, gamma=74.26334500056159 ...


[Parallel(n_jobs=-1)]: Done  21 tasks      | elapsed:   38.9s
```

```
[CV]  C=6.047013289740188, gamma=74.26334500056159, score=0.8213367609254498, total=   2.9s
[CV] C=6.047013289740188, gamma=74.26334500056159 ...

[Parallel(n_jobs=-1)]: Done  22 tasks      | elapsed:   40.1s

[CV]  C=6.047013289740188, gamma=74.26334500056159, score=0.8324046292327475, total=   2.8s
[CV] C=7.082857219471582, gamma=8.584740676392588 ...

[Parallel(n_jobs=-1)]: Done  23 tasks      | elapsed:   43.1s

[CV]  C=6.047013289740188, gamma=74.26334500056159, score=0.831118731247321, total=   2.8s
[CV] C=7.082857219471582, gamma=8.584740676392588 ...

[Parallel(n_jobs=-1)]: Done  24 tasks      | elapsed:   44.3s

[CV]  C=7.082857219471582, gamma=8.584740676392588, score=0.8530419880034276, total=   2.1s
[CV] C=7.082857219471582, gamma=8.584740676392588 ...

[Parallel(n_jobs=-1)]: Done  25 tasks      | elapsed:   46.0s

[CV]  C=7.082857219471582, gamma=8.584740676392588, score=0.853836262323189, total=   2.2s
[CV] C=6.469737291475392, gamma=8.49642449206237 ...

[Parallel(n_jobs=-1)]: Done  26 tasks      | elapsed:   47.4s

[CV]  C=7.082857219471582, gamma=8.584740676392588, score=0.8534076296613802, total=   2.2s
[CV] C=6.469737291475392, gamma=8.49642449206237 ...

[Parallel(n_jobs=-1)]: Done  27 tasks      | elapsed:   49.2s

[CV]  C=6.469737291475392, gamma=8.49642449206237, score=0.8534704370179949, total=   2.1s
[CV] C=6.469737291475392, gamma=8.49642449206237 ...
[CV]  C=6.469737291475392, gamma=8.49642449206237, score=0.8534076296613802, total=   2.2s
[CV]  C=6.469737291475392, gamma=8.49642449206237, score=0.8534076296613802, total=   2.1s

[Parallel(n_jobs=-1)]: Done  30 out of  30 | elapsed:   53.1s finished
```

```
Out[40]: RandomizedSearchCV(cv=None, error_score='raise',
                   estimator=SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
           decision_function_shape='ovr', degree=3, gamma='auto', kernel='rbf',
           max_iter=-1, probability=False, random_state=None, shrinking=True,
           tol=0.001, verbose=False),
                   fit_params=None, iid=True, n_iter=10, n_jobs=-1,
                   param_distributions={'C': <scipy.stats._distn_infrastructure.rv_frozen obje
                   pre_dispatch='2*n_jobs', random_state=None, refit=True,
                   return_train_score='warn', scoring=None, verbose=30)

In [41]: print("Best parameters: ", rbf_svm_grid_cv.best_params_)
         print("Best cross-validation score: {:.3f}".format(rbf_svm_grid_cv.best_score_))

Best parameters:  {'C': 7.696926304605969, 'gamma': 1.9372036764791933}
Best cross-validation score: 0.853
```

---

7 Conclusion

7.0.1 The best results are obtained using 'Bag of words' vectorization technique with hyper-parameter C = 10,and gamma = 0.1, the best CV score achieved is 89.6%.