```
 1 C:\Users\Samuel\PycharmProjects\pythonProject\venv\
   Scripts\python.exe "C:\Program Files\JetBrains\
   PyCharm 2020.2.1\plugins\python\helpers\pydev\pydevd.
   py" --multiproc --qt-support=auto --client 127.0.0.1
    --port 50126 --file C:/Users/Samuel/PycharmProjects/
   Doc2Vec_Movie_Analysis/main.py
 2 Connected to pydev debugger (build 212.5457.59)
 3 [nltk_data] Downloading package punkt to
 4 [nltk_data]     C:\Users\Samuel\AppData\Roaming\
   nltk_data...
 5 [nltk_data]   Package punkt is already up-to-date!
 6 Initialization Complete. Time: 0.0 min
 7 C:\Program Files\JetBrains\PyCharm 2020.2.1\plugins\
   python\helpers\pydev\pydevd.py:1476: DtypeWarning:
   Columns (9) have mixed types.Specify dtype option on
   import or set low_memory=False.
 8   return self._exec(is_module, entry_point_fn,
   module_name, file, globals, locals)
 9 Before Preprocessing:
10                                             narrative
                                                Product
11 1  I contacted Ally on Friday XX/XX/XXXX after fa
   ...                              Vehicle loan or
   lease
12 3  Thanks to the CFPB I was able to get my accoun
   ...   Credit reporting, credit repair services, or o
   ...
13 5  Hello This complaint is against the three cred
   ...   Credit reporting, credit repair services, or o
   ...
14 6  I am a victim of Identity Theft & currently ha
   ...   Credit reporting, credit repair services, or o
   ...
15 9  Two accounts are still on my credit history af
   ...   Credit reporting, credit repair services, or o
   ...
16    Dataframe Shape: (829857, 2)
17    Number of words in dataset: 158726606
18 Preprocessing Complete. Time: 4.82 min
19    Number of words in dataset: 158298539, delta: -
   428067
```

```
20 100%|████████| 580899/580899 [00:00<00:00, 1126491.
   47it/s]
21 Ready To Train Model. Time: 33.1 min
22 100%|████████| 580899/580899 [00:00<00:00, 1126598.
   77it/s]
23 100%|████████| 580899/580899 [00:00<00:00, 1100722.
   56it/s]
24 100%|████████| 580899/580899 [00:00<00:00, 1062069.
   00it/s]
25 100%|████████| 580899/580899 [00:00<00:00, 1093440.
   26it/s]
26 100%|████████| 580899/580899 [00:00<00:00, 1061715.
   88it/s]
27 100%|████████| 580899/580899 [00:00<00:00, 1062401.
   05it/s]
28 100%|████████| 580899/580899 [00:00<00:00, 1093030.
   66it/s]
29 100%|████████| 580899/580899 [00:00<00:00, 1079143.
   74it/s]
30 100%|████████| 580899/580899 [00:00<00:00, 1073479.
   61it/s]
31 100%|████████| 580899/580899 [00:00<00:00, 1093417.
   19it/s]
32 100%|████████| 580899/580899 [00:00<00:00, 1084652.
   09it/s]
33 100%|████████| 580899/580899 [00:00<00:00, 1093444.
   18it/s]
34 100%|████████| 580899/580899 [00:00<00:00, 1084116.
   86it/s]
35 100%|████████| 580899/580899 [00:00<00:00, 1105493.
   60it/s]
36 100%|████████| 580899/580899 [00:00<00:00, 1104443.
   26it/s]
37 100%|████████| 580899/580899 [00:00<00:00, 1083898.
   39it/s]
38 100%|████████| 580899/580899 [00:00<00:00, 1083446.
   76it/s]
39 100%|████████| 580899/580899 [00:00<00:00, 1082859.
   78it/s]
40 100%|████████| 580899/580899 [00:00<00:00, 1094083.
   96it/s]
```

```
41 100%|████████| 580899/580899 [00:00<00:00, 1025582.
   02it/s]
42 100%|████████| 580899/580899 [00:00<00:00, 1091613.
   93it/s]
43 100%|████████| 580899/580899 [00:00<00:00, 1079752.
   54it/s]
44 100%|████████| 580899/580899 [00:00<00:00, 1093891.
   41it/s]
45 100%|████████| 580899/580899 [00:00<00:00, 997237.
   25it/s]
46 100%|████████| 580899/580899 [00:00<00:00, 1090821.
   71it/s]
47 100%|████████| 580899/580899 [00:00<00:00, 1113849.
   64it/s]
48 100%|████████| 580899/580899 [00:00<00:00, 1103777.
   31it/s]
49 100%|████████| 580899/580899 [00:00<00:00, 1059134.
   54it/s]
50 100%|████████| 580899/580899 [00:00<00:00, 1125331.
   74it/s]
51 100%|████████| 580899/580899 [00:00<00:00, 1084466.
   22it/s]
52 Training Complete.
53 Ready To Fit Classifier. Time: 107.27 min
54 C:\Users\Samuel\PycharmProjects\pythonProject\venv\
   lib\site-packages\sklearn\linear_model\_logistic.py:
   814: ConvergenceWarning: lbfgs failed to converge (
   status=1):
55 STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
56
57 Increase the number of iterations (max_iter) or scale
    the data as shown in:
58     https://scikit-learn.org/stable/modules/
   preprocessing.html
59 Please also refer to the documentation for
   alternative solver options:
60     https://scikit-learn.org/stable/modules/
   linear_model.html#logistic-regression
61   n_iter_i = _check_optimize_result(
62 Results. Time: 109.01 min
63 Testing accuracy 0.39520722370841666
```

```
64 Testing F1 score: 0.23413934310985143
65 100%|████████| 580899/580899 [00:00<00:00, 1093632
   .16it/s]
66 100%|████████| 580899/580899 [00:00<00:00, 1092694
   .39it/s]
67 100%|████████| 580899/580899 [00:00<00:00, 1093836
   .41it/s]
68 100%|████████| 580899/580899 [00:00<00:00, 1089254
   .35it/s]
69 100%|████████| 580899/580899 [00:00<00:00, 1093105
   .20it/s]
70 100%|████████| 580899/580899 [00:00<00:00, 1062550
   .70it/s]
71 100%|████████| 580899/580899 [00:00<00:00, 1093394
   .62it/s]
72 100%|████████| 580899/580899 [00:00<00:00, 1061269
   .14it/s]
73 100%|████████| 580899/580899 [00:00<00:00, 1093717
   .58it/s]
74 100%|████████| 580899/580899 [00:00<00:00, 1083580
   .23it/s]
75 100%|████████| 580899/580899 [00:00<00:00, 1094302
   .63it/s]
76 100%|████████| 580899/580899 [00:00<00:00, 1079560
   .21it/s]
77 100%|████████| 580899/580899 [00:00<00:00, 1061009
   .41it/s]
78 100%|████████| 580899/580899 [00:00<00:00, 1068791
   .82it/s]
79 100%|████████| 580899/580899 [00:00<00:00, 1092720
   .85it/s]
80 100%|████████| 580899/580899 [00:00<00:00, 1084897
   .44it/s]
81 100%|████████| 580899/580899 [00:00<00:00, 1087394
   .40it/s]
82 100%|████████| 580899/580899 [00:00<00:00, 1060962
   .29it/s]
83 100%|████████| 580899/580899 [00:00<00:00, 1062932
   .20it/s]
84 100%|████████| 580899/580899 [00:00<00:00, 1078781
   .56it/s]
```

```
 85 100%|████████| 580899/580899 [00:00<00:00, 1093390
    .21it/s]
 86 100%|████████| 580899/580899 [00:00<00:00, 1110364
    .38it/s]
 87 100%|████████| 580899/580899 [00:00<00:00, 1094053
    .01it/s]
 88 100%|████████| 580899/580899 [00:00<00:00, 1032729
    .91it/s]
 89 100%|████████| 580899/580899 [00:00<00:00, 1062145
    .39it/s]
 90 100%|████████| 580899/580899 [00:00<00:00, 1061884
    .77it/s]
 91 100%|████████| 580899/580899 [00:00<00:00, 1062233
    .84it/s]
 92 100%|████████| 580899/580899 [00:00<00:00, 1083633
    .73it/s]
 93 100%|████████| 580899/580899 [00:00<00:00, 1093354
    .39it/s]
 94 100%|████████| 580899/580899 [00:00<00:00, 1126541
    .99it/s]
 95 100%|████████| 580899/580899 [00:00<00:00, 1093433
    .88it/s]
 96 Training Complete.
 97 Ready To Fit Classifier. Time: 215.1 min
 98 C:\Users\Samuel\PycharmProjects\pythonProject\venv\
    lib\site-packages\sklearn\linear_model\_logistic.py:
    814: ConvergenceWarning: lbfgs failed to converge (
    status=1):
 99 STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
100
101 Increase the number of iterations (max_iter) or
    scale the data as shown in:
102     https://scikit-learn.org/stable/modules/
    preprocessing.html
103 Please also refer to the documentation for
    alternative solver options:
104     https://scikit-learn.org/stable/modules/
    linear_model.html#logistic-regression
105   n_iter_i = _check_optimize_result(
106 Testing accuracy 0.6458278103133862
107 Testing F1 score: 0.6204632424005363
```

```
108 Ready To Fit Classifier. Time: 345.35 min
109 C:\Users\Samuel\PycharmProjects\pythonProject\venv\
    lib\site-packages\sklearn\linear_model\_logistic.py:
    814: ConvergenceWarning: lbfgs failed to converge (
    status=1):
110 STOP: TOTAL NO. of ITERATIONS REACHED LIMIT.
111
112 Increase the number of iterations (max_iter) or
    scale the data as shown in:
113     https://scikit-learn.org/stable/modules/
    preprocessing.html
114 Please also refer to the documentation for
    alternative solver options:
115     https://scikit-learn.org/stable/modules/
    linear_model.html#logistic-regression
116   n_iter_i = _check_optimize_result(
117 Testing accuracy 0.6421725752938247
118 Testing F1 score: 0.6130176880860239
119
120 Process finished with exit code 0
121
```