

```

1 E:\python.exe C:/Users/R0G/Desktop/00/JD_NLP-
text_classification_Project1/JD_NLP1-
text_classification_assignment1/model.py
2 ['——\n', '000\n', '0÷000\n', '"0\n', '00\n', '00\n
', ': \n', '→\n', '0 \n', '&\n', '*\n', '00\n', '~~~\n
n', '\n', '.\n', '0\n', '.0\n', './\n', '--\n', '0\n
', '0"\n', '0\n', '000\n', '00\n', '0000\n', '0000\n
', '00\n', '000\n', '0\n', '0/\n', '0\n', '0\n', '000
0\n', '0000\n', '00\n', '}\n', '00\n', '\n', '0\n
', '0000\n', '0000\n', '0000\n', '0000\n', '0000\n
', '0000\n', '00\n', '00\n', '00\n', '0000\n', '\n
', 'x×x\n', '0000\n', '000\n', '00\n', '0000\n', '000
0\n', '0000\n', '0000\n', '0000\n', '0000\n', '0000\n
', '0000\n', '0000\n', '00\n', '0000\n', '0000\n
', '0000\n', '0000\n', '0000\n', '0000\n', '0000\n
', '0000\n', '0000\n', '0000\n', '0000\n', '0.\n
', '0000\n', '.0\n', '00\n', '0000\n', '0/\n', '0000\
n', '0000\n', '0000\n', '0000\n', '0000\n', '0000\n
', '0000\n', '0000\n', '0000\n', '0000\n', '0000\n
', '0000\n', '00\n', '00\n', '://\n', '0\n', '000\n
', '0000\n', '000\n', '00\n', '...\n
', '.....\n', '.....0\n
', '000000\n', '0000\n', '0\n', '0000\n', '00'00\n
', '000\n', '0\n', '00\n', '00\n', '0\n', '0000\n
', '0000\n', '0000\n', '0\n', '00000\n', '↑\n', '.0\n
', '0000\n', '00\n', '0000\n', '0000\n', '0000\n
', '0000\n', '0000\n', '0000\n', '0000\n', '0000\n
', '0000\n', '0000\n', '00\n', '000\n', '00\n', '000
0\n', '0000\n', '0000\n', '0000\n', '0000\n', '00\n
', '0\n', '0\n', '000\n', '0000\n', '0000\n', '000\n
', '000\n', '::\n', '000\n', '000\n', '000\n', '000\n
', '000\n', '000\n', '000\n', '000\n', '.....\n', '——\n
', '?\n', '0\n', '0\n', '"\n', '"\n', '0\n', '0\n
', '0\n', '0\n', '0\n', '0\n', '0\n', '0\n', ',\n
', '0\n', '"\n', '? \n', '·\n', '——\n', '——\n', '?\n
', '-\n', '<\n', '>\n', '0\n', '0\n', '0\n', '0\n
', '[\n', ']\n', '(\n', ')\n', '-\n', '+\n', '0\n
', '×\n', '0\n', '/\n', '0\n', '0\n', '0\n', '0\n
', '0\n', '0\n', '0\n', '0\n', '0\n', '0\n', '0\n', '
B\n', '"\n', ';\n', '#\n', '@\n', '0\n', 'μ\n', '0\n
', '00\n', '× \n', '0\n', '0\n', '0\n', 'sub\n', 'exp
\n', 'sup\n', 'sub\n', 'Lex \n', '0\n', '0\n', '0\n
', '0\n', '0\n', '00\n', '00\n', '0\n', '00\n', '0\n
', '0±\n', '00\n', '0λ\n', '00\n', '00\n', '= \n', '0\

```

[illegible]

```

2 \n', '00\n', '00\n', '00\n', '0000\n', '00\n', '00\n
  ', '0\n', '00\n', '00\n', '00\n', '00\n', '0\n', '00\n', '0\n
  ', '00\n', '00\n', '0\n', '00\n', '00\n', '00\n', '00
\n', '00\n', '0000\n', '0\n', '00\n', '00\n', '0\n
  ', '00\n', '00\n', '00\n', '00\n', '00\n', '00\n', '
00\n', '00\n', '00\n', '00\n', '000\n', '00\n', '0\n
  ', '00\n', '00\n', '0\n', '00\n', '0\n', '00\n', '00\
n', '0\n', '00\n', '00\n', '00\n', '0\n', '00\n', '0\
n', '00\n', '0\n', '00\n', '00\n', '0\n', '00\n', '0\
n', '00\n', '0\n', '00\n', '00\n', '00\n', '00\n', '0
\n', '0\n', '00\n', '0\n', '00\n', '0\n', '00\n', '0\
n', '0\n', '0\n', '00\n', '00\n', '000\n', '00\n', '0
\n', '00\n', '0\n', '00\n', '0\n', '00\n', '00\n', '
00\n', '00\n', '00\n', '0\n', '0\n', '0000\n', '0\n
  ', '0\n', '00\n', '0\n', '0\n', '0\n', '0\n', '00\n
  ', '0\n', '00\n', '000\n', '000\n', '00\n', '00\n', '
0\n', '00\n', '00\n', '0\n', '00\n', '00\n', '000\n
  ', '00\n', '00\n', '00\n', '00\n', '0\n', '00\n', '0\
n', '0\n', '00\n', '00\n', '00\n', '00\n', '000\n', '
00\n', '00\n', '0\n', '00\n', '00\n', '00\n', '0\n
  ', '0\n', '0\n', '0000\n', '00\n', '0\n', '00\n', '00
\n', '00\n', '0\n', '0\n', '00\n', '000\n', '0\n', '
0000\n', '00\n', '00\n', '00\n', '00\n', '0\n', '00\n
  ', '00\n', '0\n', '00\n', '0\n', '00\n', '0\n', '0\n
  ', '00\n', '000\n', '000\n', '00\n', '0\n', '0\n', '
00\n', '0\n', '0\n', '00\n', '00\n', '00\n', '000\n
  ', '0000\n', '00\n', '00\n', '0000\n', '000\n', '000\
n', '00\n', '00\n', '00\n', '00\n', '0\n', '0\n', '00
\n', '000\n', '00\n', '00\n', '00\n', '00\n', '00\n
  ', '0\n', '00\n', '00\n', '0\n', '00\n', '0\n', '00\n
  ', '000\n', '000\n', '00\n', '00\n', '00\n', '0000\n
  ', '0000\n', '0000\n', '0000\n', '0000\n', '00\n', '0
\n', '00\n', '00\n', '00\n', '00\n', '00\n', '0\n', '
0\n', '0\n', '0\n', '0\n', '0\n', '0\n', '00\n', '0\n
  ', '0\n', '0\n']

```

```

3      label

```

```

      text

```

```

4 0      0000      00 00 00 0 0000 00 00 ||
    000 0 00 0
5 1      0000  0 0 00 || 0 || 00 || 00 NUM 0 || 00 00
    00 0 0 ...
6 2      00  0 0 00 0 0 00 00 00 0 00 0 0 || 00 0 00
    0 || 0...
7 3      00  00 00 0 00 00 000 0 || 0000 00 NUM 0

```

```

7 NUM 0 || ...
8 4      000  00 00 00 0 00 000 0 0 00 00 0 0 00 00
  00 00 0 ...
9 ...      ...
  ...
10 28368      00      00 00 ||
  0000 || 00 00
11 28369 00000      00 0 0 0 00 0
  00 00 || 00
12 28370 other 00 00 0 0 00 0 00 0 00 00 || 0 00 || 00
  00 || ...
13 28371 0000 00 00 || 0 00 0 0 000 || 0 00 00 00000
  ( 000 )...
14 28372 000      000 0 00 || 00 00 0 || 00 00 0 00 00
  || 00 0 00
15
16 [28373 rows x 2 columns]
17 0      00 00 00 0 0000 00 00 || 000 0
  00 0
18 1      0 0 00 || 0 || 00 || 00 NUM 0 || 00 00 00 0
  0 ...
19 2      0 0 00 0 0 00 00 00 0 00 0 0 || 00 0 00 0
  || 0...
20 3      00 00 0 00 00 000 0 || 0000 00 NUM 0 NUM 0
  || ...
21 4      00 00 00 0 00 000 0 0 00 00 0 0 00 00 00 00
  0 ...
22
  ...
23 28368      00 00 || 0000 ||
  00 00
24 28369      00 0 0 0 00 0 00 00
  || 00
25 28370 00 00 0 0 00 0 00 0 00 00 || 0 00 || 00 00
  || ...
26 28371 00 00 || 0 00 0 0 000 || 0 00 00 00000 ( 000
  )...
27 28372 000 0 00 || 00 00 0 || 00 00 0 00 00 || 00
  0 00
28 Name: text, Length: 28373, dtype: object
29 0      00 00 00 0 0000 00 00 ||
  000 00
30 1      0 00 || 0 || 00 || 00 NUM 0 || 00 00 00 0 0
  00...

```

```

31 2      0 0 00 0 00 00 0 00 0 || 0 || 00 00 || 00 00
    0...
32 3      00 00 00 00 000 0 || 0000 00 NUM NUM 0 || 00
    0...
33 4      00 00 00 00 000 00 00 0 00 00 00 0 0 0 00 00
    0...
34      ...

35 28368      00 00 || 0000 ||
    00 00
36 28369      00 0 0 00 00 00
    || 00
37 28370      00 00 0 00 00 0 00 00 || 0 00 || 00 || 0 ||
    00...
38 28371      00 00 || 00 0 0 000 || 0 00 00 00000 000 ||
    00...
39 28372      000 0 00 || 00 00 0 || 00 00 00 00 || 00
    0 00
40 Name: text, Length: 28373, dtype: object
41      label    ... tfidf22709
42 0      0000    ...      0.0
43 1      0000    ...      0.0
44 2      00      ...      0.0
45 3      00      ...      0.0
46 4      000     ...      0.0
47 ...      ...    ...      ...
48 28368      00      ...      0.0
49 28369 00000     ...      0.0
50 28370 other     ...      0.0
51 28371 0000     ...      0.0
52 28372 000      ...      0.0
53
54 [28373 rows x 22712 columns]
55 transform w2v
56      label    ... w2v_4_max_299
57 0      0000    ...      0.054239
58 1      0000    ...      0.366395
59 2      00      ...     -0.033314
60 3      00      ...      0.040619
61 4      000     ...      0.097385
62 ...      ...    ...      ...
63 28368      00      ...     -0.161887
64 28369 00000     ...     -0.085961
65 28370 other     ...      0.060511

```

```

66 28371    0000 ...      0.446367
67 28372    000 ...     -0.370565
68
69 [28373 rows x 25712 columns]
70          label ...          lda_10
71 0          0000 ...          [0.0]
72 1          0000 ... [0.15926872193813324]
73 2           00 ...          [0.0]
74 3           00 ...          [0.0]
75 4          000 ...          [0.0]
76 ...      ... ...      ...
77 28368     00 ... [0.41494041681289673]
78 28369 000000 ... [0.18126417696475983]
79 28370  other ...          [0.0]
80 28371    0000 ...          [0.0]
81 28372    000 ...          [0.0]
82
83 [28373 rows x 25723 columns]
84 Building prefix dict from the default dictionary ...
85 Loading model from cache C:\Users\ROG\AppData\Local\
  Temp\jieba.cache
86 Loading model cost 0.635 seconds.
87 Prefix dict has been built succesfully.
88          label ... mean_word_len
89 0          0000 ...      2.000000
90 1          0000 ...      1.647727
91 2           00 ...      1.633333
92 3           00 ...      1.904762
93 4          000 ...      1.617647
94 ...      ... ...      ...
95 28368     00 ...      2.285714
96 28369 000000 ...      1.600000
97 28370  other ...      1.653846
98 28371    0000 ...      1.757576
99 28372    000 ...      1.823529
100
101 [28373 rows x 25740 columns]
102          label
                                     text
103 0          0000  00 0 0 0 0 0 0 0 0 0 0 0 0 0 0 || 00 00 0
    00 00 ...
104 1          0000  00 0 00 NUM 0 || 0000 00 0 00 0 NUM 0
    || 00 NU...
105 2          0000  0 0 0 00 00 00 0 0 || 00 00 || 0 0 0 00

```

```

105  0 0 0...
106  3      000  00 00 00 00 0 00 || 0 0 00 00 000 0 00
      0 00 0 ...
107  4      other                      00 0 0 0 00 00 0 0
      || 0 0 00
108  ...      ...
      ...
109  1575     00  00 00 || 0 00 0 0 0 00 00 00 0 || 0 00
      0 0 0 |...
110  1576     0000  00 00 || 0 0 00 0 000 || 00 00 00 00
      || 00 0 0...
111  1577     000                      000000 || 00
      0 00 0 00
112  1578     000                      00 || 00 0 B
      0 0 000 00
113  1579     0000      00 00 0 0 00000 || 0 0 0 0 00 || 0 00
      00 00 0 0
114
115  [1580 rows x 2 columns]
116  0      00 0 0 0 0 0 0 0 0 0 0 00 00 || 00 00 0 00
      00 ...
117  1      00 0 00 NUM 0 || 0000 00 0 00 0 NUM 0 || 00
      NU...
118  2      0 0 0 00 00 00 0 0 || 00 00 || 0 0 0 00 0 00
      0...
119  3      00 00 00 00 0 00 || 0 0 00 00 000 0 00 0 00
      0 ...
120  4                      00 0 0 0 00 00 0 0 || 0
      0 00
121                      ...

122  1575     00 00 || 0 00 0 0 0 00 00 00 0 || 0 00 0 0 0
      |...
123  1576     00 00 || 0 0 00 0 000 || 00 00 00 00 || 00 0
      0...
124  1577                      000000 || 00 0 00
      0 00
125  1578                      00 || 00 0 B 0 0
      000 00
126  1579      00 00 0 0 00000 || 0 0 0 0 00 || 0 00 00
      00 0 0
127  Name: text, Length: 1580, dtype: object
128  0      00 0 0 0 0 0 0 0 00 00 || 00 00 0 00 00 00 0
      0...

```

```

129 1      00 0 00 NUM 0 || 0000 00 0 00 NUM 0 || 00
      NUM-...
130 2      0 0 0 00 00 0 || 00 00 || 0 0 0 00 0 00 00
      00 ...
131 3      00 00 00 00 0 00 || 00 000 0 00 0 0 0 0
      || 0...
132 4      00 0 00 00 0 || 0
      0 00
133      ...

134 1575   00 00 || 00 0 0 00 00 00 || 0 00 0 0 0 || 00
      0...
135 1576   00 00 || 0 00 000 || 00 00 00 00 || 00 0 0
      000...
136 1577   00000 || 00 00
      0 00
137 1578   00 || 00 0 B 0 0
      000 00
138 1579   00 00 00000 || 0 0 0 00 || 00
      00 0

139 Name: text, Length: 1580, dtype: object
140      label    ... tfidf22709
141 0      0000    ...      0.0
142 1      0000    ...      0.0
143 2      0000    ...      0.0
144 3      000     ...      0.0
145 4      other   ...      0.0
146 ...      ...   ...      ...
147 1575     00     ...      0.0
148 1576   0000    ...      0.0
149 1577     000     ...      0.0
150 1578     000     ...      0.0
151 1579   0000    ...      0.0
152
153 [1580 rows x 22712 columns]
154 transform w2v
155      label    ... w2v_4_max_299
156 0      0000    ...     -0.016342
157 1      0000    ...     -0.047609
158 2      0000    ...      0.395582
159 3      000     ...      0.180093
160 4      other   ...      0.113059
161 ...      ...   ...      ...
162 1575     00     ...     -0.282726

```



```

163 1576 0000 ... 0.074011
164 1577 000 ... -0.127310
165 1578 000 ... 0.263881
166 1579 0000 ... -0.318620
167
168 [1580 rows x 25712 columns]
169 label ... lda_10
170 0 0000 ... [0.2440698742866516]
171 1 0000 ... [0.0]
172 2 0000 ... [0.5561922788619995]
173 3 000 ... [0.0]
174 4 other ... [0.0]
175 ... ... ...
176 1575 00 ... [0.0]
177 1576 0000 ... [0.12865708768367767]
178 1577 000 ... [0.2714894413948059]
179 1578 000 ... [0.0]
180 1579 0000 ... [0.0]
181
182 [1580 rows x 25723 columns]
183 label ... mean_word_len
184 0 0000 ... 1.473684
185 1 0000 ... 2.194444
186 2 0000 ... 1.588889
187 3 000 ... 1.566038
188 4 other ... 1.416667
189 ... ... ...
190 1575 00 ... 1.500000
191 1576 0000 ... 1.848485
192 1577 000 ... 2.142857
193 1578 000 ... 1.666667
194 1579 0000 ... 1.666667
195
196 [1580 rows x 25740 columns]
197 X_train: (28373, 25738) y_train: (28373,)
198 [ 1195  2050  2916 10613  1305   912  2443  1501
    1144  1191  3103]
199 E:\Lib\site-packages\xgboost\core.py:430:
    FutureWarning: Pass `objective` as keyword args.
    Passing these as positional arguments will be
    considered as error in future releases.
200 warnings.warn(
201 E:\Lib\site-packages\xgboost\sklearn.py:1146:
    UserWarning: The use of label encoder in

```

```
201 XGBClassifier is deprecated and will be removed in a
    future release. To remove this warning, do the
    following: 1) Pass option use_label_encoder=False
    when constructing XGBClassifier object; and 2)
    Encode your labels (y) as integers starting with 0,
    i.e. 0, 1, 2, ..., [num_class - 1].
202 warnings.warn(label_encoder_deprecation_msg,
    UserWarning)
203 [22:18:09] WARNING: C:/Users/Administrator/workspace
    /xgboost-win64_release_1.4.0/src/learner.cc:1095:
    Starting in XGBoost 1.3.0, the default evaluation
    metric used with the objective 'multi:softprob' was
    changed from 'merror' to 'mlogloss'. Explicitly set
    eval_metric if you'd like to restore the old
    behavior.
204 [2 8 7 ... 3 3 5]
205 0.8037974683544303
206
207 Process finished with exit code 0
208
```