


• Submission result

Submission and Description	Private Score	Public Score
tfidf_RM.csv 5 days ago by dattyDM add submission details	0.28933	0.29016

fig1. Best submission result

• Ranking

91	—	趕工大隊		0.28933	5	5d
----	---	------	---	---------	---	----

Due to the word embedding of my w2v is too slow, I haven't caught up the deadline time. And also I encounter lots of environmental problems when doing the DL stuffs. The followings are the classification ways I tried.

1. tfidf + random forest model
2. tfidf + naive bayes model
3. tfidf + logistic regression model

Also, the KNN / SVM models and deep learning model all run so long, so I can't get the results.

• Future works

Although I can't caught up the deadline but I still try hard to work on the competition. (Never give up!!)

```

clf = RandomForestClassifier()
Y = train_df['emotion']

scores = cross_val_score(clf, X, Y)

/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_validation.py:372: FitFailedWarning:
5 fits failed out of a total of 5.
The score on these train-test partitions for these parameters will be set to nan.
If these failures are not expected, you can try to debug them by setting error_score='raise'.

Below are more details about the failures:
-----
5 fits failed with the following error:
TypeError: only size-1 arrays can be converted to Python scalars

The above exception was the direct cause of the following exception:

Traceback (most recent call last):
  File "/usr/local/lib/python3.7/dist-packages/sklearn/model_selection/_validation.py", line 680, in _fit_and_score
    estimator.fit(X_train, y_train, **fit_params)
  File "/usr/local/lib/python3.7/dist-packages/sklearn/ensemble/_forest.py", line 328, in fit
    X, y, multi_output=True, accept_sparse="csc", dtype=DTYPE
  File "/usr/local/lib/python3.7/dist-packages/sklearn/base.py", line 581, in _validate_data
    X, y = check_X_y(X, y, **check_params)
  File "/usr/local/lib/python3.7/dist-packages/sklearn/utils/validation.py", line 976, in check_X_y
    estimator=estimator,
  File "/usr/local/lib/python3.7/dist-packages/sklearn/utils/validation.py", line 746, in check_array
    array = np.asarray(array, order=order, dtype=dtype)
  File "/usr/local/lib/python3.7/dist-packages/numpy/core/_asarray.py", line 83, in asarray
    return array(a, dtype, copy=False, order=order)
ValueError: setting an array element with a sequence.

warnings.warn(some_fits_failed_message, FitFailedWarning)

```

However, I encounter a big bug when using w2v for the competition data (which has same code for homework worked well). However, I fit it at the end by finding out the wrong arrays in the list for fitting.

My mythologies are as follows:

1. w2v + Gradient boosting classifier (not enough time)

```

from sklearn.ensemble import GradientBoostingClassifier

clf = GradientBoostingClassifier(n_estimators=100, learning_rate=1.0,
                                max_depth=1, random_state=0, verbose=1).fit(X, Y)

```

Iter	Train Loss	Remaining Time
1	1.7811	123.35m

2. w2v + Random forest classifier

Submission and Description	Private Score	Public Score
w2v_RF.csv a few seconds ago by dattyDM add submission details	0.30729	0.30684

3. w2v + HistGradientBoostingClassifier

Submission and Description	Private Score	Public Score
w2v_HB.csv a few seconds ago by dattyDM add submission details	0.30716	0.30665

** I will try DP and other classifiers and submit to late submission to the leaderboard.*

- **What I learn?**

(1) NEVER STUCK TO ONE KERNAL!!!

Jupyter, kaggle, google collab notebook!!!

(2) lots of great kits and packages for data mining

ex. tqdm, space...

(3) ability of python

I used to use C++, java stuffs, but not familiar with python. Writing python should know the right package, it will take less time to get the results...