

```

1  #!/usr/bin/env python
2  # -*- coding: utf-8 -*-
3
4  import sys
5
6  # sys.path.append('/home/moegwjawel/atec_project/train/lmf/lib/python3.6/site-
  packages') # delete when rank
7
8  import logging
9
10 import numpy as np
11 import pandas as pd
12
13 from sklearn.preprocessing import StandardScaler
14
15 import joblib
16
17 input_data_path = '/home/admin/workspace/job/input/test.jsonl'
18 output_predictions_path = '/home/admin/workspace/job/output/predictions.jsonl'
19
20
21 # input_data_path = '/mnt/atec/train.jsonl' # delete when rank
22 # output_predictions_path = './predictions.jsonl' # delete when rank
23
24 def rank_v0(pd_raw_data):
25     pred_list = []
26
27     # 2. 数据预处理
28     pd_data = pd_raw_data.copy(deep=True)
29     pd_data = pd_data.drop(columns=['memo_polish'], axis=1)
30     print('drop memo', pd_data.shape)
31     # step3, 去除取值只有一个值的列, -15col
32     drop_list = ['x2', 'x55', 'x91', 'x96', 'x107', 'x184', 'x198', 'x207',
33                  'x209', 'x261', 'x319', 'x384', 'x436',
34                  'x452', 'x456']
35     pd_data = pd_data.drop(columns=drop_list)
36     print('drop col', pd_data.shape)
37     # step4, 填充缺失值
38     pd_data = pd_data.fillna(pd_data.mean())

```