Article

July 13, 2021

0.1

:

1. . ,
, (feature engeneering), ,
.
2.
50 .txt.
. 10 .
. 10 .
( ) python 3.8.10.
, , .
, .
: (:)
: 1. ,
2. ,
,
: (max_force),
(max_lenght), , (max_time),
(section_area), is_valid, 1 ,
, 0 .

:

[3]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
import os

%matplotlib inline
```

- Pandas -
- Nympy
- matplotlib
- seaborn
- os -

- %matplotlib inline

,

Pandas.

[7]:
```
experiment_data = pd.DataFrame({
        'max_force':[],
        'max_lenght':[],
        'max_time':[],
        'section_area':pd.Series([], dtype=np.dtype('int32')),
        'is_valid': pd.Series([], dtype=np.dtype('bool_')),
    })
```

whitelist, , .

[6]:
```
whitelist = ['3-2', '3-3', '3-5', '3-9', '3-10', '4-2', '5-6', '5-7', '5-10',␣
↪'6-1', '6-3', '6-8', '6-9', '7-1', '7-3', '7-7', '7-8', '7-9']
```

. .

data, .

experiment_data. 50 experiment_data

CSV - Comma-Separated Values — , .

, , .

[9]:
```
for file in sorted(os.listdir('Data_samples/')):

    column_names = ['time', 'force', 'lenght']

    data = pd.read_csv(f'Data_samples/{file}', sep='\t', skiprows=17,␣
↪decimal=',', names = column_names)

    name = file.split('.')[0]
    section_area = int(file[0]) * 10
    max_force = data.force.max()
    max_lenght = data.loc[data.index[data.force.idxmax()], 'lenght']
    max_time = data.loc[data.index[data.force.idxmax()], 'time']

    if name in whitelist:
        is_valid = True
    else:
        is_valid = False

    new_value = {
        'max_force': max_force,
        'max_lenght':max_lenght,
        'max_time':max_time,
        'section_area':section_area,
        'is_valid':is_valid
    }
```

```
    experiment_data.loc[f'{name}'] = new_value

experiment_data.to_csv('experiment_data.csv')
```

[ ]: