

Microstructure - Unsupervised Learning

MicroClustering

Manual

ClusterUI is an unsupervised learning web tool that clusters microstructures in Steel based on their physical properties. Files can be uploaded in xlsx or csv formats. The tool supports difference clustering algorithms which can be used also in combination with dimensionality reduction. The dimensionality reduction algorithm used here is Principal Component Analysis (PCA).

This web tool is developed as part of the coursework - Software Engineering, for the Chair for Functional Materials, Saarland University.

More information about the department can be found [here](#).

Setting up the environment

To build ClusterUI you need Python 3.8. The procedure is described below:

1. Clone the directory from Github repository

```
git clone https://github.com/microclustering/project_25
```

Or download it to your machine. You can also use the source files that have been provided.

2. Install dependencies (that are listed in the following section of this document) by running the following command:

```
pip install -r requirements.txt
```

For this step you need *pip* to be installed on your machine.

Launching the Webtool

a) Running from Command Line

To run the application from command line, give the following command in the same order in command prompt.

```
python app.py
```

The application will start Default Browser case it did not open automatically, you can manually open your internet explorer and got to the link which is shown in the command line.

b) Creating and running batch script

1. Navigate to the scripts folder.

```
cd project_25
```

2. Check the python path in your device and replace it in app.bat file as

```
"Absolute python path (python.exe)" "Path where app.py python script is"
```

To check python path use `where python` in command prompt under windows. Also add the absolute path of python script app.py as saved on your device.

3. After replacing the path as above, save the file as bat file

```
app.bat
```

4. Launch the application app.bat.

c) Creating a shortcut for the entire application to desktop

To create a shortcut for the application instead of running the application from the command line

1. Navigate to the `project_25` folder in your local system
2. Right- Click on the `app.bat`. In the pop up menu select send to *Desktop (create shortcut)*
3. On the desktop right-click on the newly created shortcut and select rename.

4. Give the name as “*ClusterUI*”
5. Right-Click the shortcut and go to *Properties*. Click the *Change Icon* option.
6. In *Change Icon* Pop-Up window, select *Browse* and browse to the *Project25*(the project



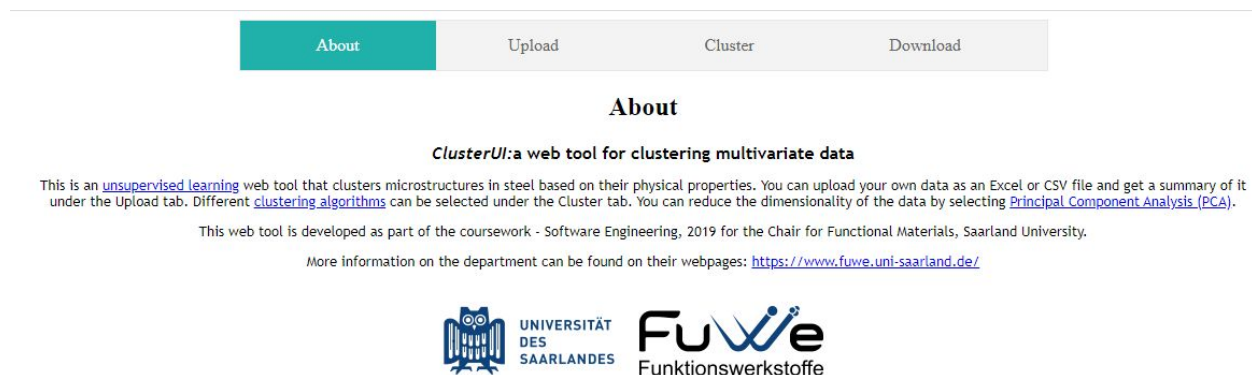
- folder). In the Folder is the icon for *clusterui* **ClusterUI** .Select it and click *Open*.
7. Click *ok* in the *Change Icon* window and click on *Apply* in the *Properties* Window.
 8. The Name and Icon of the shortcut is changed and now the application can run directly from the Desktop

Using the Webtool ClusterUI

You can navigate through the webtool and check all feature by clicking on them. In the following we explain how you can run a clustering algorithm on you own data.

After launching the application `app.bat` or through the icon or the terminal directly

1. The application opens in the `default` browser.
2. On the first page an Introduction is given for the webtool under the *About* tab.



3. Select the *Upload* tab to upload your dataset file in `.xlsx` or `.csv` formats. *Please notice that the first 5 columns of the uploaded dataset are considered to be reserved for the additional information not relevant to the clustering process.* Therefore they will be dropped and not used in clustering. It is better to keep the first column as the object number, indicating the tag for the data point. All other columns are input features (i.e. variables) used for clustering.

When uploading is done, a message “*Successfully uploaded <File Name>*” will be indicated. A summary of the uploaded dataset can be shown by clicking on *Get Summary* button. (Uploading might take some time if the file uploaded is `.xlsx`). You can scroll to the right to view all the columns. It displays the first five rows of the dataset.

Upload your file here in **xlsx/.csv** format

No file chosen

Upload your file here in **xlsx/.csv** format

No file chosen

	N ^o of object	Sample	Class	Border object	Grey value	Equiv. Diameter	Major axis length	Minor axis length	Perimeter	Equiv. radius	Max Feret diameter	Min Feret diameter	Mean Feret diameter	Convex perimeter	Area	Convex area	Filled area	Total subarea	Relative area	Subarea/Grain area	Convex area/Filled area	Axial ratio	Aspect ratio	Roundness	Circularity
0	5.0	1318 _net_Rag_0.25	1.0	1.0	255.0	1.710221	1.318887	0.743995	7.208316	0.855111	2.562406	1.806993	2.184699	7.011089	2.297177	3.166505	2.297177	0.000471	0.000213	0.000205	1.378433	0.564108	0.705194	0.445460	0.667428
1	6.0	1318 _net_Rag_0.25	1.0	1.0	255.0	8.206043	6.359600	3.750442	52.406886	4.103022	13.303000	7.742576	10.522788	33.613412	52.888039	72.140236	52.888818	0.010161	0.004913	0.000192	1.363998	0.589729	0.582017	0.380518	0.616861
2	7.0	1318 _net_Rag_0.25	1.0	1.0	255.0	26.069428	32.686754	6.286851	203.081980	13.034714	73.871248	15.789208	44.830228	156.527985	533.768416	858.712622	533.769195	0.108100	0.049585	0.000203	1.608771	0.192336	0.213740	0.124541	0.352904
3	8.0	1318 _net_Rag_0.25	1.0	1.0	255.0	6.755097	7.066641	2.012306	35.750645	3.377548	12.760737	5.317679	9.039208	30.779244	35.838764	52.670707	35.838764	0.007926	0.003329	0.000221	1.469657	0.284761	0.416722	0.280228	0.529366
4	9.0	1318 _net_Rag_0.25	1.0	1.0	255.0	2.453631	2.565155	0.765414	13.355075	1.226815	5.001500	1.883948	3.442724	11.479428	4.728336	6.959301	4.728336	0.001045	0.000439	0.000221	1.471829	0.298389	0.376677	0.240668	0.490579

3. After the file is uploaded, navigate to the *Cluster* tab and select any clustering algorithm from the list. The maximum number of clusters that can be chosen is 50. The default number of clusters is set to 2, the default clustering algorithm is set to k_means and by default Dimensionality Reduction is unchecked. If the number of features is huge, you can use the feature dimensionality reduction by checking the box for PCA algorithm.

Clustering Algorithms

Select the clustering algorithm from the drop-down list:

- k_means
- k_medoids
- hierarchical clustering
- fuzzy_clustering
- gmm_clustering

Number of clusters (between 1 and 50):

☐ Dimensionality Reduction

The clustering process can take some time based on the size of dataset and your computer specifications.


[About](#) [Upload](#) [Cluster](#) [Download](#)

Clustering Algorithms

Select the clustering algorithm from the drop-down list:

Number of clusters (between 1 and 50):

☒ Dimensionality Reduction



A circular loading spinner with the word "LOADING" in the center, surrounded by a ring of dots.

4. A visual feedback (hyperlink) with the message *"Your file has been processed using <Selected Algorithm> is ready for downloading"* informs about the end of clustering process. You can click on it to navigate to the download page by clicking on the highlighted text.

Clustering Algorithms

Select the clustering algorithm from the drop-down list:

k_means

▼

Number of clusters (between 1 and 50):

2

☐ Dimensionality Reduction

Submit

[Your file has been processed using k_means and is ready for downloading](#)

5. On the *Download* page, a summary of clustering results can displayed using *Get Summary*. Summary can be downloaded using *Download Summary(.txt)* button. Also, the results can be saved in .csv or .xlsx format locally on the basis of the file format of the file that was uploaded.

Download option goes here

Get Summary

The clustering algorithm used is k_means
 The input file have 11171 observations and 111 features
 Removed 1 observations and 0 features after preprocessing
 Dimensionality reduction is enabled : 111 features reduced to 18 features
 18 components capture 0.950000 amount of variance in the data.

Cluster#	Number of observations
0	980
1	3443
2	5112
3	1635


 Download Summary(.txt)


 Download Clustered File(.xlsx)

Running the Application

- Start the application by clicking on the application shortcut
- The application opens in the default browser.
- Select the Upload option
- On the Upload page, select input file with .x/xs or .csv format and click **Submit**
- The Upload Page also gives option of **Get Summary** which gives a summary of the dataset uploaded to application
- Click on Cluster Page Option on the Main Menu.
- Select the clustering algorithm name, number of clusters.
- Check dimensionality Reduction option to reduce the dimension of the uploaded dataset if required.
- Select **Submit**. While the algorithm runs on the page displays loading and the Terminate Clustering option which will stop the clustering.
- After the clustering is done the download option appears on the screen highlighted in blue. Clicking on this option, redirects to the download page.
- The Download page gives the option to display the summary which describes the number of clusters and the features in the form of a table
- This summary can be downloaded as a .txt file by clicking on the option Download Summary as txt file
- The clustered file can be downloaded as output.csv file into the local PC.

License

This code is released under the terms of the [MIT license]
(<https://github.com/microclustering/project25/blob/master/LICENSE>).

Dependencies required for Python

The following Python libraries are necessary for the code to run:

```
altgraph==0.17
Click==7.0
decorator==4.4.1
et-xmlfile==1.0.1
Flask==1.1.1
future==0.18.2
itsdangerous==1.1.0
jdcal==1.4.1
Jinja2==2.10.3
joblib==0.14.1
MarkupSafe==1.1.1
```

```
networkx==2.4
numpy==1.18.1
openpyxl==3.0.3
pandas==0.25.3
pefile==2019.4.18
prettytable==0.7.2
PyInstaller==3.6
python-dateutil==2.8.1
pytz==2019.3
pywin32-ctypes==0.2.0
scikit-fuzzy==0.4.2
scikit-learn==0.22.1
scipy==1.4.1
six==1.14.0
sklearn==0.0
Werkzeug==0.16.0
xlrd==1.2.0
```