**Faculty of**

**computers and**

**artificial intelligence**

Team num:

# Cover sheet

**Selected CS-1 project.**

**Team no.:**

|  |  |  |
| --- | --- | --- |
| Name | ID | Level |
| **احمد خالد محمد عبدالسلام** | **20210048** | **Three** |
| **احمد عاطف صلاح مجاهد** | **20210067** | **Three** |
| **احمد السيد السيد محمود** | **20210026** | **Three** |
| **احمد علي سالمان** | **20210082** | **Three** |
| **ندي محمد عبدالرحمن احمد** | **20211000** | **Three** |
| **اميرة حسن صبحس** | **20210185** | **Three** |

# Numerical dataset

**General information about dataset**

|  |  |
| --- | --- |
| Name | Power Plant |
| No. of classes | Dataset Regression |
| Total no. of samples | 9568 |
| No. of samples in training\validation | 7654 |
| No. of samples in testing | 1914 |
| Num of Feature | 4 |

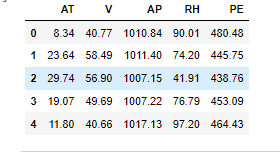
***Dataset description (Power Plant):***

The dataset contains 9568 data points collected from a Combined Cycle Power Plant over 6 years (2006-2011), when the power plant was set to work with full load. Features consist of hourly average ambient variables Temperature (T), Ambient Pressure (AP), Relative Humidity (RH) and Exhaust Vacuum (V) to predict the net hourly electrical energy output (EP) of the plant.

A combined cycle power plant (CCPP) is composed of gas turbines (GT), steam turbines (ST) and heat recovery steam generators. In a CCPP, the electricity is generated by gas and steam turbines, which are combined in one cycle, and is transferred from one turbine to another. While the Vacuum is collected from and has effect on the Steam Turbine, the other three of the ambient variables effect the GT performance.

For comparability with our baseline studies, and to allow 5x2 fold statistical tests be carried out, we provide the data shuffled five times. For each shuffling 2-fold CV is carried out and the resulting 10 measurements are used for statistical testing.

We provide the data both in. odds and in .xlsx formats.



## Linear Regression

Goal: Predict the net hourly electrical energy output (EP)

Code: used linear regression from Sklearn.linear\_model

Cross Validation: k-fold Cross Validation with K = 5

**Implementation details:**

We use Validation*the number of fold and ratio of training/validation.* = 5

**Cross validation**

**Hyperparameters**

**n\_splits=5**

A screenshot of a computer code

Description automatically generated

A white rectangular object with a black border

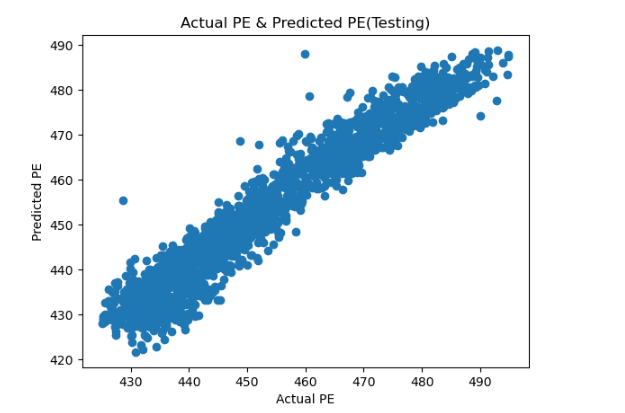
Description automatically generated

***Correlation***

A red and blue squares with numbers

Description automatically generated

***Predict for linear regression:***



***Predict for KNN:***

A white background with black and white text

Description automatically generated

A blue dotted graph with numbers and lines

Description automatically generated

***Histograms:***

A graph of a bar graph

Description automatically generated with medium confidence

**Image dataset**

**General information about dataset**

|  |  |
| --- | --- |
| Name | 102 Categories Flowers |
| No. of classes | 3 |
| Total no. of samples | 597 |
| Size of image | 620 \* 480 |
| No. of samples in training\validation | 417 |
| No. of samples in testing | 180 |

***Dataset description (Flowers):***

We have created a 102-category dataset, consisting of 102 flower categories. The flowers chosen to be flower commonly occurring in the United Kingdom. Each class consists of between 40 and 258 images. The details of the categories and the number of images for each class can be found on this [category statistics page](https://www.robots.ox.ac.uk/~vgg/data/flowers/102/categories.html).

The images have large scale, pose and light variations. In addition, there are categories that have large variations within the category and several very similar categories. The dataset is visualized using isomer with shape and color features.

We visualize the categories in the dataset using SIFT features as shape descriptors and HSV as color descriptor. The images are randomly sampled from the category.

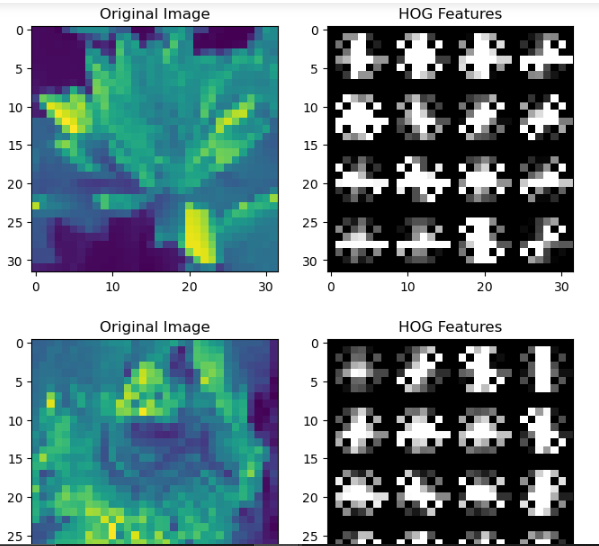
***Implementation details:***

**Features extraction**

No. of features extracted per image = 597

Dimension of resulted features = 2

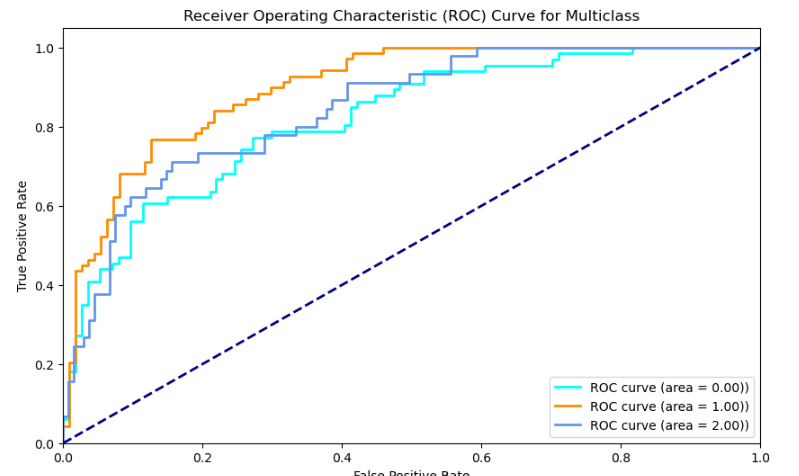
***HOG***



A screenshot of a computer program

Description automatically generated

**ROC**



**confusion matrix**

A blue squares with black text

Description automatically generated

**K-Means**

**Un supervised.**

A blurry image of a white circle

Description automatically generated

A white rectangular object with black text

Description automatically generated