Assignment 3: Machine Learning as a Service in Microsoft Azure ML

**Under guidance of: Team 1**

Prof. Sri Krishnamurthy Amulya Aankul

Ashwin Dinoriya Nand Govind Modani

Saksham Agrawal

**Step 1: Data Preparation in R**

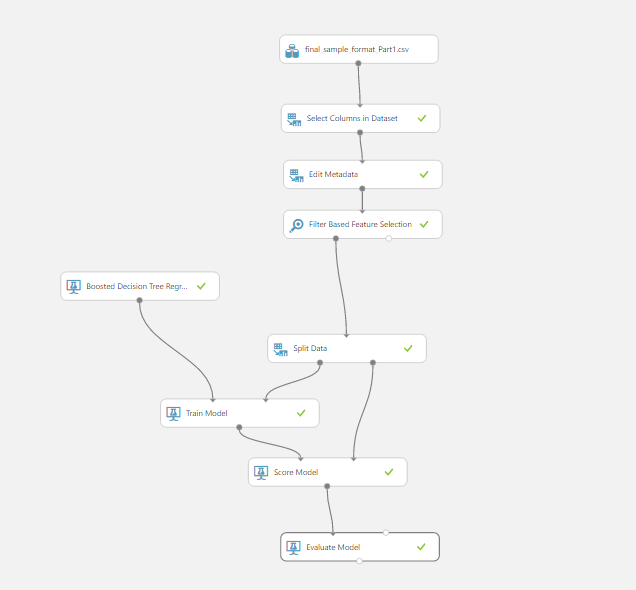
The first step is preparation of data. We have followed the same procedures as previous assignments for pre-processing and wrangling of data in R. To summarize:

* API calls for weather data and holiday data
* Pre-processing of data
* Imputation of missing values
* Removal of outliers
* The clean data from R is uploaded to Microsoft Azure ML Studio datasets

**Step 2: Microsoft Azure ML Studio – Building model**

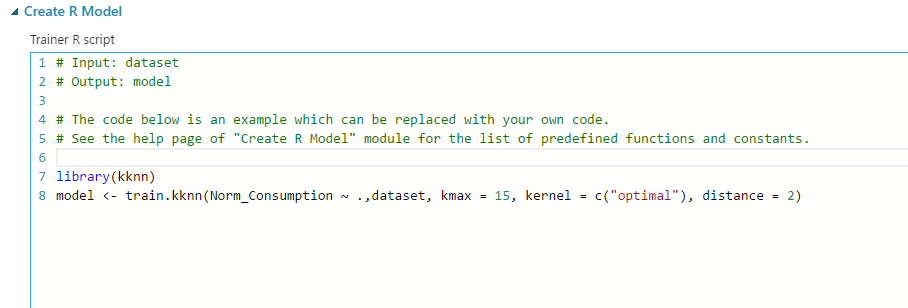
The following are the steps to create an experiment in Microsoft Azure ML Studio:

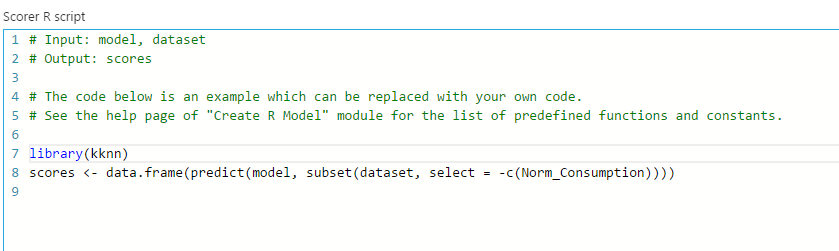
* Select the required column using **Column Selector**
* Specify the data types and categorical variables using **Edit Metadata**
* Selection of features using **Filter based Feature Selection**
* Splitting data into Training and Testing using **Split**
* Training of model using **Train Model**
* Scoring and Evaluation of model using **Score Model** and **Evaluate Model**

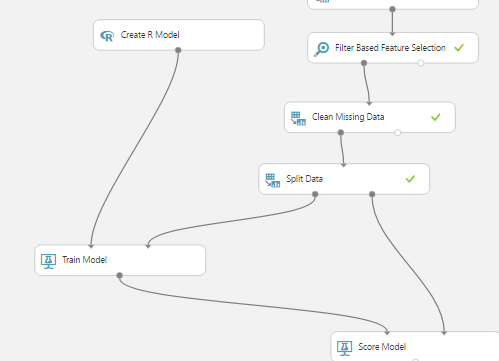


**Step 3: Building customized model in R**

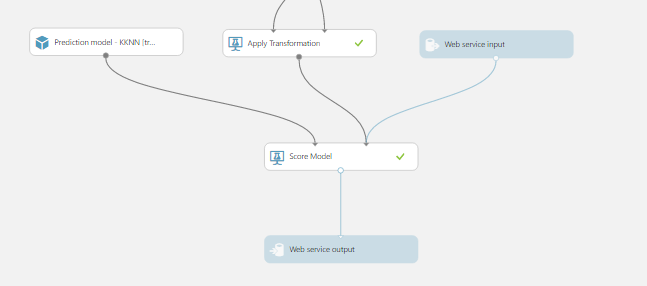
The following are the steps to create a model in R in Microsoft Azure ML Studio. We noticed that KNN and Hierarchal Clustering are not pre-present in Microsoft Azure. To implement that, we had to use **Create R model** feature in Microsoft Azure ML. It has a Trainer script and Scorer script:

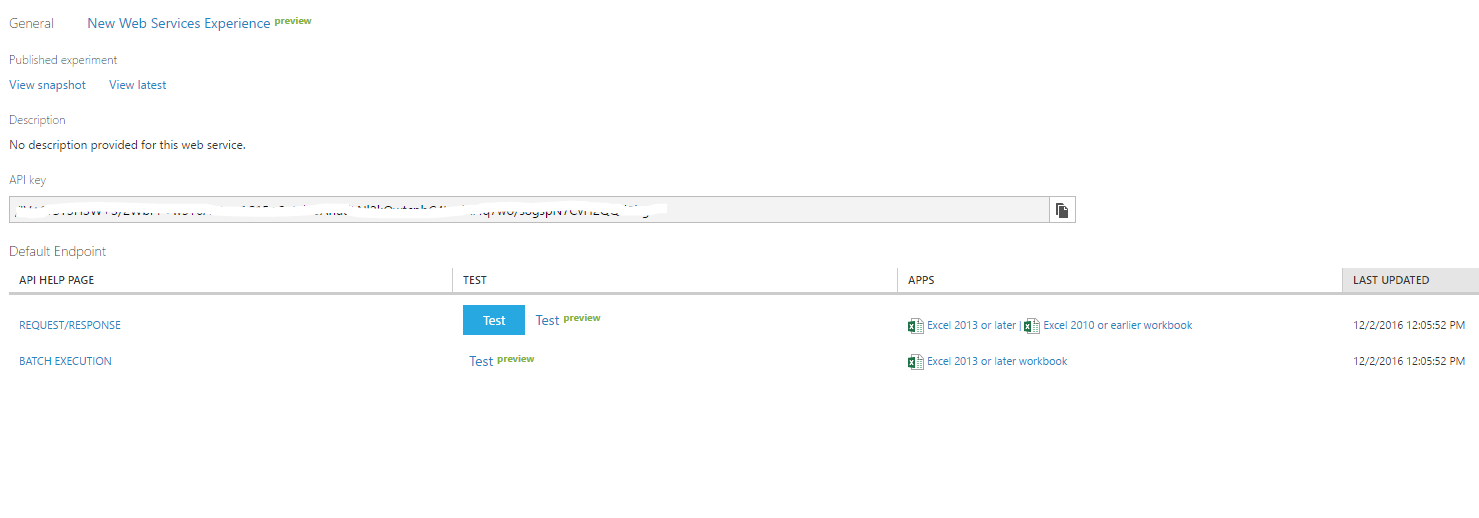






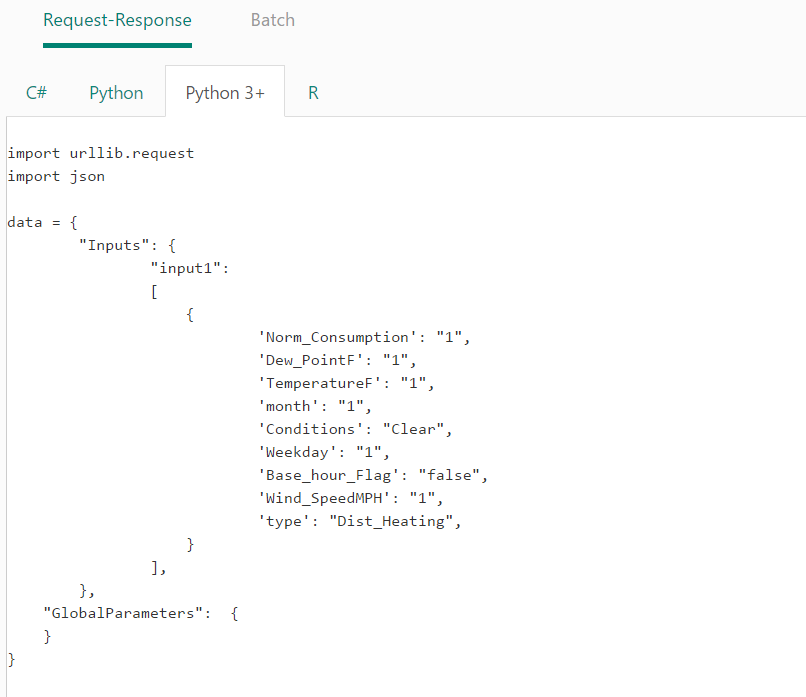
**Step 4: Set up and Deploy Web Service**

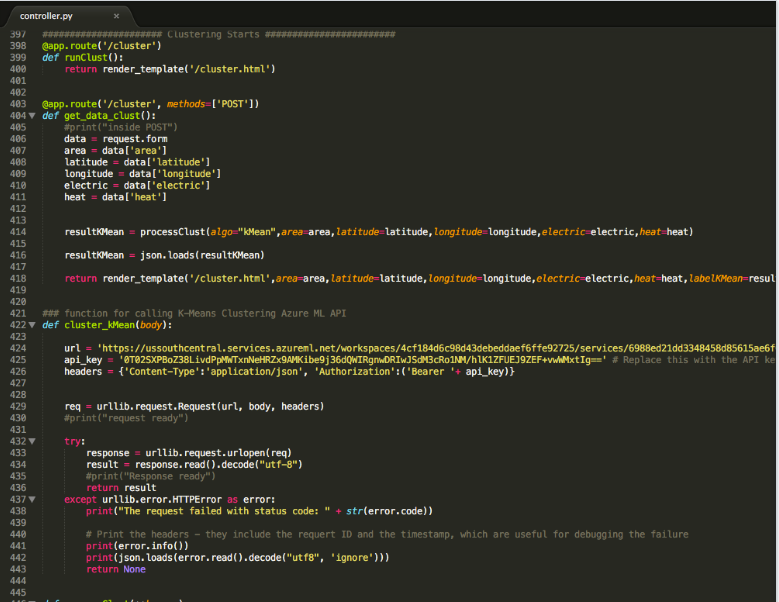


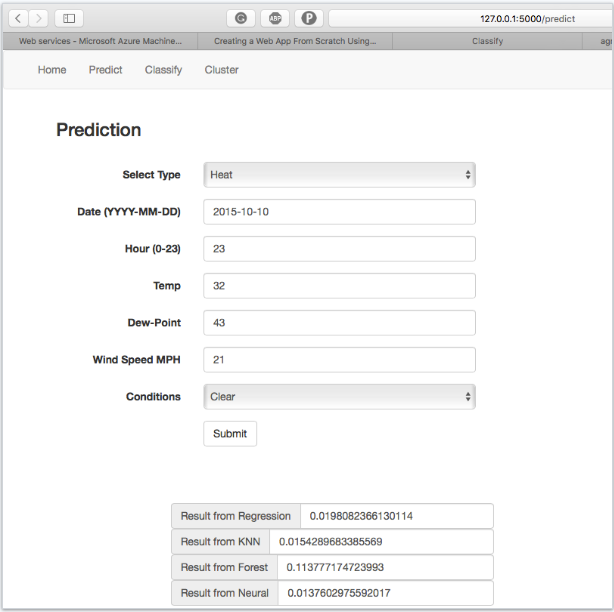
****

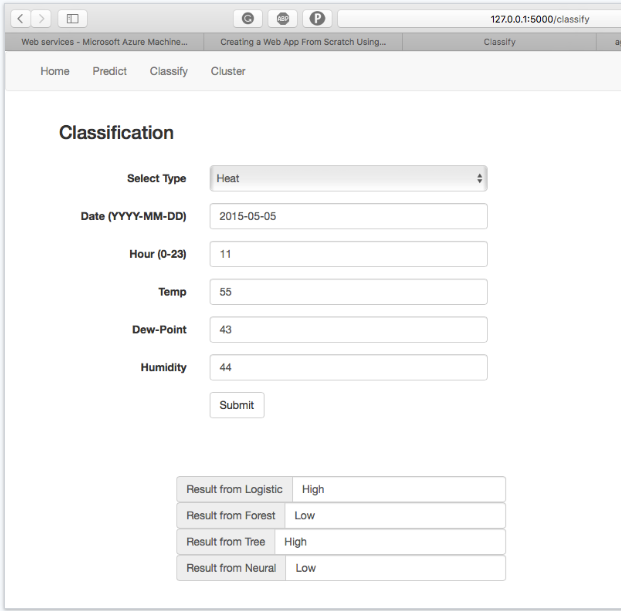
**Step 5: Building a website**

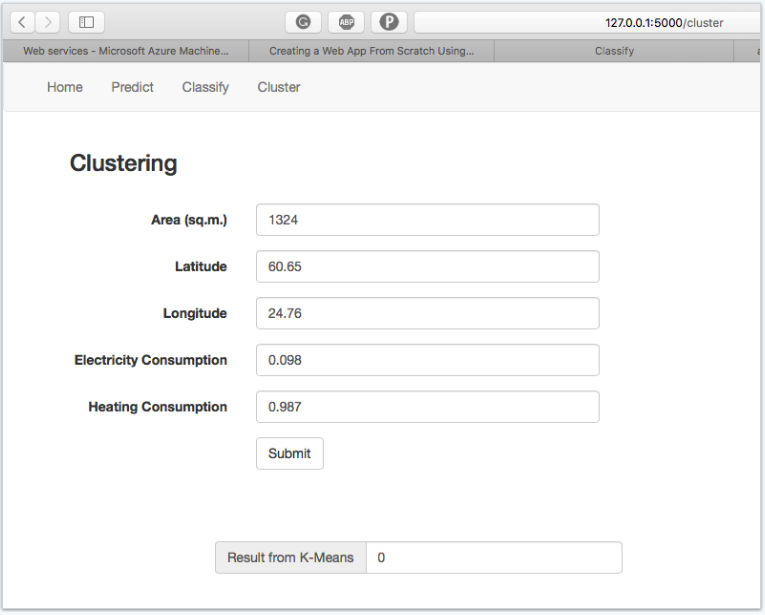
We used HTML5, bootstrap and in the backend we have used Flask framework in Python 3+.





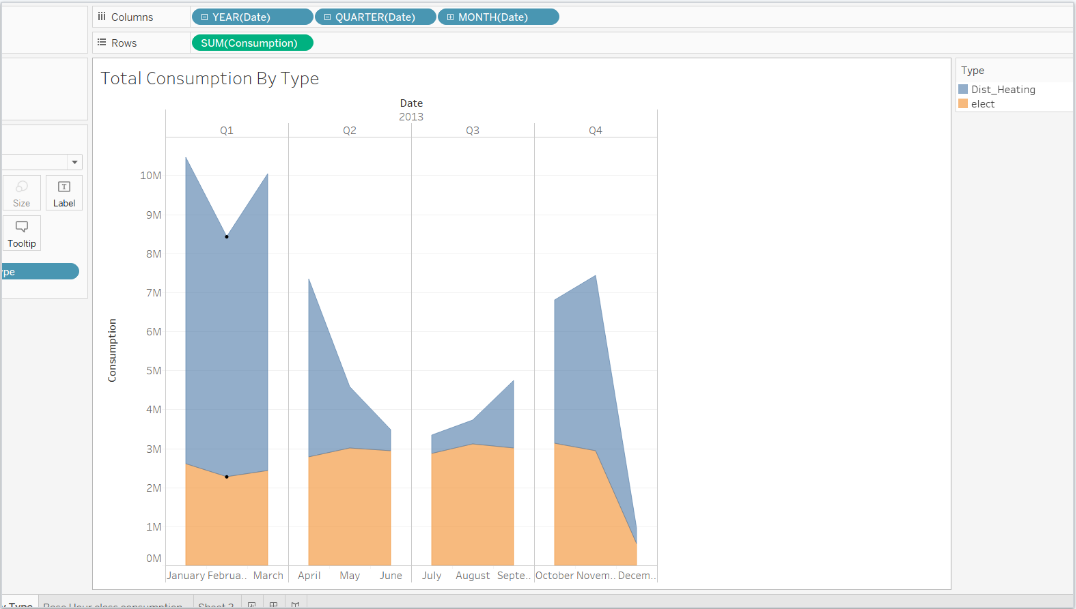




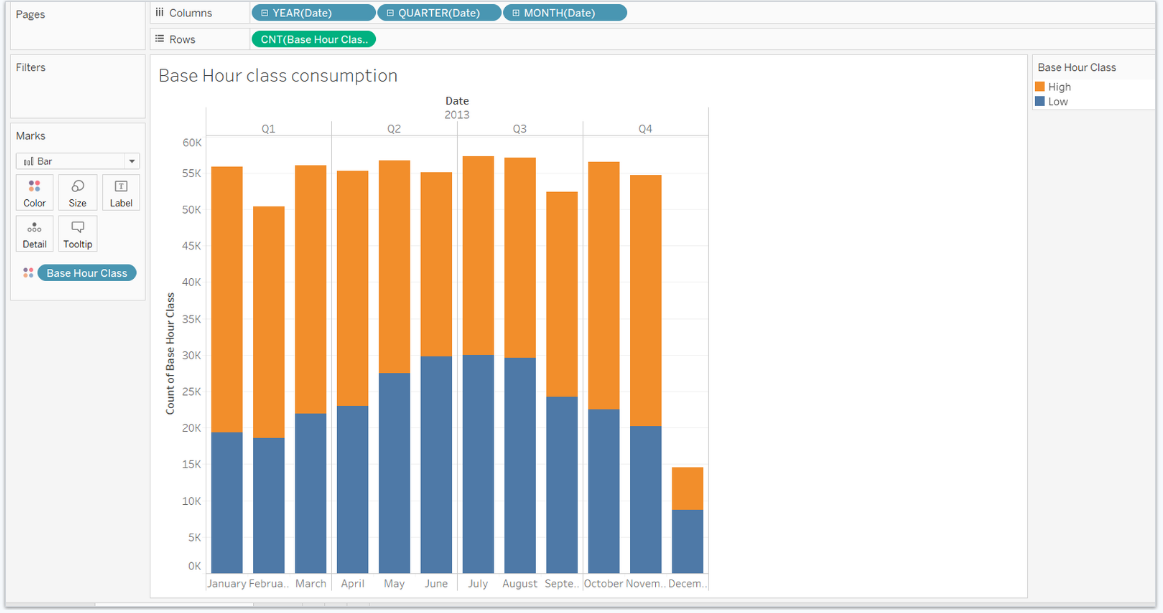




**Step 6: Data Visualization**

****

* The Plot shows the total Energy consumption by type for entire year. This gives the consumption distribution for year.

****

* The base hour class is used as a classified label for classification. The above plot shows the exploratory data analysis for count of base hour class values per month.