**Clustering on Turkiye Student Evaluation Data Set**

**Abstract:**

Our goal in this project is to group the students based on the similarity of their answers on the survey. Notice that we don’t know how many cluster (group) of students will be. In fact, we will use different methods of clustering to decide the best “natural” number of group of this dataset. We will attempt to perform k-means clustering technique to monitor and assess the student performance and behavior as well as give improvement toward e-learning system in the future. The challenge in this project is that we do not have the labelled data and our algorithm must be able to cluster such that intra cluster similarity must be high and intra cluster similarity must be low. This dataset is based on an evaluation form filled out by students for different courses. It has different attributes including attendance, difficulty, score for each evaluation question, among others. The dataset has 5820 rows and 33 columns.

**Existing System:**

The existing system does not provide a way of grouping customers and hence identifying natural clusters is difficult.

**Disadvantages of Existing System:**

The limitations of available systems are not sufficient to deal with the complex data. In this section, we present some of the limitations that are present in the existing system.

* The system uses DBMS and hence can return records based on the filters.
* The system also requires data extensive data preprocessing and Exploratory Data Analysis(EDA) inorder to perform feature engineering.

**Proposed System:**

We aim to implement K-Means, Hierarchical clustering and others and also fine tune the parameters of the model. These models would be trained on a data set which will be engineered carefully after performing the feature engineering.

**Advantages:**

* Load and explore the dataset and generate ideas for data preparation and model selection.
* Perform Exploratory Data Analysis to find correlations.
* Visualize clusters produced by the algorithms

**Software requirements:**

Operating System : Windows 7 , Windows 8, (or higher versions)

Language : Python 3.5 and other libraries likes numpy, pandas, matplotlib, seaborn and scikitlearn.

Mozilla Firefox(or any browser)

**Hardware requirements:**

Processor : Pentium 3,Pentium 4 and higher

RAM : 2GB/4GB RAM and higher

Hard disk : 40GB and higher