**Machine learning**

* Purpose of project

The team target is to analyze new datasets provided by client and decide which machine learning method is suitable for use

* Background Information and Relevant Research

The team has to figure out how many machine learning methods would be provided in our website and their respective principles and how to use them.

There is many Test suitability of open source machine learning toolkits such as H2o, weka and python/R packages. These machine learning toolkits could be used to analyze different datasets. In this stage, the team prefer to use weka to understand each machine learning tools.

When study the weka, the team find out there are many types of dataset machine learning tool could use, for instance Arff data files, CSV data files XRFF data files etc.

After some researches on the UCI Machine Learning Repository [1], the team decide to use CSV data files to store datasets into database and be used by machine learning tools. Because the CSV file is convenient for the team to transform from the .data file and .name file. The CSV files are also easier to be stored in NoSQL database.

* Initial Design Ideas

The final goal of the project is to analyze datasets and automatically suggest the best machine learning tools by using our machine learning analyzer. The Short-term goal is to give a suggestion of machine learning tools directly by not using any analyzer. The team decide to provide ten machine learning tools: BayesNet, NaiveBayes, Logistic, SMO, Kstar, IBK, IterativeClassifierOptimizer, InputMappedClassifier, DecisionTable and HoeffdingTree.

The website will provide the most suitable machine learning tool for each dataset uploaded by client. Besides, the website will analyze the dataset by using the best machine learning tool and visualize the output.

**Reference list**

[1] http://archive.ics.uci.edu/ml/datasets.html