Data mining is about explaining the past and predicting the future by analyzing the data. It is a multi-disciplinary field which combines statistics, artificial intelligence, machine learning and database technology. Data mining predicts the future by the means of modelling. Predictive modelling is a process where in which an outcome is predicted by creating a model. If the outcome is categorical, it is called *classification* and if the outcome is numerical, it is called *regression*. Classification is a data mining task of predicting the values of a categorical variable by building a model based on one or more categorical variables. *Association rules* find interesting associations among observations.

The following are the groups of classification algorithms.

Frequency table, covariance matrix, similarity functions and others.

*K Nearest Neighbors* algorithm falls under the *similarity* *functions* category.

Association rules find all sets of items that have a support greater than the minimum support and using large sets to get desired rules that have confidence greater than the minimum confidence.

The following are the algorithms I am using in the project.

Bootstrap algorithm

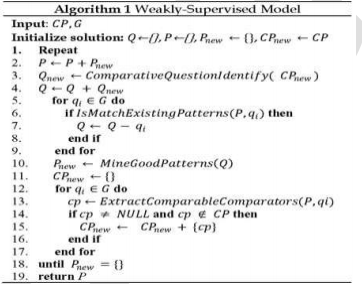
*K Nearest Neighbors* algorithm

AIS algorithm for association rules

Let us discuss now, how I have used each algorithm in the project.

Bootstrap algorithm.

Bootstrap algorithm is a process where in which we extract comparators from comparative questions.



The above picture depicts the pseudocode of the bootstrapping algorithm

Comparator mining

A weakly supervised method is used for mining comparators form comparative questions.

In this method a sequential pattern is defined as a sequence s (S1 S2 S3…. Si…. Sn) where Si can be a word or a POS tag or a symbol denoting either a comparator ($C) or the beginning (#start) or the end of a question (#end). A sequential pattern is called an indicative extraction pattern (IEP) if it can be used to identify comparative questions and extract comparators.