# Which data needs to be collected & from where

### i) For individuals

| **Data** | **Source** |
| --- | --- |
| Personal Information | Government Database |
| Credit score/ rating | Banks |
| Behavioural nature | Workplace |
| Crime records | Government Database |

### ii) For companies

| **Data** | **Source** |
| --- | --- |
| Company Information | Government Database |
| Business credit score | Banks |
| Trustworthiness | Employees and Partners |
| Ratings from suppliers and customers | Company Database |

# Algorithms to be used for mining this data & integrating the results

### **C4.5 Algorithm**

C4.5 is used to generate a classifier in the form of a decision tree from a set of data that has already been classified. Classifier here refers to a data mining tool that takes data that we need to classify and tries to predict the class of new data.

Every data point will have its own attributes. The decision tree created by C4.5 poses a question about the value of an attribute and depending on those values, the new data gets classified. The training dataset is labelled with lasses making C4.5 a supervised learning algorithm

### **K-mean Algorithm**

One of the most common clustering algorithms, k-means works by creating a k number of groups from a set of objects based on the similarity between objects. It may not be guaranteed that group members will be exactly similar, but group members will be more similar as compared to non-group members. As per standard implementations, k-means is an unsupervised learning algorithm as it learns the cluster on its own without any external information.

### **Apriori Algorithm**

Apriori algorithm works by learning association rules. Association rules are a data mining technique that is used for learning correlations between variables in a database. Once the association rules are learned, it is applied to a database containing a large number of transactions. Apriori algorithm is used for discovering interesting patterns and mutual relationships and hence is treated as an unsupervised learning approach. Thought the algorithm is highly efficient, it consumes a lot of memory, utilizes a lot of disk space and takes a lot of time.

### **Naive Bayes Algorithm**

Naive Bayes is not a single algorithm though it can be seen working efficiently as a single algorithm. Naive Bayes is a bunch of classification algorithms put together. The assumption used by the family of algorithms is that every feature of the data being classified is independent of all other features that are given in the class. Naive Bayes is provided with a labelled training dataset to construct the tables. So it is treated as a supervised learning algorithm.

### **CART Algorithm**

CART stands for classification and regression trees. It is a decision tree learning algorithm that gives either regression or classification trees as an output. In CART, the decision tree nodes will have precisely 2 branches. Just like C4.5, CART is also a classifier. The regression or classification tree model is constructed by using labelled training dataset provided by the user. Hence it is treated as a supervised learning technique

# Actions should result in a positive score and a negative score.

**• Behaviour**

Human Behaviour is a very important factor and it determines how the mind of the person behaves.

**• Trustworthiness**

Loyalty is rarer than anything else so trust factor is an important deciding factor

**• Ratings from suppliers and customers**

Every company has a rating and in layman terms, that rating represents the value of the company.

# How should people be rewarded

• Discounts on energy bills

• Rent things without deposits

• Better interest rates at banks

• Will reportedly get more matches on dating websites.

• Streamlined administrative procedures.

• Fewer inspections and audits

• Fast-tracked approvals.

# How should people be penalized

• Banning you from flying or getting on the train.

• Throttling internet speeds.

• Banning from the best schools.

• Stopping from getting the best jobs.

• Keeping out of the best hotels.

• Being publicly named as a bad citizen.

# Where will this data be stored