Predicting Employee Attrition using Machine Learning Techniques

Problem Statement:

Employees are the most important part of an organization. Successful employees meet deadlines, make sales, and build the brand through positive customer interactions.

Employee attrition is a major cost to an organization and predicting such attritions is the most important requirement of the Human Resources department in many organizations. In this problem, the main task is to predict the attrition rate of employees of an organization.

Objectives:

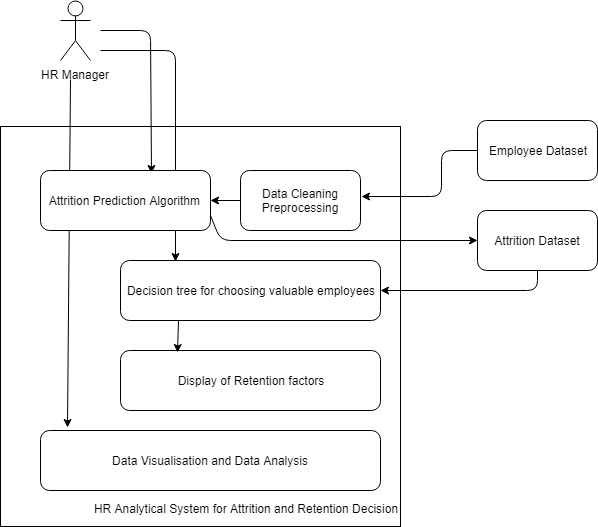
This project aims at finding the attributes responsible for employee attrition and then predict the employee attrition with the data mining technique. Once the attrition is found out, study on the factors deciding the valuable employees and after finalizing those factors, build the decision model for valuable employees.

I am finding the attributes for employee attrition with the help of previous study and research on prediction of employee attrition. The most accurate prediction model was developed. I built the decision model for the valuable employees. Then again with other assumptions I found out the factors most effective for attrition and displayed them in the file.

Relevant Technical support associated with the project

System Description:

I will describe the system with the help of a diagram. The system diagram below, shows the system design of an application which illustrates how the raw data is provided to the system and then retrieved as an output and again fed to the system as a second input and get the final report (dashboard result in this case).



Input: The input to the model was given in csv format which contains employee details

Output: The output of the model will be prediction of attrition rate and is that

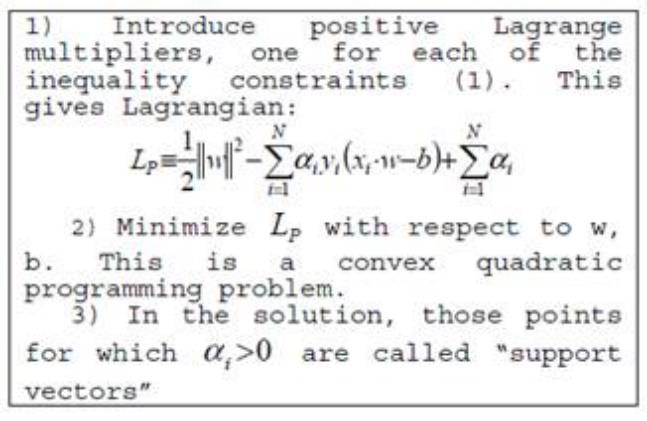
Employees are valuable to the company or not.

Mathematical Formulation

Our model is divided into four main parts in which two parts require mathematical formulation the models which will being

Used has specific calculations that differ from algorithm to algorithm.

Support vector machine:

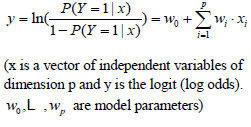


Logistic Regression

In this calculation, the probabilities portraying the possible results of a separate trial are displayed utilizing a Logistic Regression



Logistic regression is a regression model that fits the values to the logistic function. It is useful when the dependent variable is categorical



So its general formula can be determined as



K-Nearest Neighbour:Neighbors based order is a kind of instance-based learning. Characterization is figured from a basic greater part vote of the k closest neighbors of each point

So formulation is 𝐷 = Σ |𝑥𝑖 − 𝑦𝑖 |

Decision Tree:

Every node in tree can be calculated by gain which is calculated by

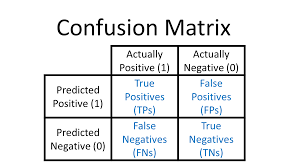


Where E(A) is entropy which can be calculated by using



## Evaluation matrix:

This matrix will entirely determine success and failure of framework. This matrix has four entries. As shown in figure which entirely decides how many predicted values are correctly predicted and which values have been not predicted accurately.



Usefulness:

This model is helpful while making the following decisions:

* Evaluation of employee requirements, their strengths and weaknesses
* Minimize cost of new talent acquisition based on the employee profiling and company requirements
* Analysis and assessment of the loss in expertise and skillsets
* Measurement of financial and productivity loss due to attrition
* Able to plan and minimize the loss
* Provides good understanding of workforce supply and demand
* Able to prepare contingency plans based on the insight and foresight provided by the model

Monthly planner chart

