Introduction

Employee attrition is when an employee of an organisation, leaves the company due to various reasons. It may be because his or her performance was not up to the mark. Or due to incompatibility with their colleagues and more so on. If the attrition happens with the will of the organisation than it is supposed to be fine. But if an organisation loses its valuable employee because the employee wanted it so, then it is a problem. The organisation might have gone through a long process of hiring and training of that employee. Both time and money of the organisation is invested on a particular employee, and when this employee leaves than it is a matter of great concern for the employer. The organisation will have to do all that investments of time and resource again to hire another employee to replace the one that left. This problem can be analysed and to a great extent be solved by using machine learning algorithms. To do such a task, the most preliminary tool required to get started is the past data of the employees who left the company along with the ones who stayed. Machine learning models learn from the past data, or we might say, past experiences to predict the future, so if we feed this data consisting of employees that stayed and the employees that left, we might get a model that takes the employee data as input and tells if the employee is going to leave or not. Further analysis of this model will give us insights of what are the plausible factors leading to attrition of employees. This will help the management to take steps in the right direction.

Machine Learning

The procedure of using computational methods to learn information directly from the data. This learning should be done without relying too much on a predetermined equation. This whole process is called Machine learning. The data samples used to train the model is very important. When the data models increase, the accuracy of this model might also increase. Since the process of Machine learning largely involves a lot of statistical computations, the data used to train such a model needs to be clean and relevant to the intended goals of the project. The data used should be consistent, that is, it should not have too many null values and understandingly the tabulated data should be correct. The relevance of the data used is important because if the data is not relevant than the model will predict wrongly for real world cases. For example, if we need a model to predict a disease in an individual, we need the medical data of that person rather than the data about his TV watching habits. The TV habits data can be useful for the model where subscription to a TV plan is predicted. There are many machine learning algorithms that use different statistical computation techniques to train a model. We have used some of these models during this research.