Algorithms

Multiple machine learning algorithms were used for this research, K- Nearest Neighbour and Multivariate Logistic Regression algorithms were used to predict the likelihood of the students getting admission into university based on their profile. Decision Tree algorithm was used to predict the rank of the college that would be suitable for the students based on their profile and suggest the list of universities accordingly.

K-NearestNeighbours: It is an algorithm which is used widely for classification and regression problems. Due to its simplicity and effectiveness, it is easy to implement and understand. It is a supervised machine learning algorithm that uses available data to create the model and further that model can be applied to classify the new data. The class of new data is determined by the class of its neighbours. Distance is calculated between the unseen data sample and the all other data samples already present in the data-set. Depending on the value of K, that many nearest neighbours are selected and their class is identified. The class of neighbours which has majority is assigned to the class of the new data sample. Generally, Euclidean distance is used to calculate the distance between the records. Multiple values of K should be tried and tested, and the value of K at which best performance is observed must be selected for the model.

Logistic Regression: Logistic regression algorithm is used to identify the probab- ility of occurrence of an event based on single predictor variable. Multivariate Logistic regression can be used to determine the probability of the occurrence of an event based on multiple predictor variables. The class variable that has to be predicted has to be bin- ary or dichotomous. Logistic Regression is also a supervised machine learning algorithm which used data with predetermined classes to create a model and perform predictive analysis on unseen data.

Decision Tree: It is a supervised machine learning algorithm. Due to its simple logic, effectiveness and interpretability it the most widely used classification algorithm. The model works by creating a tree-like structure by dividing the data-set into several smaller subsets based on different conditional logic. The main components of the decision tree are the decision nodes, leaf nodes and the branches. Nodes with multiple branches are the decision nodes, nodes with no branches are called the leaf nodes, and the top node is called the root node of the decision tree. The nodes are connected to each other via branches based which are different conditions. The root and decision nodes are created by computing the entropy and information gain for the data-set.

Shiny Library: The Shiny package in R is used to create interactive standalone

And web-based applications. It allows creating a user interface for the R programs by providing a platform to integrate the presentation code and program code in a single.