**SYSTEM ANALYSIS**

**EXISTING SYSTEM:**

Nowadays, the data sizes accumulated from various

fields are exponentially increasing. Due to the rapid growth

in information technology, engineering, banking and

medicine, data classification becomes more and more

significant. Data mining techniques have become popular in

commercial and scientific domains and offered promising

ways to uncover hidden patterns that were used to predict

future behavior within large amount of data [1]. Data mining

has various tasks where each task is concerned with specific

type of problem. Therefore, the method for solving

problems differs from one task to other

Nowadays, the data sizes accumulated from various

fields are exponentially increasing. Due to the rapid growth

in information technology, engineering, banking and

medicine, data classification becomes more and more

significant. Data mining techniques have become popular in

commercial and scientific domains and offered promising

ways to uncover hidden patterns that were used to predict

future behavior within large amount of data [1]. Data mining

has various tasks where each task is concerned with specific

type of problem. Therefore, the method for solving

problems differs from one task to other

Along with the enormous development of computer systems and the fast spread of internet, data processing and analysis have become a significant concern. Different soft computing techniques of data analysis have been introduced to extract valuable information from data. These techniques applied in different areas and reflected useful promising results. a novel decision tree algorithm combined with linear regression is proposed to solve data classification problem. The proposed method is applied to Turkey Student Evaluation and Zoo datasets that are taken from UCI Machine Learning Repository and compared with other classifier algorithms in order to predict the accuracy and find the best performing classification algorithm. The results show that the proposed method performs better than all other algorithms.

**DISADVANTAGES OF EXISTING SYSTEM:**

1. Zoo datasets is not providing valid results
2. The results are not up to the mark and promising.

* **Algorithm:** , Logistic Regression, Naive Bayes.,K-Nearest Neighbors.,Decision Tree.

**PROPOSED SYSTEM:**

Random forest select random sample from dataset. To get prediction result from every decision tree it will construct decision tree for every sample. Then, voting performed for every predicted outcome. Finally, the most voted prediction result is selected as final prediction outcome. Decision tree is used to predict data of future to produce meaningful continuous result . A decision tree asks series of questions to the data for arriving at final estimate. Each question narrowing our predictable values until the model gets sufficiently sure to make a solitary prediction. Model decides the order of question as well as their content. All questions are in the form of True or false. During decision tree algorithm training, the model is fitted with any historical data which is relevant to problem statement. The model learns any connection between actual data and target variable. To predict values of any new data points KNN algorithm uses feature similarity. It means new point is assigned a value based on how closely it resembles point in the training set. In KNN, Distance between new point and each training point is calculated. After calculating distance, closest k data points are selected. Euclidean Distance and Manhattan Distance are two different methods to calculate distance.

**ADVANTAGES OF PROPOSED SYSTEM:**

* In order to achieve this objective, we use different machine learning algorithms to predict internet bandwidth
* We used three machine learning algorithms to predict bandwidth requirement. From these three algorithms random forest provides highest accuracy in prediction.

**Algorithm:** **Decision tree, Random Forest, knn.**