ABSTRACT

Data Mining is a knowledge discovery from data and it treats as mining of knowledge from large amount of data in every field. The algorithms are implemented using MATLAB and fuzzy logic tool box and results are evaluated based on performance parameter in both algorithms. After doing this research experiment results show that how k-means and fuzzy C means implemented on protein data set. In this research work we present the problem that show proteins are highly affiliated to each other.

FCM allows one piece of data to belong to two or more clusters. Results based on different clusters in both algorithms. K-means is the centroid based technique. We are also compared k-means and FCM results in this research. Comparison results show that the k-means is better than FCM. With the help of this research we can remove complexity from data sets in future. So the result shows that proteins are close to each other and k-means algorithm remove data set complexity with high accuracy and less consuming time and found large sum of distance in among the statistics peak’s association to FCM algorithm. Data mining techniques is very important in the analysis of real environmental data. Forest fire is important to the forest ecosystem. Only few research focus on the scientific data. It is very difficult task. In this research show that the comparison results using bagging, stacking and random subspace algorithms taking place forest fire figures locate in to WEKA statistics mining suite. Bagging, stacking and random subspace algorithms are implemented using WEKA and experiments are behavior and consequences are assessing based on performance. Finally, we compared performance of bagging, stacking and random subspace algorithm. On the basis of experiments, we have found that using 20-fold cross validation then performance of the stacking with decision stump and decision table improve the prediction accuracy of classifier. So stacking is better and straightforward to interpret other. Stacking algorithm built accurate classifier model and consuming less time.

In this research firstly, we presented the comparative analysis of k-means algorithm and k-means algorithm with Gaussian mixture model to find number of iterations, Accuracy according to points of misclassified, Accuracy according to sum
of distance. Gaussian mixture model improve the quality of the final output. Gaussian mixture model is consists of a mixture. K-means gives the better solution for different clusters comparison to k-means with Gaussian mixture model.

So, when we used data points through k-means algorithm then k-means gives better solutions, but when we used data points through Gaussian mixture model then k-means not able to improve better performance. K-means is better than k-means with Gaussian mixture model. Secondly we presented comparative analysis of k-means algorithm and fuzzy c means algorithm. So, when we used data points through k-means algorithm then k-means gives better solutions, but when we used data points through fuzzy c means then FCM algorithm is not able to improve better performance.