IJCSE International Journal of Computer Sciences and Engineering Open Access
Research Paper Volume-6, Issue-3 E-ISSN: 2347-2693

An Effective K-means approach for Imbalance data clustering using Precise Reduction Sampling

Shaik.Nagul[1*], R.Kiran Kumar[2]

[1]*Department of Computer Science, Krishna University, Machilipatnam, India
[2] Department of Computer Science, Krishna University, Machilipatnam, India

Available online at: www.ijcseonline.org

Received: 24/Feb//2018, Revised: 03/Mar2018, Accepted: 23/Mar/2018, Published: 30/Mar/2018

Abstract— K-means clustering is one of the top 10 algorithms in the field data mining and knowledge discovery. The uniform effect in the k-means clustering reveals that, the imbalance nature of the data source hampered the performance in terms of efficient knowledge discovery. In this paper, we proposed a novel clustering algorithm known as Precise Reduction Sampling K-means (PRS_K-means) for efficient handling of imbalance data and reducing the uniform effect. The experiments shows that the algorithm can not only give attention to different instances of sub clusters for identify the intrinsic properties of the instances for clustering; and it performs better than K-means in terms of reduction in error rate and has higher accuracy and recall rate for improved performance.

Keywords: Data Mining, Knowledge Discovery, Clustering, K-means, imbalance data, uniform effect, under sampling, PRS_K-means.

I. INTRODUCTION

Grouping of the items in the similar groups using machine learning algorithm is one of the recent trends in the research community with a great scope for real time applicability. K-means is the one of best clustering algorithm for performing unsupervised learning. Researchers in large have not studied a real world source of data having class imbalance nature. Despite of the wide diversity of characteristics, instances in a particular data source, in most cases, have some similarities in intrinsic properties. For example, the data source of a disease may contain a large set of instances about one class, either positive or negative, making the data source imbalance in nature. Uncovering such data sources could make users an inconvenient way for knowledge discovery, deviating from real and useful mining tasks.

This is, however, a challenging task due to the imbalance nature of the data sources. The class imbalance nature is an unwanted feature for reducing the efficiency of the knowledge discovery. Improving the data source, in such a way to reflect, the real nature of the data source. Based on this idea, we develop a novel method, Precise Reduction Sampling K-means (PRS_K-means), for class imbalance data learning in order to discover the inherent characteristics of the data source.

Our major contributions include:

1. Deriving from *Precise Reduction technique*, a novel concept called *Precise Reduction Sampling K-means (PRS_K-means)* is proposed based on which a new similarity function between instances and instances is descried.
2. A new clustering algorithm called *PRS_K-means* is designed to group densely linked instances into similar clusters and identify their intrinsic characteristics.

Our experiments show that PRS_K-means is efficient and effective at uncovering hidden relations between different instances of sub clusters, which set a foundation for further mining and exploring varied data sources.

However, the behaviour of the imbalance datasets on the unsupervised approaches is to be investigated for better generalization. Wu. J [12] have studied on one of the scenario regarding the effects of imbalance datasets on k-means clustering. The results suggest that the k-means clustering approach generates uniform group of clusters from the input data of non uniform in nature. They named this phenomenon as "uniform effect". In this study, we investigated the causes and reasons for the uniform effect and proposed a novel algorithm to solve the uniform effect in k-means clustering.

The arrangement of paper is follows as. In Section 2, we present the current approaches of the k-means with imbalance data learning on uniform effect in clustering.

Section 3, describes the reason and rectifying technique for the problem of uniform effect and at last it laid the basis for the proposed technique PRS_K-means. Section 4 presents the experimental set up and the validation measures used for results analysis. In Section 5, results of the proposed approach PRS_K-means are presented with the k-means classical approach. Section 6 presents the conclusion with the extension of the future work.

II. CURRENT APPROACHES OF K-MEANS WITH IMBALANCE DATA LEARNING

Hui Xiong et al [1] have investigated on the issue of k-means algorithm generating the clusters of relative uniform size irrespective of non-uniform clusters in the existing dataset. Hui Xiong et al [2] have provided the coefficient of variation (CV) as a necessary criterion to validate the clustering results on the effect of skewed data distributions. Abhishek et al [3] have presented the pros and cons of K-Means algorithm towards uniform effect on skewed data distribution.

Farhad et al [4] have reviewed the benefit of sparse matrix towards PCA or K-means, for significantly faster processing, especially in a distributed big data setting. Fabon Dzogang et al [5] have proposed a new algorithmic approach to deal with data sources of high dimensionality. They also introduced a new objective function for analyzing different centroid regions. Kaile Zhou et al [6] have analyzed the varied effects of skewed distributed data o Fuzzy c-means approach towards uniform effect. They also analyzed the effect of reducing the variation in cluster sizes with respective to factors such as data source size, density and imbalance ratio.

Jaya Rama Krishnaiah et al [7] have used the technique of normalization for finding the specific clusters for massive datasets ad is useful for avoiding Empty Clusters. Hartono et al [8] have proposed an approach for optimizing K-Means clustering in handling class imbalance problem using the perceptron feed-forward neural network to determine coordinates of the centroid of a cluster in K-Means clustering processes. Md. Akmol Hussain et al [9] have presented an algorithm to alleviate the biasing effect of the uniform colour patches of the colour constancy compensation techniques which employs the k-means clustering algorithm to segment image areas according to their colour information.

Junjie Wu et al [10] have introduced different evaluation measures on different level of class imbalance distribution in different application scenarios. Richard Nock et al [11] have reviewed a wide range of clustering constrain satisfaction problems using supervised algorithm techniques.

III. THE PROPOSED PRECISE REDUCTION SAMPLING K-MEANS (PRS_K-MEANS) APPROACH

The proposed k-means algorithm has some of the unique features, which are elaborated in the below section.

1. K-means clustering is one of the best alternatives for generating clusters for arbitrary shapes.
2.K-means clustering technique have good applicability on the datasets of large size due to the simple technique used for clustering process.
3.Density based clustering techniques are having incremental mechanism for updating the clustering centroid for varied data sources such as class imbalance nature.

3.1 Motivating Concepts

To derive a more efficient algorithm for proposed Precise Reduction Sampling (PRS_K-means), the following definitions are first introduced.

***Definition 1 (Precise Sampling)*:**
Given a set of *m* dataset, (1 2 , , ,) *mi* instances. The instances which are pure i.e in terms of membership of class are to be categories as a subgroup. These instances are identified using the precise techniques of membership estimation and only those instances are sampled for efficient over sampling.

Definition 2 (Precise Reduction):
Given point *p* and point *q*, the nearest distance between these instances is measured using k-nearest approach. These instances may also be defined as noisy, outlier or borderline instances. The instances in the specified distance *d b*etween *p* and *q* is defined as the range for finding the noisy, outlier or borderline instances are removed using precise reduction technique.

This section presents the proposed algorithm Precise Reduction Sampling K-means (PRS_K-means), is a density based approach for maximizing the intrinsic properties of the instances in the similar groups. The K-means clustering approach depends on some of the parameters such as initially setting the number of centres.

The imbalance nature of the dataset can e reduced y either performing over sampling or under sampling. In this proposed work, we prefer to use a under sampling approach. Precise reduction technique is oriented towards removing unnecessary instances from the majority subset by proper choosing of instances. The proposed Precise Reduction technique employs, nearest neighbour contributing for an improvement over 20-30% of better classification.

The instances oriented towards class are retained in the class of interest C and deletes the noisy instances from the overall class O where O = T-C. The implementation of the above said technique is performed y using the below two stages,

The precise reduction technique is used to exactly point out the specific noisy instances for removal from A1 in O. Specifically, *n* nearest neighbour samples are used to remove, samples with a different class to the majority class of the *n* nearest neighbours, It removes samples that have different classes to at least *n-1* of *n* nearest neighbors. Subsequently, the neighbourhoods are processed again and a set A2 is created. Then, the *n* nearest neighbour samples that belong to O and lead to C samples misclassifications are inserted in the set A2. In the last stage, the fine tuned data is prepared y performing the precise reduction from all the sub classes A1 and A2, A1∪A2. The pseudo code of the Precise Reduction technique is given below,

1. Split data *T* into the class of interest *C* and the rest of data *O*.
2. Identify noisy data *A1* in *O* with edited nearest neighbour technique.
3. For each class *Ci* in *O*
4. If (*x* belongs to *Ci* in *n* nearest neighbour of misclassified *y* belongs to *C*)
5. Reduced data *S=T-(A1UA2)*

The improved majority subset after applying précised reduction technique is sampled and combined with existing minority subset to form an improved data source. The improved data source is applied to the base algorithm i.e k-means clustering technique and the required validation measures are generated.

IV. EXPERIMENTAL SETUP AND ASSESSMENT CRITERIA

Twenty two real datasets from UCI [13] data repositories are used in the following experiments. N
The experimental simulation details of the datasets used are presented in the table 1. The datasets are arranged in alphabetical order and the properties of imbalance ratio is also presented, which gives the level of the class imbalance nature in the dataset.

Table1 UCI datasets and their properties

S.no. Dataset	Ins	Attributes	IR
1. abalone19	4174	9	129.43
2. abalone9-18	731	9	16.40
3. ecoli-0-1-3-7_vs_2-6	281	8	39.14
4. ecoli4	336	8	15.8
5. glass-0-1-6_vs_2	192	10	10.29
6. glass-0-1-6_vs_5	184	10	19.44
7. glass2	214	10	11.58
8. glass4	214	10	15.46
9. glass5	214	10	22.77
10. page-blocks-1-3_vs_4	472	11	15.85
11. shuttle-c0-vs-c4	1829	10	13.86
12. shuttle-c2-vs-c4	129	10	20.5
13. vowel0	988	14	9.97
14. yeast-0-5-6-7-9_vs_4	528	9	9.35
15. yeast-1-2-8-9_vs_7	947	9	30.56
16. yeast-1-4-5-8_vs_7	693	9	22.1
17. yeast-1_vs_7	459	8	14.3
18. yeast-2_vs_4	514	9	9.07
19. yeast-2_vs_8	482	9	23.1
20. yeast4	1484	9	28.09
21. yeast5	1484	9	32.72
22. yeast6	1484	9	41.4

The validation strategy used for the experimental methodology is 10 fold cross validation. Where one set out of 10 sets is used for testing and remaining 9 sets are used for training. The testing and training folds are done for 10 runs and the average value of 10 runs is taken as the mean value in the simulations. The validation measures used in the experimental setup are AUC, precision, Recall ad F-measure.

The Area under Curve (AUC) measure is computed using the below equation (1) or (2),

$$AUC = \frac{1 + TP_{RATE} - FP_{RATE}}{2} \quad \text{---------- (1)}$$

Or

$$AUC = \frac{TP_{RATE} + TN_{RATE}}{2} \quad \text{---------- (2)}$$

The Precision measure is computed using the below equation (3),

$$precision = \frac{TP}{(TP)+(FP)} \quad \text{---------- (3)}$$

The Recall measure is computed using the below equation (4),

$$\text{Re}call = \frac{TP}{(TP)+(FN)} \quad \text{---------- (4)}$$

The F-measure Value is computed using the below equation (5),

$$F-measure = \frac{2 \times \text{Pr}ecision \times \text{Re}call}{\text{Pr}ecision + \text{Re}call} \quad \text{---------- (5)}$$

V. RESULTS

The experimental simulation is performed on Weka [14] open source environment on a system unit of windows 7 with 4.00 GB Ram with i5-2410M CPU unit. We compare PRS_K-means with benchmark K-means algorithm.

In this experiment we focus on studying classical k-means clustering approach, there are more elements which could be influenced by data characteristics for other classifiers and the uniform effect may not be visible in specific context. We have decided to K-means clustering algorithms, which have been often considered in related works and which represent benchmark in clustering.

In the first step of experiments, we compare the performance of PRS_K-means on all the 22 imbalanced datasets. Again we can present the details of selected experiments in due course of discussion. We have made simulated experiments for K-means and PRS_K-means. Table 2 presents the AUC results on real dataset for k-means and PRS_K-means.

5.1 Effectiveness

Depending on the AUC, precision, recall and f-measure measurement, the performance of k-means and the proposed PRS_K-means approach is investigated. The following table 2 to 5 with the average performance of PRS_K-means and the compared k-means for each dataset using the best performance are presented.
The study revealed that the proposed PRS_K-means model is mainly affected by the degree of imbalance of the dataset. It also showed that the PRS_K-means model was not significantly affected by the number of instances. Generally, the PRS_K-means performed the best in datasets with relatively small IR. This result confirms the results argued that PRS_K-means improves small class modelling. On the other hand, PRS_K-means also performed best in relatively high IR datasets.

The results emphasize the negative effect of IR on a non-preprocessed dataset. The results also shows that the number of features and the sample size contribute to the clustering performance , the possible reason for the drop in balance in the first dataset abalone19, IR 129.43, is the relatively high number of features and instances, 9 feature and 4174 instances.

5.2 Efficiency

The PRS_K-means average performance showed an overall datasets is similar performing with compared k-means approach. . On the other hand, the under-sampling performance demonstrated lower overall performance on some datasets. However, unlike the traditional k-means, PRS_K-means showed a steady performance against relatively high IR as shown in tables 2 to 5.

Table 2 Results of AUC on all the datasets with summary of tenfold cross validation performance

Datasets	K-means	PRS_K-means
abalone19	0.414±0.136○	0.409±0.149
abalone9-18	0.479±0.120●	0.491±0.166
ecoli-0-1-3-7_vs_2-6	0.660±0.190●	0.714±0.224
ecoli4	0.533±0.173●	0.707±0.120
glass-0-1-6_vs_2	0.489±0.074●	0.601±0.119
glass-0-1-6_vs_5	0.621±0.249○	0.542±0.241
glass2	0.495±0.047●	0.506±0.080
glass4	0.785±0.191○	0.621±0.258
glass5	0.603±0.242○	0.493±0.235
page-blocks-1-3_vs_4	0.587±0.155○	0.406±0.154
shuttle-c0-vs-c4	0.685±0.191●	0.862±0.188
shuttle-c2-vs-c4	0.563±0.233○	0.483±0.179
vowel0	0.486±0.090●	0.498±0.119
yeast-0-5-6-7-9_vs_4	0.499±0.162●	0.749±0.106
yeast-1-2-8-9_vs_7	0.533±0.174●	0.586±0.142
yeast-1-4-5-8_vs_7	0.548±0.139●	0.609±0.139
yeast-1_vs_7	0.638±0.130●	0.656±0.097
yeast-2_vs_4	0.802±0.095●	0.860±0.099
yeast-2_vs_8	0.501±0.173●	0.513±0.076
yeast4	0.768±0.089○	0.730±0.147
yeast5	0.844±0.064●	0.847±0.075
yeast6	0.802±0.088●	0.850±0.065

● Bold dot indicates the win of Proposed PRS_K-means approach;

Table 3 Results of Precision on all the datasets with summary of tenfold cross validation performance

Datasets	K-means	PRS_K-means
abalone19	0.003±0.005○	0.002±0.005
abalone9-18	0.050±0.037●	0.054±0.061
ecoli-0-1-3-7_vs_2-6	0.062±0.172●	0.158±0.246
ecoli4	0.042±0.063●	0.188±0.112
glass-0-1-6_vs_2	0.013±0.069●	0.157±0.172
glass-0-1-6_vs_5	0.098±0.140●	0.118±0.183
glass2	0.007±0.036●	0.024±0.123
glass4	0.217±0.178○	0.188±0.186
glass5	0.073±0.104	0.073±0.121
page-blocks-1-3_vs_4	0.078±0.044●	0.112±0.233
shuttle-c0-vs-c4	0.324±0.372●	0.671±0.358
shuttle-c2-vs-c4	0.065±0.170○	0.000±0.000
vowel0	0.087±0.033●	0.091±0.064
yeast-0-5-6-7-9_vs_4	0.101±0.086●	0.462±0.145

yeast-1-2-8-9_vs_7	0.035±0.023●	0.083±0.053
yeast-1-4-5-8_vs_7	0.052±0.033●	0.147±0.100
yeast-1_vs_7	0.096±0.044●	0.196±0.048
yeast-2_vs_4	0.279±0.092●	0.697±0.179
yeast-2_vs_8	0.067±0.167○	0.003±0.013
yeast4	0.092±0.022○	0.076±0.038
yeast5	0.094±0.023●	0.126±0.065
yeast6	0.069±0.019●	0.078±0.035

● Bold dot indicates the win of Proposed PRS_K-means approach;

Table 4 Results of Recall on all the datasets with summary of tenfold cross validation performance

Datasets	K-means	PRS_K-means
abalone19	0.131±0.257○	0.100±0.275
abalone9-18	0.324±0.243○	0.323±0.354
ecoli-0-1-3-7_vs_2-6	0.390±0.490●	0.530±0.502
ecoli4	0.330±0.462●	0.820±0.386
glass-0-1-6_vs_2	0.025±0.131●	0.500±0.503
glass-0-1-6_vs_5	0.380±0.488○	0.370±0.485
glass2	0.020±0.098●	0.055±0.224
glass4	0.760±0.399○	0.615±0.476
glass5	0.370±0.485○	0.300±0.461
page-blocks-1-3_vs_4	0.633±0.320○	0.223±0.297
shuttle-c0-vs-c4	0.495±0.326●	0.784±0.359
shuttle-c2-vs-c4	0.200±0.402○	0.000±0.000
vowel0	0.431±0.164●	0.459±0.259
yeast-0-5-6-7-9_vs_4	0.440±0.289●	0.735±0.213
yeast-1-2-8-9_vs_7	0.547±0.350○	0.540±0.347
yeast-1-4-5-8_vs_7	0.443±0.296●	0.500±0.312
yeast-1_vs_7	0.700±0.309●	0.750±0.168
yeast-2_vs_4	0.862±0.186●	0.807±0.202
yeast-2_vs_8	0.465±0.350○	0.040±0.197
yeast4	0.828±0.189○	0.733±0.327
yeast5	0.970±0.171	0.970±0.171
yeast6	0.898±0.183●	0.990±0.100

● Bold dot indicates the win of Proposed PRS_K-means approach;

Table 5 Results of F-measure on all the datasets with summary of tenfold cross validation performance

Datasets	K-means	PRS_K-means
abalone19	0.005±0.010○	0.003±0.009
abalone9-18	0.086±0.063●	0.092±0.101
ecoli-0-1-3-7_vs_2-6	0.089±0.180●	0.218±0.270
ecoli4	0.073±0.108●	0.301±0.167
glass-0-1-6_vs_2	0.017±0.089●	0.235±0.249
glass-0-1-6_vs_5	0.153±0.209●	0.172±0.245
glass2	0.010±0.052●	0.028±0.134
glass4	0.317±0.204○	0.278±0.243
glass5	0.119±0.167○	0.115±0.186
page-blocks-1-3_vs_4	0.138±0.076○	0.115±0.171
shuttle-c0-vs-c4	0.368±0.351●	0.708±0.341
shuttle-c2-vs-c4	0.090±0.205○	0.000±0.000
vowel0	0.144±0.054●	0.148±0.083
yeast-0-5-6-7-9_vs_4	0.162±0.125●	0.558±0.153
yeast-1-2-8-9_vs_7	0.066±0.043●	0.143±0.0910
yeast-1-4-5-8_vs_7	0.092±0.059●	0.222±0.140
yeast-1_vs_7	0.169±0.076●	0.310±0.073
yeast-2_vs_4	0.416±0.117●	0.740±0.172
yeast-2_vs_8	0.089±0.115○	0.005±0.025
yeast4	0.165±0.039○	0.137±0.067
yeast5	0.170±0.040●	0.218±0.094
yeast6	0.128±0.034●	0.143±0.059

● Bold dot indicates the win of Proposed PRS_K-means approach;

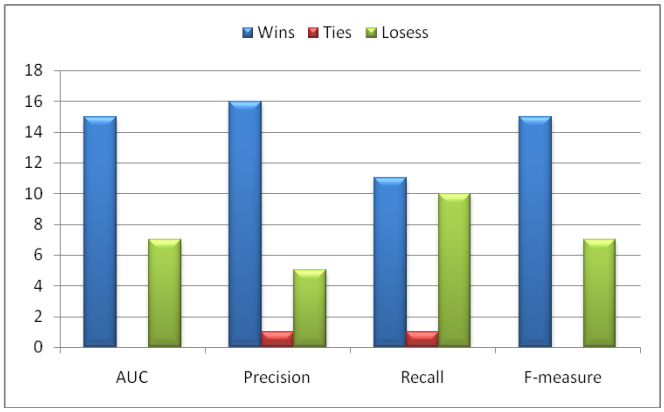

Fig. 1 Trends for K-means versus PRS_K-means on imbalance data sets on validation asures AUC, precision, Recall and F-measure

The results generated have used a statistical significance technique for performing the analysis. The one tailed paired t-test used for statistical evidence is a level of 5% significance. The results suggest that PRS_K-means performs better or similar than competing k-means method. The complete summary of the experimental analysis is shown in the table 6.

Table 6 Summary of experimental results for USDD

Results	Systems	Wins	Ties	Losses
PRS_K-means	AUC	15	0	7
versus	Precision	16	1	5
K-means	Recall	11	1	10
	F-measure	15	0	7

Figure 1 presents, the graphical representation of the proposed PRS_K-means approach against K-means against the number of wins, ties and losses. The results indicate that good numbers of wins are registered by the proposed approach on most of the datasets. The proposed approach PRS_K-means is one of the best alternatives for handling imbalance data learning for unsupervised learning techniques.

VI. CONCLUSION

In this paper, we developed a novel clustering with precise reduction of outlier's and noisy instances and thereby for the remaining instances precise sampling is performed for improved performance. The proposed PRS_K-means approach is validated using 22 imbalance real world datasets and the results suggest a significant improvement in the validation measures.

REFERENCES

[1] *Prateeksha Tomar, Amit Kumar Manjhvar, "*Clustering Classification for Diabetic Patients using K-Means and M-Tree prediction model*", International Journal of Scientific Research in Multidisciplinary Studies , Vol.3, Issue.6, pp.48-53, 2017*

[2] Hui Xiong, Junjie Wu, and Jian Chen*,"* K-Means Clustering Versus Validation Measures: A Data-Distribution Perspective", IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, VOL. 39, NO. 2, APRIL 2009.

[3] Abhishek kumar K and Sadhana,: SURVEY ON K-MEANS CLUSTERING ALGORITHM", *International Journal of Modern Trends in Engineering and Research (IJMTER) Volume 04, Issue 4, [April– 2017]*

[4] Farhad Pourkamali-Anaraki and Stephen Becker, "Preconditioned Data Sparsification for Big Data with Applications to PCA and K-means",

[5] Fabon Dzogan, Christophe Marsala, Marie-Jeanne Lesot and Maria Rifqi," An ellipsoidal K-means for document clustering", 2012 IEEE 12th International Conference on Data Mining

[6] Kaile Zhou, Shanlin Yang," Exploring the uniform effect of FCM clustering: A data distribution Perspective", Knowledge-Based Systems 96 (2016) 76–83

[7] Jaya Rama Krishnaiah VV, Ramchand H Rao K, Satya Prasad R (2012) Entropy Based Mean Clustering: An Enhanced Clustering Approach. J Comput Sci Syst Biol 5: 062-067. doi:10.4172/jcsb.1000091

[8] Hartono, O S Sitompul, Tulus and E B Nababan,: Optimization Model of K-Means Clustering Using Artificial Neural Networks to Handle Class Imbalance Problem", IOP Conf. Series: Materials Science and Engineering 288 (2017) 012075

[9] Md. Akmol Hussain, Akbar Sheikh Akbari, Ahmad Ghaffari, "Colour Constancy using K-means Clustering Algorithm", 2016 9th International Conference on Developments in eSystems Engineering.

[10] Junjie Wu, Hui Xiong and Jian Chen," Adapting the Right Measures for K-means Clustering",

[11] Richard Nock and Frank Nielsen," On Weighting Clustering", EEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 28, NO. 8, AUGUST 2006

[12] Wu.J,"The Uniform Effect of K-means Clustering", J. Wu, *Advances in K-means Clustering*, Springer Theses, DOI:10.1007/978-3-642-29807-3_2, © Springer-Verlag Berlin Heidelberg 2012.

[13] HamiltonA. Asuncion D. Newman. (2007). *UCI Repository of Machine Learning Database* (School of Information and Computer Science, Irvine, CA: Univ. of California [Online]. Available: http://www.ics.uci.edu/ ~mlearn/MLRepository.html

[14] Witten, I.H. and Frank, E. (2005) Data Mining: Practical machine learning tools and techniques. 2nd edition Morgan Kaufmann, San Francisco.

Authors Profile

Mr.Shaik.Nagul completed his Bachelor of computer applications from Aharya Nagarjuna University in 2003 and Master of Computer Applications from Aharya Nagarjuna University in 2007. Master of Technology from Aharya Nagarjuna University in 2010. He is currently pursuing Ph.D from Krishna University and currently working as Assistant Professor in Department of Computer Science and Engineering in Lendi Institute of Engineering and Technology. He is a Associate member of IEI. His main research work focuses on Data Minig Algorithms. He has 10 years of teaching experience and 4 years of Research Experience.

Dr.R.Kiran Kumar completed his PhD (Computer Science and Engineering) from Acharya Nagarjuna University, India. M.Tech (Computer Science and Engineering) from JNTU, Kakinada, India. MCA (Master of Computer Applications) from Andhra University, India. B.Sc. from Andhra University, India. And currently working as senior Assistant Professor in Krishna University,Machilipatnam,Andhra Pradesh, India

