A Heuristic Effort Estimation Method using BAT Algorithm through Clustering

M. Padmaja, D. Haritha

Abstract*: The most challenging task within software industries or organizations is project handling. Effort estimation has to be made accurately, before handling the project in order to win the bidding among several competitions. To estimate this effort an efficient method has to be selected. In this paper, we propose two novel methods: Grey Relational Analysis (GRA) method and Meta heuristic approach as BAT algorithm (BA) for predicting the effort of a specific project. These two methods are applied on suitable parameters or features that are obtained through one-way ANOVA and Pearson Correlation Coefficient (PCC) methods to forming a reduced data. The experimentation of this paper proves the efficiency of the anticipated method. The proposed method is applied over original data set and dimensionally reduced data set which are compared and tabulated. From the experimental results it is evident that the proposed method produced less error and more prediction.

Index Terms*: Effort Estimation, Grey Relational Analysis, BAT Algorithm, K-Means Clustering.

I. INTRODUCTION

Effort estimation is an important aspect of software engineering projects. The development of a project by a company begins with a thorough study of requirements. To develop the project selection of parameters from the history projects or analogy approach is crucial [1]. The estimation of the project becomes easy if the customer provides all requirements clearly, but this cannot be done practically, as the customer‘s lack of knowledge about the developing product. The estimation becomes the main goal of the developer when the development of the project proceeds to further stages of detailed planning and implementation. If the project management estimates improperly that leads to several problems for stakeholders. Before handle the project, the management team must estimate accurately when compare with customer constraints. If customer can specify all requirements then correctly estimate the project, otherwise within insufficient knowledge, information estimation can leads to problems.

Deng Ju-long introduced Grey System Theory in 1980 [3]. It is a novel method applicable in the study of uncertained problems with a subset of data and/or lack of information [2]. Grey Relational Analysis (GRA) works with partially known and partially unknown information by drawing out valuable information [4]. Grey Relational Analysis is included in the Grey System Theory and used to find out projects that are similar to a new project and constitute the similar parameters required to effort estimation or cost of a project. In GRA methodology, With Grey Relational Grades (GRG), know the relationship of current project with historical projects [5,8]. In order to estimate the effort of a project, first we find the required parameters. To achieve this we selected two methods, ANOVA and PCC methods based on effort value from the historical value [6]. These two methods were executed in excel with effort and found the efficient features as maximum values from the graphical representation [7]. Optimal method i.e., BAT algorithm has been chosen to show the optimal result with optimization algorithm. Optimization is a process of providing the optimal solution of problem under the considered situations. The essential task of optimization is to minimize time or maximize desired benefit of a given engineering system. All systems that are to be optimized have an objective function and several decision variables that affect the objective function. Stochastic algorithms are categorized into two groups as heuristic and metaheuristic algorithms. Heuristic means to produce the results with optimization by trial and error method. Meta means “beyond, in an upper level” and - so the term metaheuristic refers to a higher level of heuristics [10]. Some Heuristic algorithms are moved by the nature such algorithms are known as nature inspired algorithms. One of the Meta heuristic approaches has been selected to estimate the effort such as BAT Algorithm (BA) in number of iterations [11]. Bat Algorithm (BA) proposed by Xin-She Yang in 2010, is a one of the swarm based metaheuristic optimization algorithm inspired by a goods known as echolocation. This BA is works based on the echolocation behavior of micro bats with varying pulse rates of emission and loudness. Micro bats echolocation capability helps them to detect preys, distinguish different kinds of insects. Echolocation is a type of sonar that guides bats in their flying and hunting behavior [12]. Bats can move in complete darkness to get a prey.

II. RELATED WORK

The aim of a Project Manager is to deliver the efficient software product or project with more quality, less cost and reach to customer needs. Responsibility of the manager is to estimate the effort and cost accurately before handling the project, with this estimation the developer can develop the product and satisfy the customer. Qinbao Song et al., [5] had used GRACE method on different datasets and compare the results with other models. Sun-Jen Huang et al., [4] have combined Grey

Revised Manuscript Received on July 05, 2019.

M.Padmaja, Assistant Professor, Dept. of CSE, GIT, GITAM (Deemed to be University), Visakhapatnam, Andhra Pradesh, INDIA.

Dr D.Haritha, Professor, Dept. of CSE, University College of Engineering, JNTUK, Kakinada, Andhra Pradesh, INDIA.

Retrieval Number: I8242078919/19©BEIESP
DOI:10.35940/ijitee.I8242.078919

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication

Relational Analysis with Genetic Algorithm to estimate the effort of a project with help of existed projects. It was applied on different data types and evaluated the results. In addition to this, M.Padmaja et al., [8] have used Grey Relational Analysis on Kemerer data set to estimate the effort. This model shows the minimum error rate on this data than other models. Jin-Cherng Lin et al., [6-7] have used Particle Swarm Optimization (PSO) on particular dataset with clustering technique. To do this, they used ANOVA and PCC methods to select suitable parameters. M.Padmaja et al., [9] applied GRA method on COCOMO81 dataset with k-means clustering. In this, suitable parameters with ANOVA and PCC methods were selected, and this sufficient data is known as reduced data. On original and reduced data the proposed method was applied and estimated the effort. This method shows the minimum error with required parameters than all parameters on clustered projects and also extended with Taguchi method is known as Grey-Taguchi used to know the contribution of parameters to estimate the effort of a current project [15]. Neha Gupta [13] used bat algorithm on COCOMO63 dataset on different types of modes of projects were tested and obtained results were compared with the obtain using the current COCOMO model coefficients. Muhammad Ibrahim [14] used bat algorithm to evaluate the software projects of NASA60 dataset. The results indicated that MMRE have reduced in comparison with COCOMO model and other models. M.Padmaja et al., [16] applied BAT algorithm (BA) on Kemerer data performed the results compared with GRA model. Estimated effort with BA and evaluate the results with metrics as MMRE. This was shown minimum error than GRA. Malik Muneeb Abid et al., [17] provided survey of heuristic approaches to solve the travelling salesman problem (TSP). And compare the results with other method and selected approach to proved high quality and consumption of time to solve the TSP. Haiying Dong et al., [18] worked to diagnosize the faults using rough sets and grey relational analysis (GRA) methods. In this paper, decision table was selected and applied these methods to identify the faults and compare the results with other models. The proposed method proved the results effectively.

III. PROPOSED METHODOLOGY

In order to estimate the effort of a current project by using Grey Relational Analysis (GRA) is used to choose nearer projects from the historical data. The analogy method was used to select projects which are similar to the current project. If the historical data has large number of projects it becomes difficult to select nearer projects. So clustering method was chosen to divide into clusters. In this we used COCOMO81 data set. This dataset has sixty three (63) projects and each project has fifteen (15) features. It has a number of projects, so k-means clustering method was applied on 15 parameters and effort, such parameters are rely, data, cplx, time, stor, virt, turn, acap, aexp, pcap, vexp, lexp, modp, tool, sced to divide into a three number of clusters using WEKA tool. Here we selected two methods, Grey Relational Analysis (GRA) and one of the Meta heuristic algorithm i.e., BAT Algorithm (BA) were applied on original data to estimate the effort. Description of GRA as explained in [8] and working of BA is explained in [16]. Then applied k-means clustering method on original data to divided into k number of clusters. GRA and BA methods are applied on original data and clustered data. In the data set every project has more number of parameters so, only influenced parameters should be selected. To achieve this, ANOVA and PCC methods were selected and the procedure of these methods is described in [9]. By using these methods essential parameters were taken which are eight in addition to the actual effort: rely, data, turn, acap, aexp, pcap, lexp and modp. With these parameters that data is called as reduced data set. On original data and reduced data, k-means clustering technique had been applied using WEKA tool to divide into three clusters. On same clustering projects, BA and GRA methods were applied to predict the effort of every project and found the error rate which is MMRE and prediction at particular level that is PRED as described in Figure 1. Determination of suitable method is decided based on metrics such as MMRE, MdMRE and PRED. After applying any method to estimate the effort, if MMRE value is minimum and PRED value is maximum then that method is sufficient. So we use these metrics following the Eq.1, Eq.2 & Eq.3.

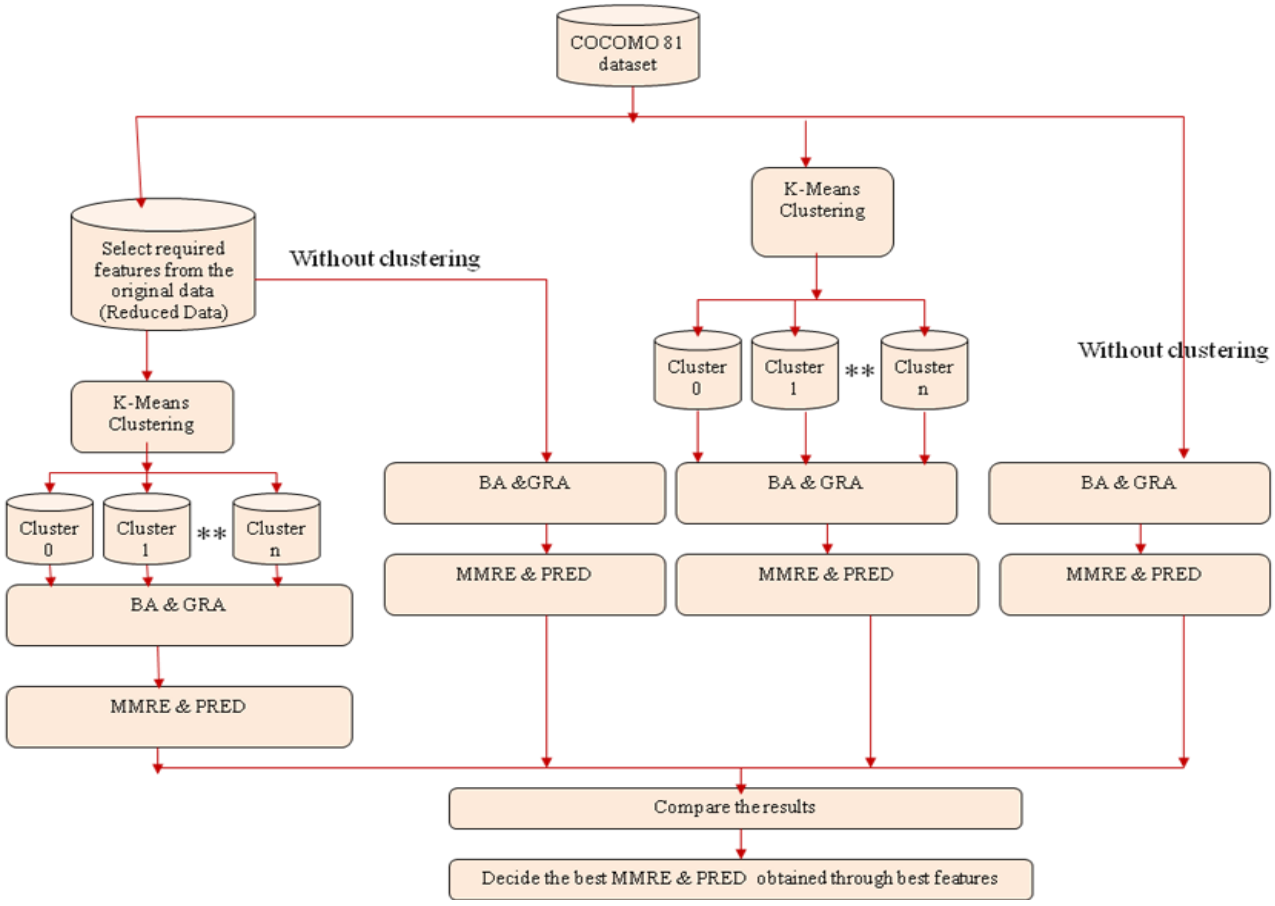

Fig. 1 Flow chart of proposed system using BAT algorithm (BA) & GRA

The BAT Algorithm (BA) is used to calculate effort estimation, and also checked if this method is efficient or not by using metrics. MMRE means Mean Magnitude Relative Error and it is deliberate with Eq. 1. It is used to determine the error rate of the method.

$$MMRE = \frac{1}{N}\sum_{i=1}^{n} MRE \tag{1}$$

Here N means number of projects.

MRE means Magnitude Relative Error

$$MRE = \frac{\left|ActualEffort_i - EstimatedEffort_i\right|}{ActualEffort_i}$$

In this methodology, MMRE is the fitness value to evaluate the method.

$$MdMRE = Median(MRE) \tag{2}$$

In case of large data, we choose another metric as MdMRE means median of MRE. It can be calculated by using Eq. 2.

$$PRED(l) = \frac{k}{n} \tag{3}$$

Here l means on particular level i.e., l=25%

k means number of projects at MRE <=0.25

n means total number of projects

We choose another metric to provide better accuracy i.e., PRED. Commonly 'l' value is selected as 25%. From the experimental results, we can decide on which method getting better results such as minimum error rate (MMRE, MdMRE) and maximum prediction (PRED). The results are shown in the section below.

IV. EXPERIMENTAL ANALYSIS

COCOMO81 dataset is used for evaluation and the respective results are tabulated below. On this dataset, GRA method and BAT algorithms are applied to estimate the effort. The step by step implementation of the proposed methodology is described below. One-way ANOVA method and Pearson correlation coefficient (PCC) methods are also used to find influential features. From these methods, most influenced parameters such as rely, data, turn, acap, aexp, pcap, lexp, modp and effort were considered to estimate the effort [9]. Those required parameters on same number of projects are known as reduced data. Now, on this reduced data both methods were applied to estimate the effort. The original data has a large number of projects. So calculation of effort is difficult. To estimate the effort chooses k-means clustering technique was implemented on original dataset and as well as reduced dataset. By using k-means clustering three clusters were chosen. Every cluster has some number of projects. Then in the next process of the proposed methodology both the methods were applied on these two datasets are divided into 'k' number of projects to calculation of effort. From the proposed method, first GRA method was applied on COCOMO81 original data with k-means clustering to estimate the effort. Here the original data was divided

into three clusters. On these three clusters the proposed method to estimate the effort was performed.

Table 1. Evaluation results on original data with clustering using GRA method

	No. of Projects	MMRE	MdMRE	PRED
Cluster 0	6	0.303285	0.279159	0.33333
Cluster 1	40	0.271319	0.212437	0.575
Cluster 2	17	0.303614	0.213741	0.588235
Avg	Total (63)	0.292739	0.235112	0.498855

Evaluated the results using metrics such as MMRE, MdMRE and PRED from which the result with minimum error and maximum prediction than other models was proven. The results are shown in Table 1. Then the same method (GRA) was applied on the only required parameters. The required parameters were selected using combine methods such as ANOVA and PCC. After getting the required features k-means clustering method was applied to divide into three clusters. On these clusters again GRA method was applied and the results were evaluated.

Table 2. Evaluation results on reduced data with clustering using GRA method

	No.of Projects	MMRE	MdMRE	PRED
Cluster 0	26	0.131357	0.117983	0.923077
Cluster 1	30	0.31481	0.207323	0.53333
Cluster 2	7	0.251605	0.237869	0.571429
Avg	Total (63)	0.232591	0.187725	0.67595

With required features, the effort using GRA method on clustered projects was estimated. The results are shown in the Table 2. These results were compared with the other model.
Next we selected Meta heuristic approach such as BAT algorithm which was applied on the original data with clustered projects of every cluster to predict the effort and evaluate the result using some metrics. Evaluation results are shown in Table 3.

Table 3. Evaluation results on original data with clustering using BAT algorithm

	No.of Projects	MMRE	MdMRE	PRED
Cluster 0	6	0.265301	0.26845	0.33333
Cluster 1	40	0.253841	0.21015	0.6842
Cluster 2	17	0.276423	0.19546	0.60213
Avg	Total (63)	0.265188	0.22468	0.53988

Then the effort was estimated with required parameters. From this, new dataset was prepared. This data set has same projects but has only nine features. With these features are known as reduced data. Then applied K-Means clustering to data and divided into three-clusters. On every clustered data proposed method was applied to predict the effort. Evaluation results as shown in the Table 4.

Table 4. Evaluation result on reduced data with clustering using BAT algorithm

	No.of Projects	MMRE	MdMRE	PRED
Cluster 0	26	0.15324	0.10478	1.2354
Cluster 1	30	0.25486	0.19542	0.53333
Cluster 2	7	0.23145	0.21498	0.58746
Avg	Total (63)	0.213183	0.171727	0.7854

From the Tables 1 and 2, the efficiency of the results on both types of data sets (i.e., original data and reduced data) with clustering technique was observed. From the experiments MMRE is a minimum of 0.213183 with influenced parameters; and prediction is a maximum of 0.7854 with influenced parameters on clustered data. Finally, the results were compared with one metric as MMRE on two methods with clustering and without clustering on original data and reduced data. From the experiments better results are proved with clustering with most influenced parameters using BAT algorithm than GRA method as shown in Figure 2.

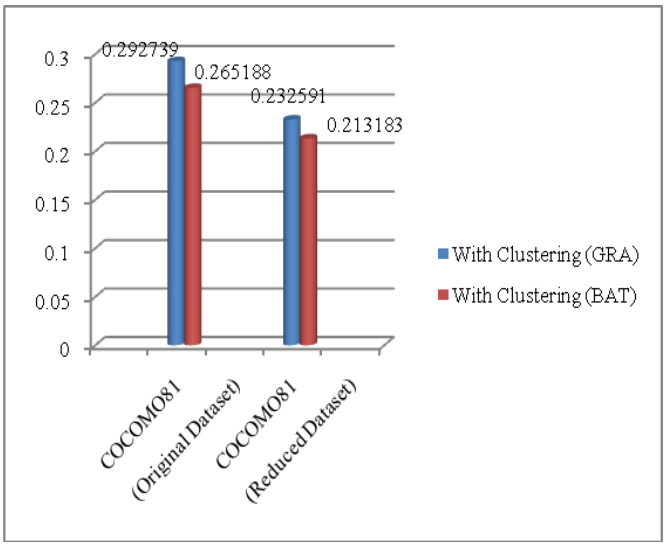

Fig 2. Efficiency of MMRE with anticipated model

On both types of datasets, both methods were applied to estimate the effort with clustering and evaluate the results, the BAT algorithm works efficiently than GRA on reduced data with clustering method with minimum error rate as 0.213183 and maximum prediction as 0.7854. These results are depicted in Figure 3.

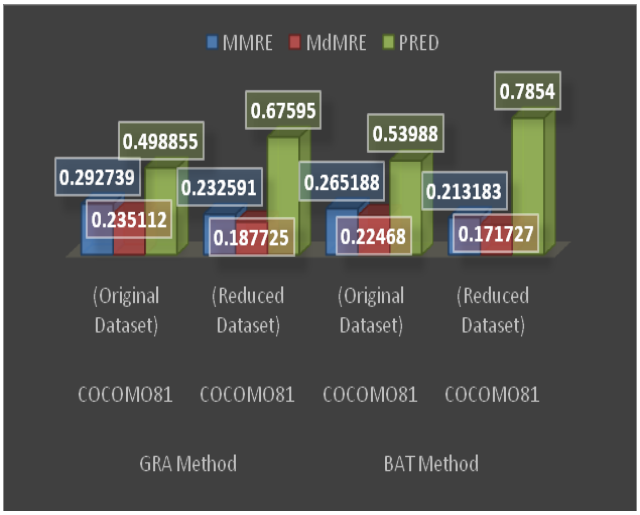

Fig 3. Comparison of results with proposed method

To know the efficiency of a proposed method in all experiments, the final result with some metrics such as MMRE, MdMRE and PRED was verified. Here, MMRE and MdMRE always a minimum and PRED is to be a maximum. Whichever method has minimum MMRE and maximum PRED, that method is efficient. It can be observed that, for these criteria, BAT algorithm (BA) works with minimum error as in MMRE, MdMRE 0.213183, 0.171727 respectively and maximum PRED as 0.7854 on reduced data with clustering than GRA are as prescribed in Figure 3. The results with GRA method was shown in [9], compare the results in all considerations with BAT algorithm (BA) in this paper. Finally, the BA was proved to show better results than GRA. So this method is used to estimate the effort accurately along with efficiently.

V. CONCLUSION

Software projects development is a very challenging task to be handled by any industry. Effort estimation for the developing projects has to be predicted before developing the software that is the newly developed project must ensure more quality with low cost. To assure this, team members must choose an appropriate model producing the mentioned qualities to deliver the project in time. This paper presents accurate and reliable effort estimation is necessary in accordance to the given project conditions. One of the Meta Heuristic method i.e., BAT Algorithm (BA) is applied to know the estimation of a particular assignment and the results show the error rate is low than with other algorithmic model i.e., Grey Relational Analysis (GRA) individually and using clustering technique on the COCOMO81 data and on clustered data. From the experiments, proposed model with clustering performs accurate estimation and show the less error rate than other model. In future this work also extended with other methods to selection of suitable parameters on different datasets.

REFERENCES

1. Martin Shepperd, Chris Schofield and Kitchenham "Effort Estimation using Analogy", *IEEE*, 2009
2. Lin. Yi, Liu. Sifeng "A Historical Introduction to Grey Systems Theory", *IEEE International Conference on Systems*, 2004
3. Deng Julong "Introduction to Grey System Theory", *The Journal of Grey System*, pp. 1-24, 1989
4. Sun-Jen Huang, Li-Wei Chen and Nan-Hsing Chiu, "Integration of the Grey Relational Analysis with Genetic Algorithm for Software Effort Estimation", *European Journal of Operational Research*, pp. 898–909, 2008
5. Qinbao Song, Martin Shepperd and Mair, "Using Grey Relational Analysis to Predict Software Effort with Small Data Sets", *11th IEEE International Software Metrics Symposium*, 2005
6. Jin-Cherng Lin, Yueh-Ting Lin, Han-Yuan Tzeng and Yan-Chin Wang, "Using Computing Intelligence Techniques to Estimate Software Effort", *International Journal of Software Engineering & Application*, Vol: 4, No: 1, pp. 43-53, Jan 2013
7. Jin-Cherng Lin , Han-Yuan Tzeng, "Applying Particle Swarm Optimization to Estimate Software Effort by Multiple Factors Software Project Clustering", *IEEE*, 2010
8. M.Padmaja, Dr D.Haritha, "Software Effort Estimation using Grey Relational Analysis", *International Journal of Information Technology and Computer Science*, Vol. 9, No: 5, pp. 52-60, May 2017
9. M.Padmaja, Dr D.Haritha, "Software Effort Estimation using Grey Relational Analysis with K-Means Clustering", *4th International Conference on Information System Design and Intelligent Applications, 15th - 17th, June, 2017, Duy Tan University, 3 QuangTrung, Da Nang, VietNam, Springer, AISC Series*, Jan 2018
10. Xin-She Yang, "Nature-Inspired Metaheuristic Algorithms", 2nd Edition, *Luniver Press*, July 2010
11. Xin-She Yang, "A New Metaheuristic Bat-Inspired Algorithm", *Nature Inspired Cooperative Strategies for Optimization, Part of Studies in Computational Intelligence*, Volume: 284, pp. 65-74, 2010
12. SelimYılmaz, uksille and Ecir U. Kuc¸, "A New Modification Approach on BAT Algorithm for Solving Optimization Problems", *Applied Soft Computing*, pp. 259-275, 2015
13. Nehagupta, Kapil Sharma, "Optimizing Intermediate COCOMO Model Using BAT Algorithm", *IEEE*, 2015
14. Ibrahim Muhammad, "A New Model for Software Cost Estimation Using BAT Algorithm", *International Journal of Academic Research in Computer Engineering*, Vol. 1, No. 1, pp. 53-60, September 2016
15. M.Padmaja, Dr D.Haritha, "Optimization of Process Parameters Using Grey-Taguchi Method for Software Effort Estimation of Software Project", *International Journal of Image, Graphics and Signal Processing*, Vol: 10, No: 9, pp. 10-16, Sep 2018
16. Dr D.Haritha, M.Padmaja, "Software Effort Estimation using Meta Heuristic Algorithm", *International Journal of Advanced Research in Computer Science*, Vol: 8, Issue: 5, May-June 2017, pp. 196-201
17. Iqbal Muhammad, Malik Muneeb Abid, "Heuristic Approaches to Solve Traveling Salesman Problem", *IAES, Indonesian Journal of Electrical Engineering*, Vol: 15, No: 2, pp. 390-396, 2015
18. Haiying Dong, Xiaonan Li and Zhanhong Wei, "Substation Fault Diagnosis Based on Rough Sets and Grey Relational Analysis", *Institute of Advanced Engineering and Science, Indonesian Journal of Electrical Engineering*, Vol: 12, No: 2, pp. 1162-1168, 2014

AUTHORS PROFILE

M.Padmaja is working as Assistant Professor, Dept.of CSE, GIT, GITAM Deemed to be University, Visakhapatnam. She is doing Ph. D from JNTUK, Kakinada. Her Research area is Software Engineering, Computer Networks, and Machine Learning.

D. Haritha is a working as Professor, Dept.of CSE, University College of Engineering, JNTUK, Kakinada. Her research area is Image Processing, Networking and Software Engineering.

Retrieval Number: I8242078919/19©BEIESP
DOI:10.35940/ijitee.I8242.078919

Published By:
Blue Eyes Intelligence Engineering
& Sciences Publication