Estimating the Accuracy of Software Cost Using Machine Learning Techniques

Dr.V.Vignaraj Ananth [1] and M.Bhuvaneshwari [2] and Dr.S.Srinivasan [3] and S.Ragul Kesavan [4] and S.Aditya [5]

[1] *Assistant Professor, Department of Computer Science and Engineering*
Thiagarajar College of Engineering,Madurai,India
[2] *Research Student, Department of Computer Science and Engineering*
Thiagarajar College of Engineering,Madurai,India
[3] *Associate Professor, Department of Computer Science and Engineering*
RMD Engineering College,Chennai,India
[4,5] *UG Student, Department of Computer Science and Engineering*
Thiagarajar College of Engineering,Madurai,India
[1] *[e-mail:vignaraj@tce.edu]*
[2] *[e-mail: bhuvimurugan654@gmail.com]*
[3] *[e-mail:ssn.cse@rmd.ac.in]*
[4] *[e-mail: srk71998@gmail.com]*
[5] *[e-mail: aditya77388@gmail.com]*

ABSTRACT

Software cost estimation is the above all else venture to begin any extend These days, cost estimation transforms into a liberal factor in programming developement. Exact estimation can overhauls execution in managing the endeavor plan, human resource assignment, cost estimation, etc. There are a couple of algorithmic and non-algorithmic procedures were accessible. COCOMO II is customarily used model for foreseeing the expense. There are a huge amount of procedures open to assess the product development regardless of the way that a couple of strategies can't manage huge datasets. Accomplishment of programming improvement depends upon cost estimation. It is incredibly difficult to meet unpleasant expense with authentic cost. There are different spending assessment techniques to process cost of the new development and Function point Analysis was one of the method which learning the product part. Primary preferred position is that it can avoid source code botch while picking distinctive programming languages.The key goals of this assessment we are enrolling spending plan of undertaking subject to Machine Learning procedure in which we will figure work purposes of each module. To contrast and evaluate the outcomes and proposed strategy which incorporates classification algorithms, it has been seen that when we have stood out from execution measures with exactness and error rate and by then proposed work gives better results.

Key words*: Cost Estimation, Machine Learning, Classification Algorithms, Error rate, COCOMO-II.*

1.INTRODUCTION

The creating essential for application writing computer programs' what's more, the creating size and inconvenience of these software's, requires a reasonable perspective for looking over the spending plan of programming upgrades in programming creation affiliations. Assessing the spending plan of programming headways has an immense influence in affiliation effectiveness. It could be said that mistaken assessments can cause enormous cash related hardships and in various conditions can sign to indicate expand dissatisfaction. As needs be, the rightness of assessed cost is basic. In any

case, concerning the inventive and hypothetical nature of programming headways, the calculation of cost and time becomes problematic. The system of assessing the spending plan of programming enhancements is basic, so that before exactly on schedule to convey besides, develop an item adventure every alliance initially oversees estimation of animal assets and accessible facilities.Precise cost estimation is critical because of the going with reasons:

• It have the choice to make sense of what advantages for center around the endeavor and how well these benefits will be used.
• It be able touse to assess the impact of changes and upkeep preplanning.
• Projects can be quiet to administer and control exactly when resources are well co-ordinated to authentic requirements.
• Customers acknowledge certified progression expenses to be as per obvious costs.

Cost estimation in programming structuring is ordinarily worried over the money related expend on the imperativeness to make and test the thing, this can in like way join basics outline, upkeep, arranging, coordinating and purchasing additional mechanical assembly, laborers and programming. Different techniques have been conveyed for assessing programming costs for a given undertaking. Figure 1 shows the techniques used in the estimation of programming cost.

Figure 1:Software Cost Estimation Techniques

Cost estimation techniques are prevalently of two sorts:
Algorithmic and non-algorithmic. Algorithmic technique use an equation to process the product cost estimation. The condition is made from models which are made by joining related cost factors. Furthermore, the quantifiable system is used for model turn of events. Non-algorithmic procedure don't use a condition to figure the product quote.
Algorithmic models never give an exact cost estimation. We get exact cost by using the perceptive model. The dataset is progressively huge for the non-algorithmic models. A particularly scattered getting ready instructive file is one of the difficult issues in the farsighted model. The adequacy of the desire procedures is handicapped by virtue of the data dispersal. Precise cost estimation is critical considering the way that it gives the undertaking chief more trust in the cost and time expected to plan similarly for the endeavor.

2.RELATED WORK

2.1:Literature survey on optimization and classification algorithms

In [1] the creator utilizes two methodologies which are profound neural network(DNN) and Ant province optimization(ACO).He made a movement of fundamental and complex system were made to decide CC of 74 copper mining adventures in this. Thusly, the ACO calculation was applied to smooth

out made models. Finally, an ideal half and half model was portrayed with predominant execution than different models was defined.The creator had improved the proposed model with

more preferred position when contrasted with others. A downside is that the limits need incredible idea in assessing and reviewing the CC open-pit minning adventures.

In [2] the creator proposed two-dimensional (2D) and three-dimensional (3D) convolutional neural framework (CNN) getting ready pictures and voxel information methods for a cost estimation of an amassing technique are proposed. Besides, the effects of different voxel objectives, tweaking techniques, and information volumes of the planning CNN are investigated. It was found that appeared differently in relation to 2D CNN, 3D CNN shows incredible execution concerning the relapse issue of a cost estimation and achieves a high application esteem.The disservice is that 2D CNN can't get acquainted with all segment information of the parts

In [3] the creator proposed (CUCKOO-FIS) calculation that fuses two opitmization approach,a meta-heuristic inquiry calculation and Fuzzy Inference System, numerical structure relies upon fluffy rationale. He successfully surveyed on datasets. Part of strategies had proposed previously yet expectation using referenced figuring which utilizes customary procedures showed up result with expanded precision .Disadvantage of this paper is that the information getting ready advances have diminished the idea of the assessment.

In [4] the creator utilized different calculation that improved the result. He utilized different datasets to check. Test show that strategy had been progressively fruitful evaluating exertion and time improvement being developed endeavors than cuckoo search algorithm.Drawback here is extension boundaries worth can't be adequate surveying all undertakings of dataset utilized

In[5]the creator proposed two sorts of ANN-feedforward back-engendering neural system,Elman neural system.It was applied to a dataset that has adventure information from remarkable programming stores that separate additionally light up . Also, this strategy uses more particular execution estimations for instance mean extent of relative blunder (MMRE), mean square error(MSE) and expectation (PRED(x)) to investigate execution.Drawback is that,designed system doesn't act likewise valuable for others accumulated on heterogeneous programming improvement methods .

In [6] the creator proposed a well known procedure for SDEE for instance a half breed model of wavelet neural framework (WNN) and metaheuristic algorithm.Both calculation for instance firefly calculation and bat calculation are used. The presentation of WNN with compromise with all of the metaheuristic strategies investigated. There were varieties wavelet limits—Morlet and Gaussian were used for inception works . The checked techniques likely evaluated on PROMISE SDEE storehouses and seen that real clarified model with referenced strategy outmaneuvered outcomes of programming exertion conjecture incontrast with regular. It doesn't progressed using any metaheuristic procedure. The downside is he envision that presence example of entire endeavor remains unaltered.

In [7] the creator proposed Multi Layered Feed Forward Artificial Neural Network Technique. Capacity of Artificial Neural Network was it will modify a basic gathering relationship between the diverse needy and free factors. The use of ANN (Artificial Neural Network) precisely pushed using Multi Layered Feed Forward Neural Network by utilizing Back Propagation getting ready technique.performance measures were used for execution estimation records. This break down yields propose the suggested strategy will give improved results.

In [8] a hyprid model has been made to combine MCDM (for dealing with weakness) and AI estimation (for managing imprecision) approach to manage envision the exertion even more precisely.Fuzzy systematic chain of command process (FAHP) had used suitably to feature situating. Positions made were facilitated into weighted piece least square suppport vector machine for deciding the attempt. Structured model was actually affirmed from data storehouses.Lot of burdens made by FAHP and the outspread premise function(RBF) divide had achieved continuously exact exertion surveys in relationship with honey bee colonyoptimisation and other.the disadvantage of this paper is, scarcely any key boundaries that ought to be set viably to achieve the best order results.

In [9] the estimation of cost has been finished by utilizing the calculation called bolster vector regression(SVR). To survey the viability, GPS datasets accumulated from 221 panelized private errands over a period of 8 months are used to set up the estimate model and are differentiated and veritable

transportation costs assessed for all intents and purposes. Th strategy gives a continuously

definite transportation cost estimation result for various panelized advancement adventures, and the method improves the appreciation of tremendous extension spatial and transient equipment data while growing the utilization of the GPS data viably available. The creator declare that they have no known battling money related premiums or individual associations that could have appeared to affect the work.

In [10] RF model is organized and progressed tentatively by changing a few estimations in key boundaries. the display was diverged from old style relapse tree. to recognize the most careful strategies, they used three for the most part acknowledged accuracy measures.the results shows the smoothed out self-assertive forest area defeats relapse trees model for different appraisal rules.the disservice of this paper is that the quality of the RF model was somewhat certified.

In [11] to anticipate the capital cost , the model subject to Support Vector Regression (SVR) was made. Generous structure incited insignificant bumble of estimation of CAPEX desire technique. As showed by the results, the abilityto presented framework survey different measures in a wide extent of given breaking point was illustrated. Thusly, as a whole, with a viewpoint on appraisal , it was assessed. The drawback is that SVR calculation isn't reasonable for enormous datasets. SVR doesn't perform well indeed, when the informational index has more commotion.

2.2:Overall summary of Algorithms:

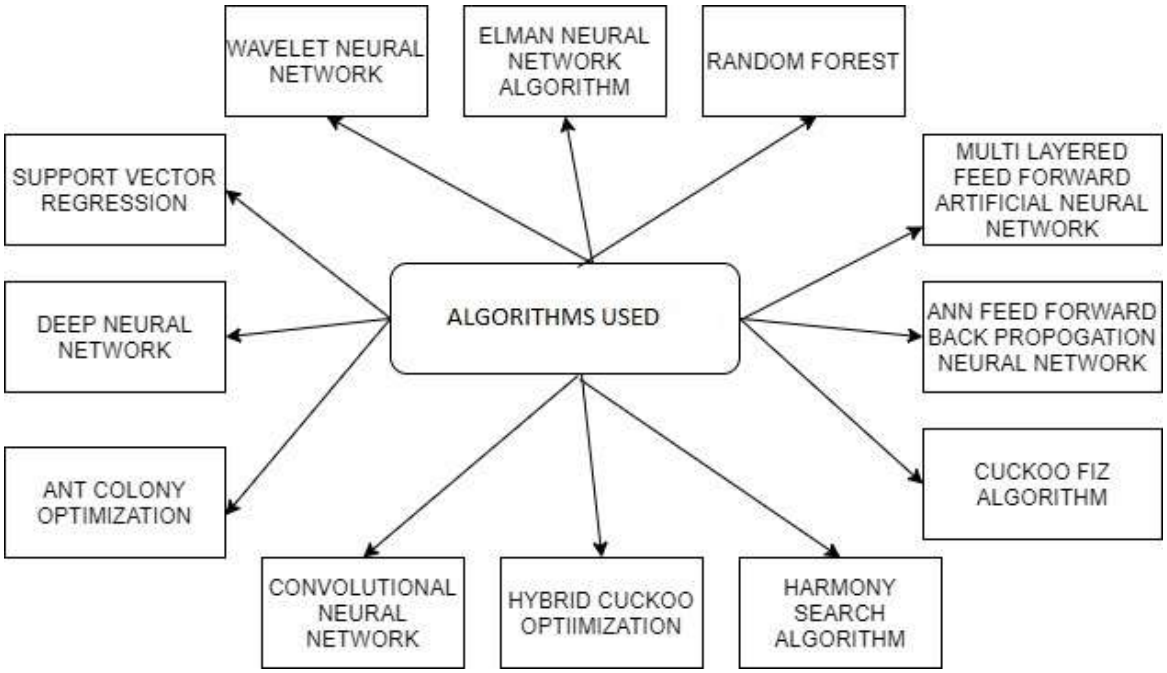

Figure 2:Algorithms used in the Literature Papers

The methods which are used in the literature papers is given in the above Figure 2**.**

3.PROPOSED METHODOLOGY

3.1.Algorithm 1-HCOC(Hybrid Cloud Optimised Cost Organize Algorithm):

Calculation relies upon going with generalsteps:

1. plan the workprocess in privatecloud D;
2. At the point when the idleness is greater than cutoff time:-
   a. pick endeavors to reassign;-
   b.choose resources from open mists ;-

c.Reassign picked tasks in T.

different essential steps of computation are the se-lection of endeavors for reallocate and decision of the re-sources from open source structure the half and half cloud. the past picks that endeavors may have their usage time diminished by using more measure of information from open cloud, last decides execution and expenses related with the new calendar.

HCOC (Hybrid Cloud Optimised Cost Organize Algorithm)
1: X= accessible information in cloud 2: allot An in cloud with the assistance of ZSE 3: when makespan(X)>F& < size(X)do 4: I =i + 1 5: pick hub n i,e ji is more prominent and ni∋ D 6:Y=X 7: D=D∪ d 8: num _group = number of gatherings of hubs in D 9: while num_group > 0do 10:choose asset a from the open mists with the end goal that amnt(i)/num_co(i)*p(i) is minimumand num_co(i)<=num_group. cost n num_group 11: Y=Y ∪ {ai} 12: num_group =num_group −num_coi 13: end while 14: for all n∈ D do 15: reassign n in v∈ Y then EF D is little 16: Recheck ES D s and EF D s 17: end for 18: end while.

Table 1:Algorithm for HCOC

The second line of Algorithm1 which is given in Table 1 characterize the using PCH. Starting there forward, while F is manhandled, computation repeats until cutoff time met or cycle amount equivalent the center amount. Cycle computation picks center point n(i) with ultimate objective Pi is most extraordinary where ni was not planned to bundle D. center n(i) were incorporate to set D, which made out all center points reallocate to individual dag, which figures out which various resources referenced previously. the fact of the matter was, an accentuation is reiterated until the quantity of picked focuses shows up at the amount of gatherings picking open assets that has littlest amnt/numcores with num_co<=num_group. picked assets was then included.it relegates each datum has a place with assets a has a place with Y which gives least outcome.

3.2.Algorithm 2- Raven Roosting Optimization Algorithm:

Raven Roosting Optimization (RRO) is an estimation spurred by the duplicating behavior of ravens anyway it has the issue of inauspicious mixing. This calculation bases on the social lead of ravens and the movement of information between people from the masses in order to find food. Figure3shows the progression of RRO calculation.

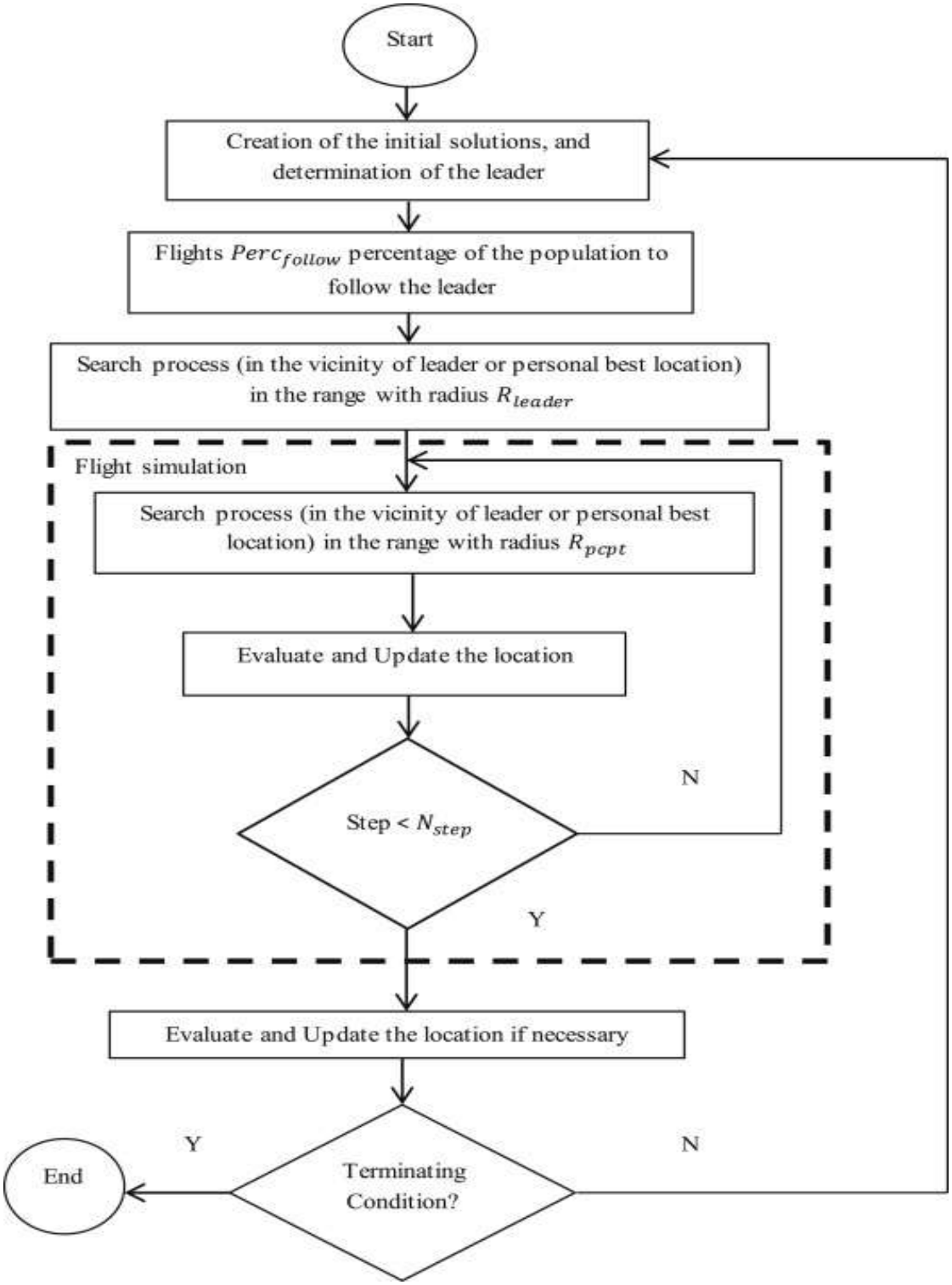

Figure 3: Flowchart for RRO algorithm

Algorithm 2:Raven Roosting Optimization algorithm(RRO)

Haphazardly select a raven site;
rehash
The N scrounging ravens are given out to Y subjective zones in the request space;
figure the wellbeing of each raven territory;
Update the individual best territory of each raven;
The territory of the best game plan is shown as inscription;Enroll Percfollow level of the N foragers from the roosting site which will glance in the locale of the pioneer (inside the extent of range Rleader) and the rest of the ravens will attempt to take off to their own best zones;
Set stage= 0;
while step < Nsteps do
While on the way to its objective (whether or not the objective is the pioneer's district or their own best region), each raven flies for quite a while and look in the area of its recurring pattern position (inside the range with clear Rpcpt);
in case an unrivaled course of action is found than the winged animal's own one of a kind best, by then There is a Probstop chance the raven will stop;
Update the individual best zone;
else It continues flying;
end
stage =stage + 1;
end
For the ravens which finally appear at their objectives (the pioneer's locale or their own best), update their own best territories if indispensable;
to the wellbeing of the zone) and evaluate the status for each forager;
Update zone of the best course of action found as of recently if imperative;
until completion condition

Table 2: **Algorithm raven roosting optimization.**

In this algorithm(in Table 2), the underlying advance is discretionary assurance of a roost in the chase space. Each masses of N people is put self-assertively in the chase space. The wellbeing regard is resolved for the circumstance of each winged creature, and the flying animal with the best game plan is demonstrated as the "pioneer". Some part of the raven masses leaves the roost and follows the pioneer. The fan winged creatures search inside a hyper circle that sets up the range around the region as of late found by the pioneer at which to rummage. People from rest of the masses, fly toward their own best region and continue searching there. Flying ravens are separated into N steps. The length of every movement is picked heedlessly, and each feathered animal's circumstance in flight is invigorated.

3.3.Performance Measures

The precision of the technique is evaluated using particular execution estimations. Tremendous irregularities between certified effort and foreseen effort can influence programming improvement costs.

Mean Relative Error,

$$\text{MRE}= \frac{actual\ effort - predicted\ effort}{actual\ effort} \quad (1)$$

$$\text{MER}= \frac{actual\ effort - predicted\ effort}{predicted\ effort} \quad (2)$$

An appropriate MRE is a value of ≤ 0.25.

Mean Magnitude Relative Error,

$$MMRE = \frac{\sum_{i=1}^{n} MRE_i}{N}, \text{ N=Total number of projects. } MMER = \frac{\sum_{i=1}^{n} MER_i}{N} \quad (3)$$

$$MdMRE = median(MRE_1, \dots, MRE_N) \quad (4)$$

$$MdMRE = median(MRE_1, \dots, MRE_N) \quad (5)$$

$$\text{PRED}=\frac{L}{N} \quad (6)$$

Where,

L=Number of projects whose MRE ≤0.25,

N=Total number of projects.

4.CONCLUSION

This paper has presented a diagram of a grouping of programming estimation strategies, giving a review of a couple of standard estimation models starting at now open. Expectation of programming cost saw essential activity needs the utilization of right systems and methods so as accomplish a not too bad results. fundamental reason,here considered a couple anticipating approaches and a short time later surveyed contemplated proposed calculation. To contrast and survey the outcomes and presented approach and it has been seen that when we have appeared differently in relation to execution measures with precision and mistake rate and by then proposed work gives better results. These two methodologies moreover decrease the pace of slip-up. In plot, this paper contemplated that the joining of these two approachs defeats each and every other system for evaluating effort.

Later on, We may differentiate our proposed methodology and various methodologies for evaluating the effort utilizing AI strategies. In addition, to assess the theory of the revelations, loosen up our proposed approach to manage programming adventure enlightening assortments across various territories.

5.Refrences

5.1.Journal Article:

[1]. Hong Zhang a,b , Hoang Nguyen c, Xuan-Nam Bui d,e , Trung Nguyen-Thoi f,g , Thu-Thuy Bui h , Nga Nguyen i , Diep-Anh Vu j , Vinyas Mahesh k,l , Hossein Moayedi m "Developing a novel artificial intelligence model to estimate the capital cost of mining projects using deep neural network-based ant colony optimization algorithm" Elsevier, journal of Resources Policy 66 (2020) 101604.

*[2].Fangwei Ninga , Yan Shia, *, Maolin Caia , Weiqing Xua , Xianzhi Zhangb, "Manufacturing cost estimation based on a deep-learning method" Elsevier, Journal of Manufacturing Systems 54 (2020) 186–195.*

[3]. Alifia Puspaningrum, Riyanarto Sarno "A Hybrid Cuckoo Optimization and Harmony Search Algorithm for Software Cost Estimation"Elsevier,Journal of Procedia Computer Science 124 (2017) 461–469.

[4]. Anupama Kaushik • Shivi Verma • Harsh Jot Singh• Gitika Chhabra1, "Software cost optimization integrating fuzzy system and COA-Cuckoo optimization algorithm".springer, Int J Syst Assur Eng Manag(2017).

[5].Saurabh Bilgaiyan,Samaresh Mishra, Madhabananda Das, "Effort estimation in agile software development using experimental validation of neural network models". Springer, Int. j. inf. tecnol. https://doi.org/10.1007/s41870-018-0131-2.(2018).

[6].Anupama Kaushik, Niyati Singal, "A hybrid model of wavelet neural network and metaheuristic algorithm for software development effort estimation "Springer, Int. j. inf. tecnol. https://doi.org/10.1007/s41870-019-00339-1(2019).

[7].Poonam Rijwani and Sonal Jain, "Enhanced Software Effort Estimation using Multi Layered Feed Forward Artificial Neural Network Technique".Elsevier Procedia Computer Science 89 (2016) 307 – 312.

[8].Sumeet Kaur Sehra , Yadwinder Singh Brar, Navdeep Kaur,· Sukhjit Singh Sehra, "Software effort estimation using FAHP and weighted kernel LSSVM machine" Springer Soft Computing (2019) 23:10881–10900 https://doi.org/10.1007/s00500-018-3639-2.

[9].SangJun Ahna , SangUk Hanb,∗ , Mohamed Al-Husseinc, Improvement of transportation cost estimation for prefabricated construction using geo-fence-based large-scale GPS data feature extraction and support vector regression..Elsevier Advanced Engineering Informatics 43 (2020) 101012.

*[10].Zakrani abdelalia *, Hain Mustaphaa , Namir Abdelwahedb , "Investigating the use of random forest in software effort estimation". Elsevier Procedia Computer Science 148 (2019) 343–352.*

[11].Hamidreza Nourali, Morteza Osanloo, Mining capital cost estimation using Support Vector Regression (SVR) Elsevier Resources Policy xxx (xxxx) xxx–xxx(2018).

[12].Hidmi, Omar, and betulerdogdusakar. "Software development effort estimation using ensemble machine learning."Int J computcommuninstrumeng 4, no. 1 (2017): 143-147.

[13].Pospieszny, Przemyslaw, BeataCzarnacka-Chrobot, and AndrzejKobylinski. "An effective approach for software project effort and duration estimation with machine learning algorithms." Journal of Systems and Software 137 (2018): 184-196.

[14].Phannachitta, passakorn, and kenichimatsumoto. "Model-based software effort estimation–a robust comparison of 14 algorithms widely used in the data science community." International journal of innovative computing information and control 15, no. 2 (2019): 569-589.

[15]. .Prabhakar, MaitreyeeDutta, and MaitreyeeDutta. "Prediction of software effort using artificial neural network and support vector machine." International Journal of Advanced Research in Computer Science and Software Engineering 3, no. 3 (2013).

5.2.Conference Proceedings

[16].Sarro, Federica, AlessioPetrozziello, and Mark Harman. "Multi-objective software effort estimation."In 2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE), pp. 619-630.IEEE, 2016.

[17].Minku, Leandro L., and Xin Yao. "Ensembles and locality: Insight on improving software effort estimation." Information and Software Technology 55, no. 8 (2013): 1512-1528.

[18].Idri, Ali, Mohamed Hosni, and Alain Abran. "Improved estimation of software development effort using Classical and Fuzzy Analogy ensembles." Applied Soft Computing 49 (2016): 990-1019.

[19].Phannachitta, Passakorn. "Robust comparison of similarity measures in analogy based software effort estimation." In 2017 11th International Conference on Software, Knowledge, Information Management and Applications (Skima), pp. 1-7.IEEE, 2017.

[20] Mustapha, Hain, and NamirAbdelwahed. "Investigating the use of random forest in software effort estimation." Procedia computer science 148 (2019): 343-352.

[21].V.KhatibiBardsiri, D.N.A.Jawawi, S.Z.M.Hashim, E.Khatibi, "Increasing the accuracy of software development effort estimation using projects clustering", IEEE transactions of software engineering, vol 38, no 6, pp 1403-1416, Nov/Dec 2012.

[22].EkremKocaguneli, Tim Menzies and Jacky W.Keung, "On the value of ensemble effort estimation", IEEE Transactions of software engineering, vol 38, no 6, pp 1403-1416, Nov/Dec 2012.

[23].Karel Dejaeger, WouterVerbeke, David Martens and Bart Baesens, "Data Mining Techniques for Software Effort Estimation: A Comparative Study", IEEE Transactions on software engineering, vol 38, no 2, Mar-Apr 2012.

[24].EkremKocaguneli, Tim Menzies, Jacky Keung, David Cok, and Ray Madachy, "Active Learning and Effort Estimation: Finding the Essential Content of Software Effort Estimation Data", IEEE Transactions on software engineering, vol 39, no 8, pp 1040-1053, Aug 2013.

[25].Yeong-SeokSeo, Doo-Hwan Bae and Ross Jeffery, "AREION: Software effort estimation based on multiple regressions with adaptive recursive data partitioning", Information and Software Technology, vol 55, issue 10, pp 1710-1725, Oct 2013

Author's Biographies:

Dr.V.Vignaraj Ananth is a Assistant Professor,Computer Science and Engineering in Thiagarajar College of Engineering,Madurai,Tamilnadu,India. His research interests in Software Effort Estimation,Internet of Things,Data Analytics.He has published Research Article in 15 International Journals and 16 International Conferences.He has secured 98% in Machine Learning:Classifictaion,a course organized by The University of Washington.He has also filed a patent related to Software Estimation.

M.Bhuvaneshwari is a Postgraduate student,Computer Science and Engineering in Thiagarajar College of Engineering,Madurai,Tamilnadu,India.Her research interests in Machine Learning, Software Effort Estimation, Data Analytics. She has published 14 papers in SCI/Scopus journals and 06 International conferences.