

Systematic literature Review-Software Cost Estimation using AI techniques

Sharanpreet Kaur, Computer Science and Engineering, Chandigarh University, Gharuan

Article Info

Volume 82

Page Number: 15984 - 15991

Publication Issue:

January-February 2020

Article History

Article Received: 18 May 2019

Revised: 14 July 2019

Accepted: 22 December 2019

Publication: 28 February 2020

Abstract:

In developing any software whether small or large scale one of the phase that plays a vital part in the development is prior estimation of development cost known as Software cost estimation. In this how much effort as well as cost will be required under the concept of cost and staff. The success of SCE model depends on the equalisation of estimated and real cost. At the time of just starting the development it is very difficult to have an accurate model because as that not much information is there. But as we proceed further information starts increasing which helps for accurate estimation. So in other words it can be said that at the initial stage less accuracy is there but as we go on proceeding there is increase in the accuracy as well. In this paper different AI techniques are discussed which are either alone or hybridised with other techniques for estimating cost in different types of software project.

Keywords: Cost estimation; SCE; Artificial Intelligence; Artificial Neural Network; Case Based Reasoning

INTRODUCTION

Evaluating the cost of any software that needs to be developed is one of the important concept however it is very challenging task to be performed. In the early epoch of software development life cycle, very less number of instructions were there in the software, with the increase in its size the accuracy of estimated cost also increased. There are number of high scale software industries that are unable to cope up with the deadlines that have been assigned to them, in accordance to the cost plan. Now the result of such mismatch is that the projects that have been assigned are aborted, reason being the miscalculation of the cost required for any software development.

In the information system, Software Cost Estimation(SCE) Model is a methodology that is used by any of the respective organization for the purpose of predicting the budget of software development. Reckoning of the resources as well as the duration time required is the vital part for any of the project, but on the same hand there are number of chances that the cost prediction for the whole

development are usually overcome. All the SCE Models have there strong base that depends on software size metric.

In comparison to any other sector, building the SCE model has always gone through the large number of hurdles. All the current techniques that are available largely lay on the amount of information that is available for the project that needs to be developed. As we are moving towards the development of software, our model gets more accurate reason being the information becomes more closer to accurate and reliable values. Estimating the cost of the development process or procedure is always the ongoing process taking into consideration the information that goes on increasing till the process completes.

II. ARTIFICIAL INTELLIGENCE

There are various methods used for estimating the development of software. Out of them one is Artificial Intelligence methods that have shown tremendous results when used for building software

cost estimation model(SCE).In recent years, there has been large amount of efforts that have done for actually making the Software engineering concepts to be get interacted with the techniques coming under Artificial Intelligence. Not only confined to SE the methods available under AI have shown successful results in the field of automatic programming.

Now these efforts have given a way to design of different branch of systems known as Computer Aid Software Engineering (CASE). The objective of CASE is developing an system known as “expert system” whose task will be to generate a design in addition to that generating high-level instructions from the available data set that contains the complete specification detail of the requirement that are demanded by the user. This is in the same spirit as developing a compiler to generate machine code from a high-level language.

A conventional ES will consists of a knowledge base, that comprises facts and rules which will be capturing the reasoning of human experts in a mentioned problem. In addition to this there is also an inference machine, that is capable of deducting new facts from the knowledge base. At the time of development, number of “knowledge-based software assistants “ covering the different phases of cycle, consists These CASE tools that are designed are actually not replacing any technique but on the other hand are enhancements that will help the products that are developed can maintain there standard quality. The design of a CASE tool for developing SCE model is known as Intelligent Assist to Software Cost Estimation (IASCE).

This systematic study has been done for covering different software cost estimation techniques. AI techniques are reviewed and finding their issues with the help of systematically selecting the researching databases, selecting article and deciding their selection or rejection criteria. This study will help the researchers as well as practitioners for selecting suitable software cost estimation model.

For this systematic literature study article is divided into multiple sections which are as:

- ❖ Methodology for literature study
- ❖ Literature Study
- ❖ Algorithms
- ❖ Results
- ❖ Final Conclusion and future work

III. LITERATURE STUDY METHODOLOGY

In this research a systematic literature review has been performed by following a proper procedure. It has been tried to record all the relevant information that falls between the inclusion criteria. The range has been set on the basis of research questionsThe systematic literature review has been utilized to do this research. Hence this analysis centralised on the following cases like Research question identification, Review protocol development, and selection and rejection criterion and Search process.

A. Research Question Identification

Before starting the review process, research questions are formulated and from the accepted papers on different criteria answers to these questions are found which are described below in the table in result section

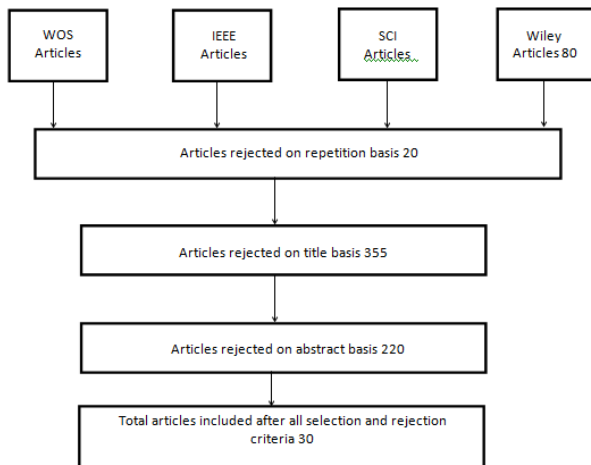
Research questions formulated are:

- RQ-1: IS Artificial technique used for SCE
- RQ-2:Whether AI technique used is single or hybrid with any other technique
- RQ-3:Is the propsed methodology evaluated by any performance metrics
- RQ-4 Publication year of the proposed methodology

B. Review Protocol

Next comes to find out some protocols for considering only the relevant articles as per the domain. Now for this process we have developed the procedure that contains the search process, selection and rejection criteria, quality assessment and data extraction.

C. Search Process



From the diagram above it can be seen that four logical databases (i.e. WOS, IEEE, SCI and Wiley) for completing the literature review in a systematic way. These logical databases maximum result regarding our topic and we have selected the required article via the self-defined process described below. In this the articles have been selected from "2000–2019" from our selected logical databases. In this way the result from logical databases are refined i.e. non-relevant articles are removed in this way.

NO. of articles			
Library	Total Papers	Accepted	Rejected
WOS	85	19	66
IEEE	150	7	143
SCI	40	3	37
Wiley	20	1	19
Total	295	30	265

Table I Summary of database searched and accepted

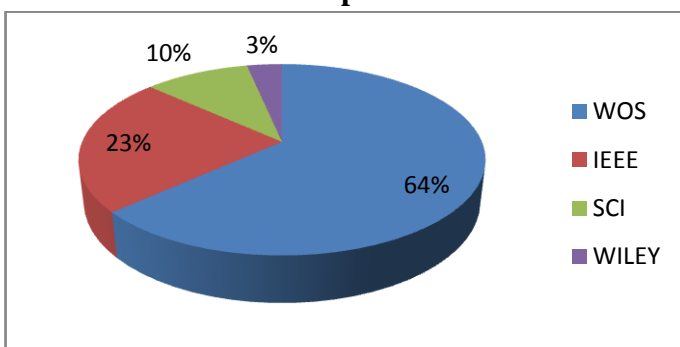


Fig. 1 Statistical graph showing the numerical propotion of all selected papers from the chosen data bases

IV. SELECTION CRITERIA

In this the criteria for choosing the articles are done on the basis of the 6 parameters that are described below:

A. Subject-Relevant

In this those articles are chosen in which cost estimation of a software has been done using any of the AI techniques. The technique can be a single one or the combination of two more techniques. In the combination non AI techniques can also be there or AI techniques can be there. But the condition is that the main role in implementing SCE model should be usage of AI technique.

B. Year Wise Range 2000-2019

While selecting the articles one more parameter has been selected for reducing overall results. That parameter is year wise range. The range that has been chosen in this review from 2000-2019. All the articles below this range have not been selected.

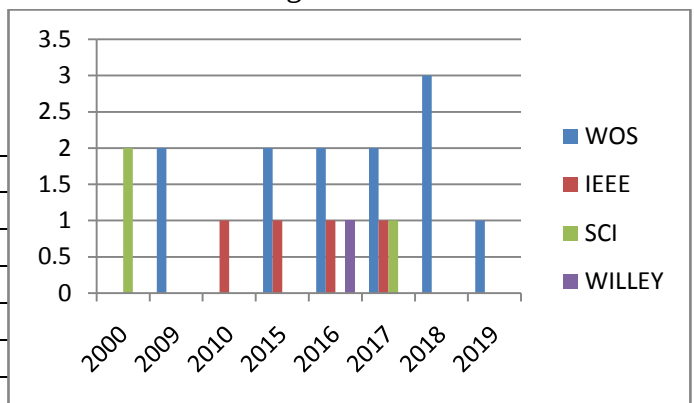


Fig. II- Year wise publications of articles selected

C. Publisher

Selected research work must be published in one of the four renowned scientific databases i.e. IEEE, WOS, SCI and Willey

D. Crucial-Effects

The research work that has been selected in this work is only those articles in which cost estimation has been done only by considering the software under SPM approach. All the proposals that are not contributing to SCE model are not selected.

E. Results-Oriented

In this only research articles are chosen and not the survey articles. From the result section it can be seen that every article that has been chosen has results associated with it. Any other article in which the results are not verified are dismissed from the study.

V. REJECTION CRITERIA

A. Repetition

There are chances that one research article is repeated in more than one database. So those articles which have been repeated then only single article is considered among them and all other are rejected.

B. Title Basis

Another rejection criteria that has been considered is on the basis of the title of the articles. Titles which are not matching with the domain, those articles are rejected.

C. Abstract Basis

There arises a condition when it is not easy to take decision on the basis of title. So at that time reading an abstract becomes essential for getting the required information concerning the research questions built.

VI. LITERATURE STUDY

Developing the SCE Model is one of the vital concept while developing any software. Predicting the SCE model is a challenging task for the company as well as for the client involved. False estimation in SCE model that can be underestimation or overestimation may arise many issues which may affect the reputation of the company. It may also affect that the company may not be able to win the projects due to false estimation. Implementing accurate SCE Model is important because:

- It helps the business to plan accordingly for classifying various projects and then assigning priorities to them.
- It helps in estimating costs accurately which will further be helping in finding resources for the completion of the specific project.

- It also helps in finding out the result if any change is introduced and is also supportive for preplanning.
- One most important concept is that the customers always wants that there should not be much difference between the actual cost and the cost that has been estimated.

Factors on which software measurement depends are:

- Effort estimation of Software
- Time completion duration of project
- Cost estimation of software

VII. ALGORITHMIC

In the algorithmic cost estimation methods, Artificial techniques are used either alone or hybrid with any AI technique or non AI technique for the purpose of estimation [8]. The algorithmic models are studies on large scale like:

Case Based Reasoning

It is the way of finding out one or more than one projects that are identical to the project whose SCE model has to be built. For implementing CBR 3 different parameters are to be considered.

- Similarity measure- It finds out exactness within the past and present projects.
- Number of analogies- It calculates the number which gives the value of the projects that are similar and can be used for implementing SCE.
- Analogy adaptation- Now after the 2nd step of finding the similar projects, SCE is built which will predict the final estimation by calculating different statistic parameters.

Artificial Neural Networks (ANN)

It is a model purely based on primary framework of an individual's neurological system. In this the linked nodes within the neurosoistransfers the desired information. In this computational machine is built which will be proving more successful as compare to the old systems. Now ANN has proven to be very

successful in sector of industry as well as in field of research. ANN can be implemented for approximately all functions in need. It has the capability of doing distinct functions like clustering, regression classification etc. For implementing SCE model it has proven to be a very successful choice of every researcher.

Cuckoo Optimization Algorithm

In year 2009, one of the researcher developed COA whose function was to optimize the solution by replacing the better with the best solution. This algorithm is based on the daily life of a bird named Cuckoo whose feature is to lay eggs on other bird's nest. Now in any technique used the egg in a nest is a solution and another solution will be that bird cuckoo's egg. In this the concept is only that the old solution that was not best is replaced with another solution that has proven to be best. The protocols that are followed in this algorithm are:-

- (i) At the initial step an egg will laid in nest that has been selected.
- (ii) Now the best nest i.e. the best solution replacing all the previous solutions will be selected and continued.
- (iii) Now the probability for throwing the egg i.e. destroying the old solution and finding the new one lies between 0 and 1.

Fuzzy Logic

Fuzzy logic has the concept of number known as fuzzy number. It is a number the value of which is not exact like the usual single numbers. Now this number is thought as a function known as membership function in which domain and range is specified in the closed interval of 0 and 1. Now to the domain a numeric value is given, on the same hand smallest value is represented by 0 and largest value is represented by 1. The same concept is used in software cost estimation for building the SCE model.

Genetic Algorithm

It is a search method used for selecting optimised solutions to find out the problems that revolves

around selecting the solutions on evolutionary biology. This technique has proved excellent on large size as well as complex databases. In SCE model also it is used for finding optimised solutions even if the complex parameters are considered. In this any solution can be solved whether parameters are stiffed or easy.

Technique/Database	WOS	IEEE	SCI	Wiley	Total	Percentage
Case Based Reasoning	1,17	3,4			4	13.33%
Artificial Neural Networks	2,8,11,13,14,16,19	2,7	2,3	1	11	36.37%
Cuckoo Optimization Algorithm	2,4,6		1		4	13.33%
Fuzzy Logic	3,4,7,10,12,15	6			6	20%
Ant Colony Optimisation	5,8				2	6.67%
Genetic Algorithm	9,14,18	1			4	13.33%
Harmony Search Algorithm		1	1		2	6.67%

Table II- Statistic usage of machine learning algorithms

VIII PERFORMANCE MEASURES

All the techniques used in all the articles for implementing SCE Model are evaluated in comparison with actual values. The performance measures that are widely used Mean Relative Error(MRE), Magnitude Mean Relative Error(MMRE) and Prediction(Pred.). These are the maximum usable performance measures in number of articles.

But in addition to this other performance measures are Median of Magnitude of Relative error (Md MRE), Magnitude of Error Relative to estimate (MER), Mean Absolute Error (MAE), Mean Magnitude of Error Relative to estimate (MMER), Mean Square Error (MSE), Root Mean Square Error (RMSE) and Standard Deviation (SD) etc.

IX. RESULT

Study no.	RQ-1	RQ-2	RQ-3	RQ-4
S1	YES	Single	YES	2019
S2	YES	Hybrid	YES	2018
S3	YES	Single	YES	2018
S4	YES	Hybrid	YES	2017
S5	YES	Single	YES	2017
S6	YES	Hybrid	YES	2018
S7	YES	Hybrid	YES	2016
S8	YES	Hybrid	YES	2016
S9	YES	Single	YES	2015
S10	YES	Hybrid	YES	2015
S11	YES	Both	YES	2013
S12	YES	Hybrid	YES	2012
S13	YES	Hybrid	YES	2012
S14	YES	Both	YES	2012
S15	YES	Hybrid	YES	2011
S16	YES	Both	YES	2011
S17	YES	Hybrid	YES	2009
S18	YES	Single	YES	2009
S19	YES	Hybrid	YES	2009

Table III Answers to Research questions for WOS Database

Study no.	RQ-1	RQ-2	RQ-3	RQ-4
S1	YES	Hybrid	YES	2016
S2	YES	Single	YES	2011
S3	YES	Single	YES	2010
S4	YES	Hybrid	YES	2014
S5	YES	Hybrid	YES	2015
S6	YES	Hybrid	YES	2017
S7	YES	Hybrid	YES	2014

Table-IV Answers to Research questions for IEEE Database

Database	Study no.	RQ-1	RQ-2	RQ-3	RQ-4
SCI	S1	YES	Hybrid	YES	2017
	S2	YES	Hybrid	YES	2000
	S3	YES	Hybrid	YES	2000
Wiley	S1	YES	Hybrid	YES	2016

Table- V Answers to Research questions for SCI and Willey Database

X. CONCLUSION

After performing this systematic survey the conclusion that are drawn from this study are:-
Implementing SCE Model is not an easy task as it deals with number of parameters which cannot be ignored, otherwise the chances of getting the accurate results gets decreases. On the other hand all the techniques that are implemented are having number of pros and cons. Some SCE models that were built proves to be successful for enormous

projects and on the other hand some gave good results for small scale projects.

The cost estimation is a complex task and all the methods proposed have multiple advantages and disadvantages.

- All the factors, which may or may not influenced needs to be considered developing SCE Model.
- The prognostic attributes from the data base needs to be considered for having constant and best results. For selecting these attributes backward input approach can be used.
- In this study, multiple AI techniques are considered and the best efficiency is observed among many of them.
- Artificial techniques when used performed better maximum times when compared with traditional method.They have given good percentage accurate results either when used standalone or hybrid with any other technique.

The future work in cost estimation is proposing a novel model using hybrid approach. There is still a lot to do with these AI techniques for getting an accurate and reliable SCE Model.

REFERNCES

- [1] V. S. Desai And R. Mohanty, "Ann-Cuckoo Optimization Technique To Predict Software Cost Estimation," In *2018 Conference On Information And Communication Technology (Cict'18)*, 2018.
- [2] Z. H. Wani, K. J. Giri, And R. Bashir, "A Generic Data Mining Model For Software Cost Estimation Based On Novel Input Selection Procedure," *Int. J. Inf. Retr. Res.*, Vol. 9, No. 1, Si, Pp. 16–32, 2019.
- [3] Y. Pratama And E. Rasywir, "Automatic Cost Estimation Analysis On Datawarehouse Project With Modified Analogy Based Method," In *2018 International Conference On Electrical Engineering And Computer Science (Icecos)*, 2018, Pp. 171–176.
- [4] A. Kaushik, S. Verma, H. J. Singh, And G. Chhabra, "Software Cost Optimization Integrating Fuzzy System And Coa-Cuckoo Optimization Algorithm," *Int. J. Syst. Assur. Eng. Manag.*, Vol.

- 8, No. 2, 2, Pp. 1461–1471, Nov. 2017.
- [5] G. Kumar And P. K. Bhatia, “Empirical Assessment And Optimization Of Software Cost Estimation Using Soft Computing Techniques,” In *Advanced Computing And Communication Technologies*, 2016, Vol. 452, Pp. 117–130.
- [6] Z. H. Wani And S. M. K. Quadri, “Artificial Bee Colony-Trained Functional Link Artificial Neural Network Model For Software Cost Estimation,” In *Proceedings Of Fifth International Conference On Soft Computing For Problem Solving (Socpros 2015)*, Vol 2, 2016, Vol. 437, Pp. 729–741.
- [7] M. Algabri, F. Saeed, H. Mathkour, And N. Tagoug, “Optimization Of Soft Cost Estimation Using Genetic Algorithm For Nasa Software Projects,” In *2015 5th National Symposium On Information Technology: Towards New Smart World (Nsitnsw)*, 2015.
- [8] A. Kaushik, A. K. Soni, And R. Soni, “A Hybrid Approach For Software Cost Estimation Using Polynomial Neural Networks And Intuitionistic Fuzzy Sets,” *Int. J. Comput. Appl. Technol.*, Vol. 52, No. 4, Pp. 292–304, 2015.
- [9] A. Kaushik, A. K. Soni, And R. Soni, “Radial Basis Function Network Using Intuitionistic Fuzzy C Means For Software Cost Estimation,” *Int. J. Comput. Appl. Technol.*, Vol. 47, No. 1, Pp. 86–95, 2013.
- [10] E. Papatheocharous And A. S. Andreou, “A Hybrid Software Cost Estimation Approach Utilizing Decision Trees And Fuzzy Logic,” *Int. J. Softw. Eng. Knowl. Eng.*, Vol. 22, No. 3, Pp. 435–465, May 2012.
- [11] A. Kaushik, A. K. Soni, And R. Soni, “An Adaptive Learning Approach To Software Cost Estimation,” In *2012 National Conference On Computing And Communication Systems (Ncccs)*, 2012, Pp. 3–8.
- [12] T. S. Sethi, C. V. M. K. Hari, B. S. S. Kaushal, And A. Sharma, “Cluster Analysis And Pso For Software Cost Estimation,” In *Information Technology And Mobile Communication*, 2011, Vol. 147, P. 281+.
- [13] Y. F. Li, M. Xie, And T. N. Goh, “A Study Of Mutual Information Based Feature Selection For Case Based Reasoning In Software Cost Estimation,” *Expert Syst. Appl.*, Vol. 36, No. 3 Part 2, Pp. 5921–5931, 2009.
- [14] J. S. Pahariya, V. Ravi, And M. Carr, “Software Cost Estimation Using Computational Intelligence Techniques,” In *2009 World Congress On Nature & Biologically Inspired Computing (Nabci 2009)*, 2009, Pp. 848–853.
- [15] J. K. Lee And K. T. Kwon, “Software Cost Estimation Using Svr Based On Immune Algorithm,” *10th Acis Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distributed Comput. Snpd 2009, Conjunction With Iwea 2009 Weacr 2009*, Pp. 462–466, 2009.
- [16] F. S. Gharehchopogh, “Neural Networks Application In Software Cost Estimation: A Case Study,” *Inista 2011 - 2011 Int. Symp. Innov. Intell. Syst. Appl.*, Pp. 69–73, 2011.
- [17] G. Jiang, Y. Wang, And H. Liu, “Research On Software Cost Evaluation Model Based On Case-Based Reasoning,” *Proc. - 2010 2nd Wri World Congr. Softw. Eng. Wcse 2010*, Vol. 2, Pp. 338–341, 2010.
- [18] Q. Liu, S. Shi, H. Zhu, And J. Xiao, “A Mutual Information-Based Hybrid Feature Selection Method For Software Cost Estimation Using Feature Clustering,” *Proc. - Int. Comput. Softw. Appl. Conf.*, Pp. 27–32, 2014.
- [19] S. P. Singh And A. Kumar, “Software Cost Estimation Using Homeostasis Mutation Based Differential Evolution,” *Proc. 2017 11th Int. Conf. Intell. Syst. Control. Isco 2017*, Pp. 173–181, 2017.
- [20] F. S. Gharehchopogh, R. Rezaii, And B. Arasteh, “A New Approach By Using Tabu Search And Genetic Algorithms In Software Cost Estimation,” *9th Int. Conf. Appl. Inf. Commun. Technol. Aict 2015 - Proc.*, Pp. 113–117, 2015.
- [21] L. V. Patil, R. M. Waghmode, S. D. Joshi, And V. Khanna, “Generic Model Of Software Cost Estimation: A Hybrid Approach,” *Souvenir 2014 Ieee Int. Adv. Comput. Conf. Iacc 2014*, Pp. 1379–1384, 2014.
- [22] A. Puspaningrum And R. Sarno, “A Hybrid Cuckoo Optimization And Harmony Search Algorithm For Software Cost Estimation,” *Procedia Comput. Sci.*, Vol. 124, Pp. 461–469, 2017.
- [23] A. Lee, C. H. Cheng, And J. Balakrishnan, “Software Development Cost Estimation: Integrating Neural Network With Cluster

- Analysis,” *Inf. Manag.*, Vol. 34, No. 1, Pp. 1–9, 1998.
- [24] B. Samson, D. Ellison, And P. Dugard, “Software Cost Estimation Using An Albus Perceptron (Cmac),” *Inf. Softw. Technol.*, Vol. 39, No. 1, Pp. 55–60, 1997.
- [25] S. S. Wang, “A Study Of Automated Estimating Of Software Cost,” 1993.
- [26] M. A. Saleem *Et Al.*, “Systematic Literature Review Of Identifying Issues In Software Cost Estimation Techniques,” *Int. J. Adv. Comput. Sci. Appl.*, Vol. 10, No. 8, Pp. 341–346, Aug. 2019.
- [27] A. Idri, A. Zahi, E. Mendes, And A. Zakrani, “Software Cost Estimation Models Using Radial Basis Function Neural Networks,” *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, Vol. 4895 Lncs, Pp. 21–31, 2008.
- [28] K. V. Kumar, V. Ravi, M. Carr, And N. R. Kiran, “Software Development Cost Estimation Using Wavelet Neural Networks,” *J. Syst. Softw.*, Vol. 81, No. 11, Pp. 1853–1867, Nov. 2008.
- [29] B. Zhang And R. Zhang, “Evaluation Model Of Software Cost Estimation Methods Based On Fuzzy-Grey Theory,” *Proc. - 4th Int. Conf. Internet Comput. Sci. Eng. Iicse 2009*, No. 5, Pp. 52–55, 2010.
- [30] D. Manikavelan And R. Ponnusamy, “Software Cost Estimation By Analogy Using Feed Forward Neural Network,” *2014 Int. Conf. Inf. Commun. Embed. Syst. Icices 2014*, No. 978, Pp. 1–5, 2015.
- [31] H. D. S. Budiono And M. Lassandy, “Cost Estimation Model In Early Stage Of Design Through Implant Miniplate Complexity Process Index,” In *3rd International Tropical Renewable Energy Conference Sustainable Development Of Tropical Renewable Energy (I-Trec 2018)*, 2018, Vol. 67.
- [32] E. E. Miandoab And F. S. Gharehchopogh, “A Novel Hybrid Algorithm For Software Cost Estimation Based On Cuckoo Optimization And K-Nearest Neighbors Algorithms,” *Eng. Technol. Appl. Sci. Res.*, Vol. 6, No. 3, Pp. 1018–1022, Jun. 2016.
- [33] T. R. Benala, S. Dehuri, S. C. Satapathy, And S. Madhurakshara, “Genetic Algorithm For Optimizing Functional Link Artificial Neural Network Based Software Cost Estimation,” In *Proceedings Of The International Conference On Information Systems Design And Intelligent Applications 2012 (India 2012)*, 2012, Vol. 132, P. 75+.
- [34] S. Sabbagh Jafari And F. Ziaaddini, “Optimization Of Software Cost Estimation Using Harmony Search Algorithm,” *1st Conf. Swarm Intell. Evol. Comput. Csiec 2016 - Proc.*, Pp. 131–135, 2016.
- [35] D. Kashyap And A. K. Misra, “Software Cost Estimation Using Particle Swarm Optimization In The Light Of Quality Function Deployment Technique,” In *2013 International Conference On Computer Communication And Informatics*, 2013.
- [36] Y. F. Li, M. Xie, And T. N. Goh, “A Study Of Genetic Algorithm For Project Selection For Analogy Based Software Cost Estimation,” In *2007 Ieee International Conference On Industrial Engineering And Engineering Management*, Vols 1-4, 2007, Pp. 1256–1260.