ISSN: 0974-5823 Vol. 7 No. 1 January, 2022

International Journal of Mechanical Engineering

IMPROVED ARTIFICIAL NEURAL NETWORK MODEL FOR SOFTWARE PROJECT COST ESTIMATION IN EARLIER STAGE

Shweta.KR
Ph.D Research Scholar, Dept. of Computer Science, Chikkanna Govt. Arts College, Tirupur-2, Tamil Nadu.

Dr. T. Latha Maheswari
Associate Professor, Dept. of Computer Science & Engineering, Sri Krishna College of Engineering & Technology, Coimbatore.

Dr. S. Duraisamy
Assistant Professor, Dept.of Computer Science, Chikkanna Govt. Arts College, Tirupur-2, Tamil Nadu.

Abstract

Machine learning is the latest buzzword that plays a crucialrole in various field of medical, research and an industrial application. It is difficult to weigh the real values or worth of software. One of the best way to estimate the software development cost, effort, size and time is based on the previous experience of software development. To measure the standard worth of the software, as a unit of software worth, machine-learning algorithms are utilized for increasing the end user's satisfaction level through correct and prompt computation of software cost and effort estimation. In this research work, a novel cost assessment for software project management using improved artificial neural network model was developed. Two publicly available datasets with different machine learning algorithms are compared and results show that the proposed model having the high accuracy and low error rate in earlier stage prediction of the cost and effort assessment.

Keywords: Cost Assessment, IANN, Machine learning, Accuracy, End user.

I. INTRODUCTION

TheResearch and consultative enterprise in Artificial Intelligenceat Emerj, defines Machine Learning (ML) is the science and technology discipline of buildingprocesserscram and performance like individuals by nurturinginformation and evidence without being obviouslyautomatic. Machine learning algorithms such as decision tree, random forest, linear regression, logistic regression and K-nearest neighbours are more powerful and applied in prediction of software cost estimation. Machine learning algorithms are used in the variety of business world to solve variety of problems especially face recognition used for security. Voice recognition is the common area in these days using Siri and Cortana that is an example of ML. Healthcare industry is another big area where ML adapted in a very big role like diagnostics, analysis of X-ray, MRI images and increasingly for shortage of doctors ML and Artificial intelligence (AI) are used. Weather forecasting and Netflix are another area actually come up with new use cases. Most of the users using snapshots to apply filters to their photos. Snapshot actually does this using facial recognition which in turn using machine learning for predicting the end user photo even though there are multiple faces available in snapshots. Machine learning algorithm detects the features on user face like the nose, the eyes and it knows exactly where user eyes are and nose are available and accordingly apply the filters.

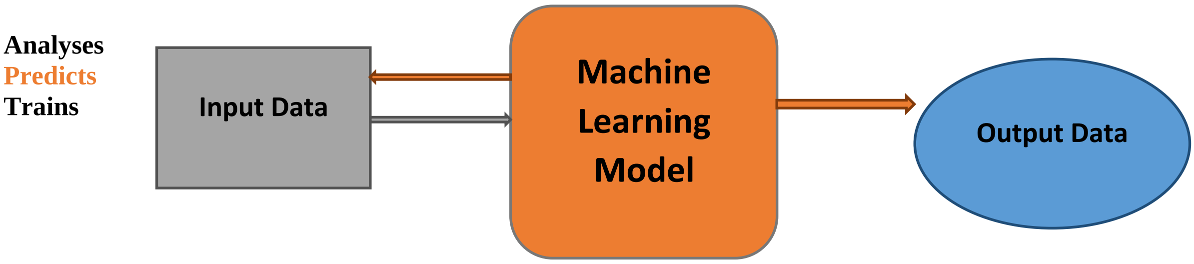


Figure 1. Steps involved in Machine learning model

Step1: Data Gathering

Machine learning needs past historical sufficient data analysing information for supervised learning.

Step2: Data Pre-processing

The raw data cannot be used directly. So data pre-processing is necessary.

Step3: Choose Model

Identifying what type of model, what type algorithm is going to be used.

Step4: Train Model

Before training the model it is called as blank model and once train the model need to test the model.

Step5: Test Model

It is used to give the training results with minimum error.

Step6: Tune Model

If it has minimum error but timely fine tune, the model is necessary and it is the iterative process to improve the performance of the model.

Step7: Prediction

After fine tune the model, then it is deployed to prediction.

Types of ML:

Three basic types of machine learning algorithms are Supervised Learning (SL), Unsupervised Learning (UL) and Reinforcement Learning (RL).

1) Supervised Learning (SL)

Supervised learning is used when user have historical data and user have labelled data, which means that user know how the data is classified, user know the classes, values (labels) in regression and classification with historical data's.

2) Unsupervised Learning (SL)

User do not have past historical labelled data. Association and clustering is used for form the clusters, new classes may be to overcome the specified drawback.

3) Reinforcement Learning (RL)

The system learns pretty much from system scratch. There is an agent and there is an environment. It allows the agent to automatically determine the ideal behaviour within a specific context and it has to maximize the performance.

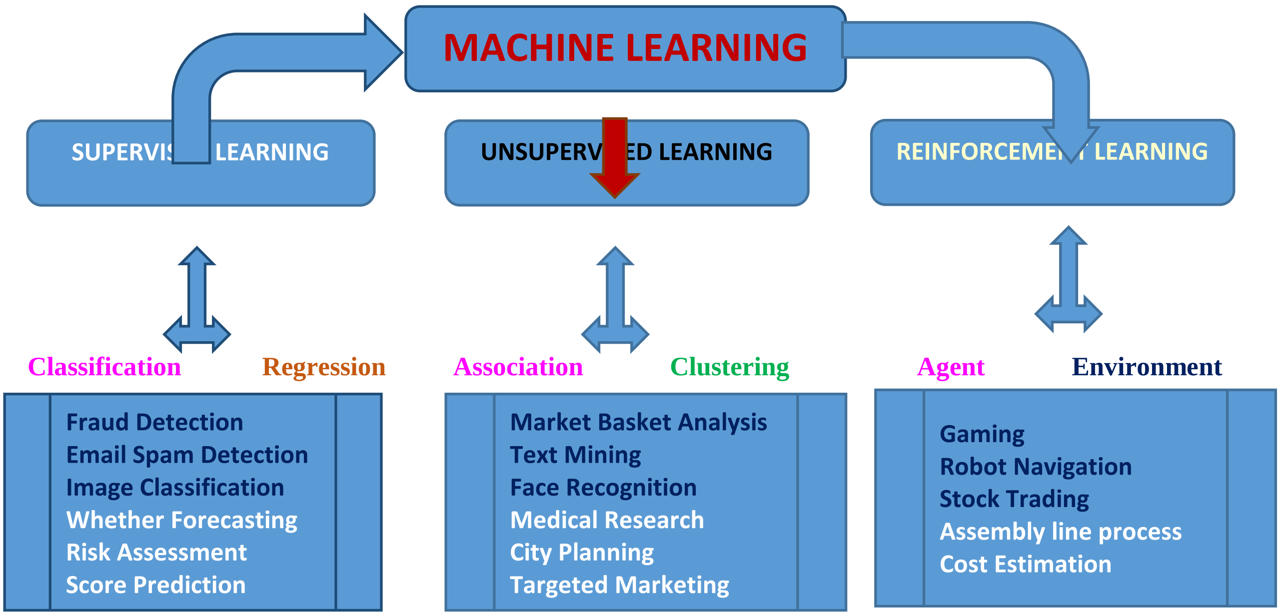


Figure 2. Types of Machine learning model

Expedient and precise task cost assessment is basic to guarantee consumer loyalty and rehash business. Notwithstanding, it stays one of the most difficult errands in programming, particularly when the tasks managed are mind boggling, broad, and in the applied stage. The progressions in Artificial Intelligence and Machine Learning innovations are changing the conventional strategy of programming cost assessment into an adaptable and perceptiveapproach. A convincing use case is a wise answer for recognize and take out copies, break down and pinpoint uncertainty, and cycle RFPs in changed configurations with negligible human intercession, for productive and assisted venture cost assessment.

The largest expenditure for a software development project is for the salaries of the development staff. Effort estimation involves approximating the level of effort needed for a project in terms of person-weeks, person months, or hours. When person months are used, the average monthly salary of a software engineer multiplied by the estimated person-months is the dollar amount the project costs. If 10 person-months are estimated, and the project timeline spans 5 months than the estimate would cover two fulltime development staff members. Effort estimation is needed for project managers to plan and schedule project activities. The effort estimation approaches using machine-learning algorithms are very prominent for producing results with high accuracy.

Today, contrasting known properties and a reference class of tasks, and influence project administrators' involvement with managing comparable undertakings previously perform programming advancement cost assessment. The current technique can lead to incorrect gauges because of the accompanying obstructions:

- Failure to factor in vague prerequisites
- Idealism predisposition that sets in because of the presumption of all assets working at most extreme usefulness
- Huge wastage of time and exertion in distinguishing and sifting through copy gauges
- Cushioning appraisals to make sufficient space to take care of sudden expenses
- Skirting key expense drivers because of tension from partners to carry out the figures carelessly
- Conditions on innovative progressions and changing economic situations

II. Literature Review

Gouthaman et al (2021) said as of late, programming project disappointments have been expanding because of absence of arranging and financial plan limitations [1]. Dareen et al (2020) applied expense assessment for development the board utilizing ANN model [2]. Fangwei Ning et al (2020) proposed three-dimensional CNN for viable expense assessment [3]. Erik Matel et al (2019) recommended quotes permit project directors to assess the attainability of activities and control costs viably [4]. Mahmood et al (2019) forms a product cost assessment model utilizing AI approach [5]. Michal et al (2018) applied neural organization calculation to cost development [6]. Przemys et al (2017) proposed different AI calculations for exertion and time assessment 7]. T.M.S. Elhag et al (1998) proposed ANN for development of programming projects [8]. Richa Yadav et al (2016) examinations the accomplishment of any venture undertaking is characterized by further developed amount and cost assessment strategy that works with ideal use of assets [9]. Murat Gunaydin et al (2004) examines the utility of neural organization system to beat cost assessment issues in beginning stages of building configuration processes [10].

III. Methodology

Artificial Neural Network (ANN)

The neural network is an iterative approach and it's solve problem through an iteration.

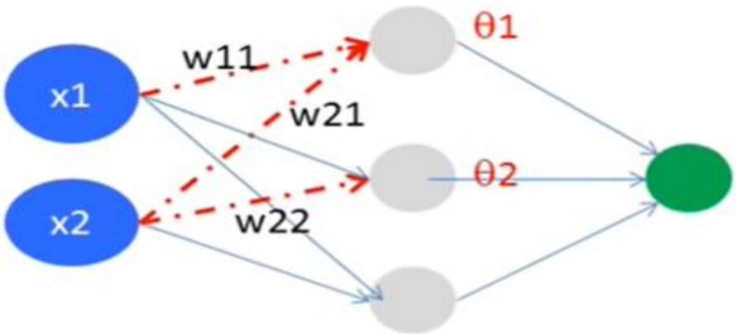


Figure 3. ANN

Need for cost estimation:

After training our proposed model, the user need to see how badly estimated model is performing while accuracy function tell us how well our model is performing, they do not provide us with insight on how to better improve them. In Figure 4, the proposed IANN is used to minimizing the errors occurred in the input data and increasing the performance of calculating the weights. Hence, anyone need a correctional function, which can help user compute when proposed model is the most accurate.

Improved ANN (IANN)

Improved Artificial Neural Network is a machine-learning algorithm, which takes multiple inputs, runs then through an algorithm and essentially sums the output of the different algorithms to get final output. Cost estimation of software project of an IANN will be the sum of minimized errors in each layer. This is done by finding the error at each layer first and then summing the individual error to get total error.

Step1: Keep input nodes equal to number of independent variable.

Step2: Assign random weight to synapsis (links)-usually start with random number between (-0.5-0.5).

Step3: Assign random biases-usually start with random number between (0.5-0.5).

Step4: Calculate output $Y_P = Y_{<predicted>}$

Step5: Calculate error $Y_a = Y_{<actual>} - Y_P$

Step6: Back propagate error to adjust weights

Gradient Descent (GD)

It is an algorithm, which is used to optimize the cost estimation with minimum error rate and maximum accuracy. It is used to find the minimum value of error possible in cost estimation.

$$CE = \frac{1}{j}\sum_{i=0}^{j}(d\theta a^{j}) - b^{i})^{2} \quad (3.1)$$

$$GD = \theta_{CE} = \theta_{CE} - \alpha \frac{\alpha CE}{\alpha\theta} \quad (3.2)$$

$$GD = \theta_{CE} - \frac{\alpha}{j}\sum_{i=0}^{j}((d_{\theta}(a^{j}) - b^{i})^{2} \quad (3.3)$$

$$GD = \theta_{CE} = \frac{\alpha CE}{\alpha\theta} \quad (3.4)$$

$$CE = \frac{1}{j}\sum_{i=0}^{j}(a^{j} - mb^{i}+b))^{2} \quad (3.5)$$

In the above equation, m denotes weight of the layer and b denotes the constant bias.

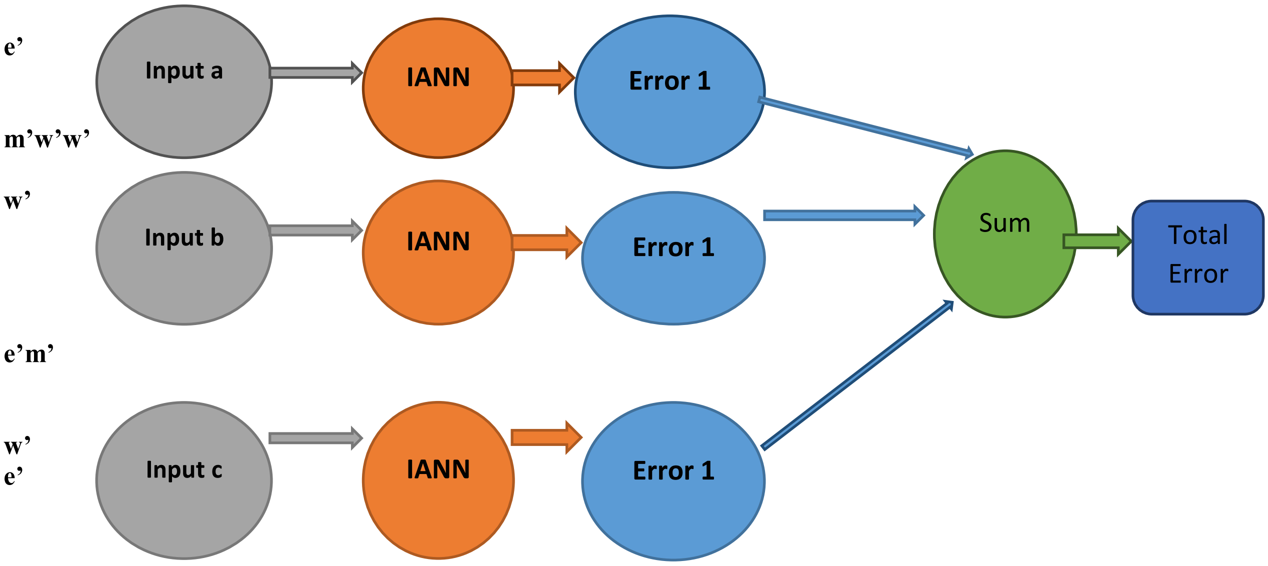


Figure 4. IANN

Figure 4, the corrected weight of a layer to further tweak the weight of the layer before that and get the desired output for IANN called backpropagation.

Dataset Description

The datasets utilized in this work are ready for programming tests that are accessible publically. Two datasets will be utilized to test and look at the Machine Learning procedures,the first dataset is called Usp05-ft and the second is Usp05. Dataset (Usp05-ft) contains 76 occasions and comprises of the 15 credits. AppType, Method, DBMS, TeamSize, AppExpr, ToolExpr, Tools, Lang, UFP, DataOut, DataFile, DataEn, IntComplx, Effort and ID are attributes used in Usp05-ft data set. Notwithstanding the 15 credits, the second dataset (Usp05) contains 203 occurrences and comprises of two additional qualities which are object type and requirements functionality percentage.

IV. Results and Discussion

Weka device rendition 3.8.5 has been utilized to assess the pre-owned calculations. The 10 folds' cross-approval procedure is utilized to prepare the information on 90% and test it on the excess 10% until the entire information (100%) is being utilized as a test information through 10 cycles. Six sorts of classifiers have been utilized in Weka to test the 13 calculations. The pre-owned classifiers are Naïve Bayes, Bayes Network, J48, Decision Table, Random woods and Improve Artificial Neural Network (IANN). Kappa Statistic, Mean Absolute Error, Root Mean Squared Error, Relative Absolute Error and Root Relative Squared Error are utilized to assess the exhibitions of proposed model. The 6 strategies have been tried on two datasets, the first is (Usp05-ft) with 15 credits and 76 examples. The subsequent one (Usp05) is with 17 credits and 203 occasions.

Table 1. Usp05-ft Dataset for tested 6 methods.

Model	Correctly Classified Instances	Incorrectly Classified Instances	Kappa Statistic	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Root Relative Squared Error

Naïve Bayes	60	12	0.6073	0.0596	0.2336	40.9403	89.8998
Bayes Network	61	11	0.6612	0.0585	0.2106	40.1818	81.0467
J48	62	10	0.5652	0.0593	0.2021	40.7916	77.7835
Random Forest	64	8	0.6624	0.0549	0.1587	37.7177	61.0988
Decision Table	67	5	0.8108	0.1281	0.2194	88.0492	84.4327
IANN	69	3	0.8897	0.0196	0.1157	13.4956	44.5338

Table 2. Efficiency of Usp05-ft Dataset for tested 6 methods.

Model	Accuracy	Precision	Recall	F-Measure	MCC	ROC	PRC
Naïve Bayes	83.3333%	0.962	0.909	0.935	0.750	0.935	0.962
Bayes Network	84.7222%	0.960	0.873	0.914	0.696	0.971	0.989
J48	86.1111%	0.873	0.932	0.938	0.759	0.938	0.980
Random Forest	88.8889%	0.887	0.893	0.940	0.789	0.994	0.998
Decision Table	93.0560%	0.948	0.973	0.884	0.956	0.971	0.981
IANN	95.8333%	0.979	0.987	0.982	0.923	0.998	0.999

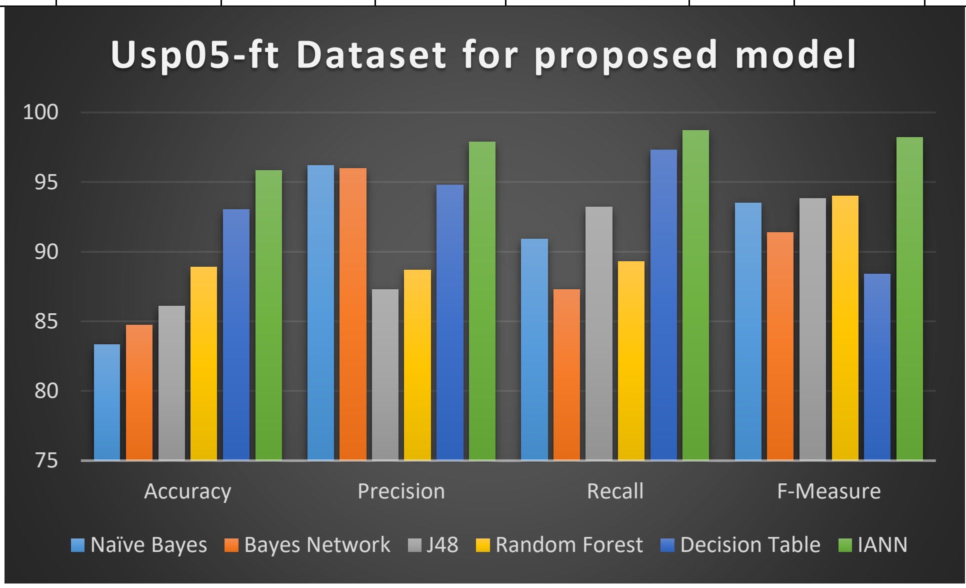


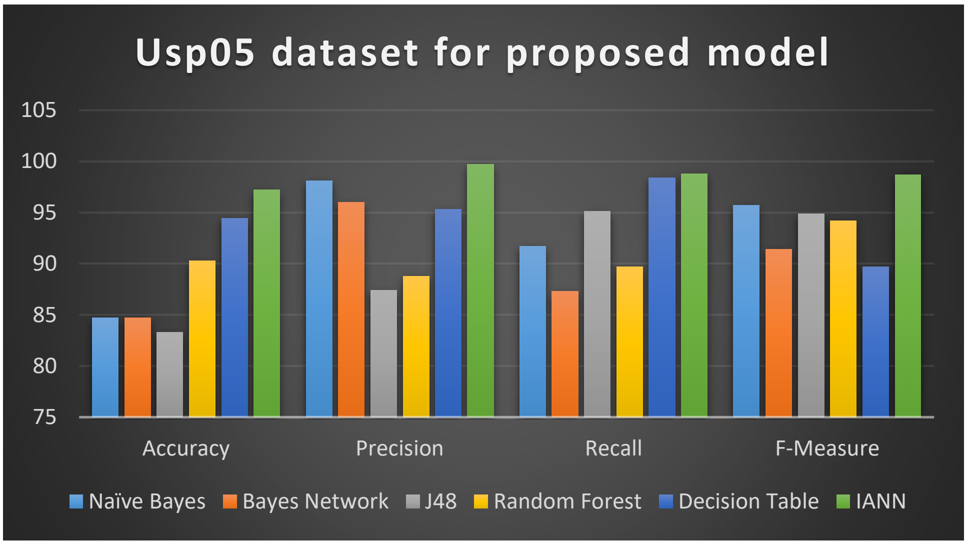
Table 3. Usp05 Dataset for tested 6 methods.

Model	Correctly Classified Instances	Incorrectly Classified Instances	Kappa Statistic	Mean Absolute Error	Root Mean Squared Error	Relative Absolute Error	Root Relative Squared Error
Naïve Bayes	61	11	0.6123	0.0524	0.2436	39.4413	87.7871
Bayes Network	61	11	0.6612	0.0585	0.2106	40.1818	81.0467
J48	60	12	0.5431	0.0586	0.1021	42.1416	76.7825
Random Forest	65	7	0.6647	0.0687	0.1489	36.6197	60.0088
Decision Table	68	4	0.8008	0.1179	0.2119	58.0492	54.4467
IANN	70	2	0.8442	0.0191	0.1157	11.1453	24.2647

Table 4. Efficiency of Usp05 Dataset for tested 6 methods.

Model	Accuracy	Precision	Recall	F-Measure	MCC	ROC	PRC

Naïve Bayes	84.7222%	0.981	0.917	0.957	0.787	0.942	0.974
Bayes Network	84.7222%	0.960	0.873	0.914	0.696	0.971	0.989
J48	83.3333%	0.874	0.951	0.949	0.759	0.929	0.990
Random Forest	90.2777%	0.888	0.897	0.942	0.823	0.997	0.998
Decision Table	94.4444%	0.953	0.984	0.897	0.963	0.980	0.988
IANN	97.2222%	0.997	0.988	0.987	0.927	0.998	0.999



Conclusion

In this research work, six Machine Learning algorithms have been evaluated using two datasets (Usp05-ft) & (Usp05). The evaluation criteria used in this work are Kappa Statistic (KS), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), Relative Absolute Error (RAE)and Root Relative Squared Error (R^2SE). The aim of the proposed model is to predict the cost and effort using dataset attributes and compare them with the actual cost and effort in order to measure the error using different criteria. The proposed IANN model achieved the high accuracy in the first experiment using (Usp05-ft) dataset and second experiment using (Usp05). In future different datasets with different machine learning algorithms are proposed for giving the more accurate cost and effort estimation in software project management.

References

[1] Gouthaman P, Suresh Sankaranarayanan, Prediction of Risk Percentage in Software Projects by Training Machine Learning Classifiers", computers and electrical engineering (2021)
[2] DareenRyied Al-Tawal, MazenArafah. GhalebJalilSweis, "A model utilizing the artificial neural network in cost estimation of construction projects in Jordan", ECAM (2020)
[3] Fangwei Ning, Yan Shia, MaolinCai, Weiqing Xu, Xianzhi Zhang, "Manufacturing cost estimation based on the machining process and deep learning method", journal of manufacturing systems (2020)
[4] Erik Matel, FaridaddinVahdatikhaki, SiavashHosseinyalamdary, Thijs Evers,HansVoordijk, "An artificial neural network approach for cost estimation of engineering services", International journal of construction management (2019)
[5] Mahmood Mohd Al Asheeri, Mustafa Hammad, "Machine Learning Models for Software Cost Estimation" International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT) IEEE (2019)
[6] Michal Juszczyk ,AgnieszkaLesniak, Krzysztof Zima, "ANN Based Approach for Estimation of Construction Costs of Sports Fields", wiley (2018)
[7] Przemys_awPospieszny ,BeataCzarnacka-Chrobot , Andrzej Kobylin_ski, "An effective approach for software project effort and duration estimation with machine learning algorithms,The Journal of Systems & Software (2017)
[8] T.M.S. Elhag, A.H. Boussabaine, "An Artificial Neural System for Cost Estimation of Construction Projects", In: Hughes, W (Ed.), 14th Annual ARCOM Conference, 9-11 September Vol. 1, 219-26. (1998)
[9] Richa Yadav, Monica Vyas, Vivekanada Vyas, Sanket Agrawal, "Cost Estimation Model (Cem) for Residential Building using Artificial Neural Network", (IJERT) Vol.5(1) (2016)
[10] H. Murat Gunaydin, S. ZeynepDogan, "A neural network approach for early cost estimation of structural systems of buildings", International journal of project management (2004)

[11] Durgabai, R. P. L., and P. Bhargavi. "Pest management using machine learning algorithms: a review." International Journal of Computer Science Engineering and Information Technology Research (IJCSEITR) 8.1 (2018): 13-22.
[12] Nawalagatti, Amitvikram, and R. KOLHE PRAKASH. "A Comprehensive Review on Artificial Intelligence Based Machine Learning Techniques for Designing Interactive Characters." International Journal of Mathematics and Computer Applications Research (IJMCAR) 8.3 (2018): 1-10.
[13] Danthala, S. W. E. T. H. A., et al. "Robotic Manipulator Control by using Machine Learning Algorithms: A Review." International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) 8.5 (2018): 305-310.
[14] Kumar, B. Satish, and Y. Kalyan Chakravarthy. "Prediction Of Optimal Torques From Gait Analysis Applying The Machine Learning Concepts." International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) 9. 4, (2019), 685-698
[15] Chandrabai, Drt, And Drk Suresh. "Assessment Of Machine Learning Techniques For Gold Price Predictions." International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) 10.3 (2020) 5879-5886
[16] Hagar, Abdulnaser A., Deepali G. Chaudhary, And A. L. I. A. Al-Bakhrani. "Big Data Analytic Using Machine Learning Algorithms For Intrusion Detection System: A Survey." International Journal of Mechanical and Production Engineering Research and Development (IJMPERD) 10.3 (2020) 6063-6084