

Regression Based Software Project Effort Estimation with Reusability for Startups



O. Rajalakshmi alias KarthiKa, C. Rekha

Abstract: Prediction of effort for software development is the major task for the effective management of any software industry. At present, there are several estimation techniques and tools that are available for a software project. However, choosing the correct and efficient effort prediction of a particular software project is the most demanding task for a start ups. This study focuses on Multiple Linear regression for Software Effort Estimation technique which will provide higher classification accuracy of a software project and thus improves the prediction efficiently. The main aim of the technique is to help developers in startups to classify the project compared to existing software metrics and thereby increasing the software effort prediction rate.

Keywords: Software, Effort, Multiple Linear regression, Projects, startup.

I. INTRODUCTION

Software effort estimation is a basic assignment performed by project designers in the early times of programming improvement. The information of various programming project attributes are required to stay away from imprecision and uncertainty which influences the achievement of software projects. The loose programming exertion estimation causes a economic drop and intrudes on project target. The effort estimate is likewise utilized for distinguishing the item size with clear programming improvement lifecycle, and process which demonstrates displaying, development and examining of programming. The basic commitment of software project cost and effort estimation is to systematically approximate the required remaining task at hand and its resulting costs in software system life cycle. Software cost estimation is a troublesome activity which requires data about various key properties which impact the aftereffects of software projects freely and its exhibition.

Both Software cost and exertion estimation utilized in IT industries for improving the improvement of software projects are in a proficient way. The underlying advance for project requesting, accounting and arranging are programming estimation. At the point when assets and plans are excessively unenthusiastic, business events are lost and it additionally prompts significant misfortunes. The Software project estimation comprises of three phases specifically: size estimation, exertion estimation, cost estimation. Software effort estimation is the way toward recognizing the most down to earth utilization of exertion required for broadening or supporting the advancement of the software project. Effort estimates are utilized to assess the effort in terms of Person-Months (PM) for the Software improvement work components of the Work Breakdown Structure (WBS). Figure 1.1 shows the Software Project Estimation process.

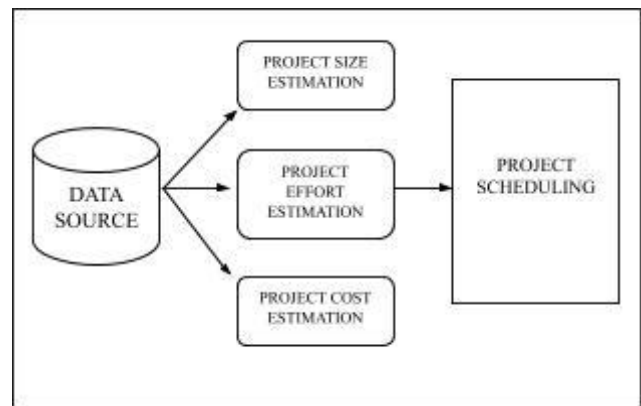


Fig. 1. Software Project Estimation process

As shown in Figure 1.1, project estimation incorporates size, exertion and cost estimations for improving software improvement to finish a project. Various Linear Regression is utilized for structuring the estimation of a quantitative variable relying upon its direct relationship with one or more interpreter. It is also utilized to quantify the connection between two or more descriptive variables and a response variable through a fitting direct condition to the handy information. The Linear regression models impacts on linear correlation dependent variable and every translator. Different Linear Regression system is regularly utilized,

- At the point when the quantity of cases is significantly raised than the quantity of predicted parameters
- During the conditions of consistent data behavior, few missed information and

Manuscript received on February 10, 2020.
Revised Manuscript received on February 20, 2020.
Manuscript published on March 30, 2020.
Correspondence Author

O. Rajalakshmi@karthika, Department of Computer Applications, Madurai Kamaraj University College, Madurai, India.
C. Rekha, Department of Computer Applications, Madurai Kamaraj University College, Madurai, India.

© The Authors. Published by Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP). This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

- When the little number of independent variables is sufficient for distinguishing yield factors to achieve an interpretable portrayal.

Regression techniques are utilized as a typical model and examination device for evaluating the software effort to keep up the forthcoming software projects. In regression techniques, the primary presumptions to be estimated are,

- Linearity for deciding the connection between two predictors for adequate recognition of data behavior and
- Mistake part which is independent and typically dispersed variable with stable fluctuation and the mean value is zero.

The determination of independent variables in the Multiple Linear Regression model is mostly of two types. The first is through the incorporation of whole independent variables which are applicable and the different utilizes stepwise strategies, for example, forward regression, backward regression, and stepwise regression.

II. LITERATURE REVIEW ON SOFTWARE ESTIMATION MODELS

Ricardo de A Araujo et al. (2017) described Multilayer Dilation-Erosion-Linear Perceptron (MDELP) for describing the issues of software development effort estimation. Vahid Khatibi Bardsiri et al. (2013) structured a Localized Multi-Estimator Model (LMES). In LMES, software projects were classified and relied upon their attributes. Karel Dejaeger et al. (2012) dissected tree/rule-based models, linear techniques, nonlinear methods, and estimation techniques. Zia et al. (2011) presented another product cost estimation method. It was utilized for software projects implemented with the guide of a component-based fourth-age language condition. The model was directed by using the experimental information and effectiveness was likewise assessed contrary to traditional model used in specific conditions. Ziauddin et al. (2012) introduced a Software Effort Estimation Model utilizing Fuzzy Logic and molecule swarm enhancement. The Fuzzy sets were utilized for designing uncertainty and imprecision in effort estimation. Anupama Kaushik et al. (2016) built up the product cost estimation model utilizing Functional Link Artificial Neural Networks (FLANN). The Intuitionist Fuzzy C-Means Clustering (IFCM) was utilized to achieve an expanded clustering precision for upgrading the product expectation results with the assistance of FLANN.

Geetika Batra and Kuntal Barua (2013) reviewed cost estimation strategies and metrics. The fundamental point of this investigation was to distinguish the issues and clarifications by presenting a steady methodology for the minimum usage of cost and effort. Tirimula Rao Benala et al. (2012) examined Computational Intelligence (CI) techniques utilized for estimation of programming cost. CI strategies were spoken to as fit methods since an enormous number of speculation factors were connected with programming ventures. Suri and PallaviRanjan (2012) dissected various classes of programming cost estimation models and techniques. An appraisal system was utilized in expanding

project tracking and evaluation. Raja Ramesh Merugu and Venkat Ravi Kumar Dammu (2012) looked into delicate figuring strategies. A top to bottom assessment of programming and task estimation techniques in the business alongside benefits and negative marks were examined. Dragicevic Srdjana et al. (2017) presented a Bayesian Network (BN) model for prediction of exertion. BN organize utilized in the arranging stage and the parameter estimation, for example, MMRE, duration, and multifaceted nature was routinely gotten from a dataset. Rahul Kumar Yadav and Niranjan (2013) investigated Fuzzy Logic applications in Software improvement exertion estimation models. Various benefits of Fuzzy Logic were talked about for improving the expectation models. Mohammad Azzeh et al. (2011) considered a product venture likeness measure and Fuzzy number adjustment system. Sandeep Kad & Vinay Chopra (2012) considered delicate figuring techniques for building a model to upgrade the exertion estimation process. Various parameters of the Constructive Cost Model (COCOMO) II were fuzzified for steady and precise exertion guess. Brajesh Kumar Singh and A. K. Misra (2012) talked about the improved variant of the Constructive COSt MOdel (COCOMO). This form of the COCOMO was utilized as an algorithmic model and the significance of the counterfeit neural system strategy was confirmed. A statistical and machine learning technique was brought about by Ruchika Malhotra and Ankita Jain (2011) for foreseeing the software effort. The one referenced above is accomplished by performing Linear Regression, Artificial Neural Network, Decision Tree, Support Vector Machine, and Bagging on datasets of software project datasets. However, those procedures neglected to give an increasingly exact estimation on a software project and was an exceptionally hard undertaking. So as to determine this issue, Multiple Linear Regression system was proposed. In this way, Multiple Linear Regression for Software Effort Estimation strategy is acquainted with guarantee the best exertion forecast of the product extends in a useful way. Because of its power and flexibility, the proposed procedure is a basic strategy for measurable investigation in different fields. It is the undertaking of the project manager to pick the best device for a software project. In the wake of acquiring the product exertion, venture director arranges the undertaking improvement, controls the expense and guarantees the quality more accurately. Zahra Shahpar et al. (2016) depicted a hereditary calculation for choosing highlights in features in software project effort estimation.

III. MULTIPLE REGRESSION FOR SOFTWARE EFFORT ESTIMATION

Ruchika Malhotra and Ankita Jain (2011) formulated a statistical and machine learning technique for anticipating the exact software effort. Performing Linear Regression, Artificial Neural Network, Decision Tree, Support Vector Machine, and Bagging on programming software project prompted the achievement of the method guaranteeing the best exertion expectation of the product extends in a useful way. Inferable from its power and flexibility, the proposed MLR-SEE procedure is a fundamental methodology of measurable examination in different fields.

To pick the best instrument for a software project is the obligation of the task supervisor in a startup. In the wake of getting the product effort, it is the undertaking of project manager to compose the project development, controlling the cost and ensuring the quality all the more precisely. The fundamental target of the proposed system is to assist engineers to isolate the task contrasted with existing programming measurements and in this way expanding the software effort prediction rate. Figure 1.2 shows the proposed Multiple Linear Regression Model for programming exertion estimation.

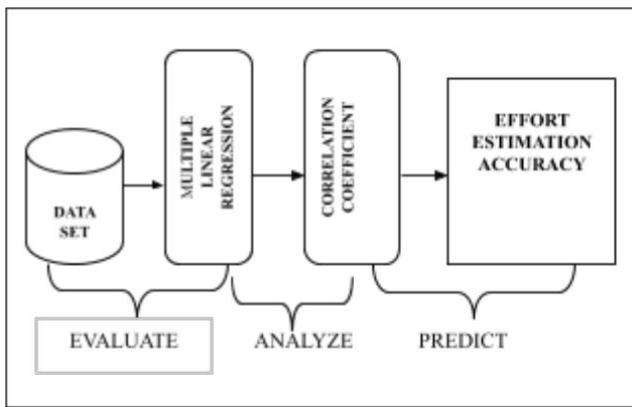


Fig. 1. Proposed Multiple Linear Regression Model
Proposed Multiple Linear Regression Model

The Proposed Multiple Linear Regression strategy effectively predicts the most authentic effort for programming development proposition by contrasting earlier programming plans. In this manner, this system is utilized to illuminate the product issue based on flawed, uncertain and loud info. Initially the quantity of project activities is taken as input the dataset. Next, the system assists with surveying the response variable and the arrangement of predictors in terms of linear function and afterward to group the undertaking as far as assessing the correlation coefficient. This advances processing the relationship among all the qualities of the venture so as to acquire exertion estimation with lesser information shakiness. Subsequently, we can get the better arrangement results with lesser mistake work. Therefore, improving the prediction accuracy in software project effort estimation is cultivated by the proposed method.

Considering the inputs as software projects “ $P_i = P_1, P_2, P_3, P_4, P_5, \dots, P_n$ ” in the dataset.

The proposed Multiple Linear Regression system is concocted for anticipating the response factors by relegating at least one anticipated factors for effort estimation. The registered characteristics including the product size, exertion, etc are utilized for improving the unwavering quality in software effort estimation. During the characterization procedure, the proposed method is utilized for registering relationship coefficient esteem so as to get the best effort forecast correlation coefficient value in order to acquire the best effort prediction. The simple linear regression models are of the the form $Y = a + bX$ Where

$Y =$ dependant variable (that is the variable that goes on y axis)

$X =$ independent variable (plotted on x axis)

$b =$ slope of the line

$a =$ y intercept

The proposed strategy calculates the regression coefficients with at least one predictor variables that measure the best exertion estimation of response variables of software project. The mathematical formula of multiple linear regression for effort estimation is composed as follows.

$$Y_i = W_0 + W_1 X_{i1} + W_2 X_{i2} + \dots + W_k X_{im} + E_i$$

Y_i denotes the i th component of the response variable

where

$$i = 1, 2, 3, \dots, n$$

$W_0 =$ Constant value

$X_{im} =$ i th component of m predictor variable

$W_1, W_2, W_3, W_4, \dots, W_k$ is the regression coefficients

E_i refers to the error or residual, where $E_i \sim N(0, \sigma^2)$ N refers to the number of distributions with mean value 0 and variance σ^2

Then it chooses to calculate the correlation coefficients so as to minimize the error function by using the least squares method. By applying the above proposed technique, an error function is defined as the difference among the actual value

Y_i and the estimated value a_i which is expressed as follows.

$$E_i = Y_i - a_i$$

which is evaluated as the difference between the estimator targeted values and the estimator observed values on a software project. Besides, the least square regression method is used to evaluate the minimum Sum of Squared Error (SSE) by proposed technique which is given as follows.

$$SSE = \sum_i E_i^2 = \sum_i (\hat{Y}_i - a_i)^2$$

SSE denoted as the sum of squared error between the actual and estimated value. When there is presence of multicollinearity, the estimated coefficients are correlated with each other. Thus, the proposed technique is estimated by correlation coefficient with lesser error function.



Followed by this, the correlation coefficient is used to calculate the relationship among the characteristics of a software project which is mathematically represented by

$$CC = \frac{Cov(U,V)}{\rho_U \rho_V}$$

CC alludes to the correlation of coefficient which is variable amount and it fluctuates between - 1 and 1. On the off chance that the connection esteem is certain (i.e., 1), at that point it gets solid direct connection and in this way lesser exertion is required for running the software project. Or there will be consequences, on the off chance that the connection esteem is negative (i.e., - 1), at that point there is a need of more exertion for the product undertaking to run Based on the correlation coefficient result, the order accuracy for effort estimation is improved by proposed MLR-SEE. This helps in the powerful utilization of software project for acquiring the higher forecast.

IV. EXPERIMENTAL RESULT

The trial work is directed by taking an alternate number of activities are taken as info which shifts from 5 to 50.

V. CONCLUSION

An improved proposed Multi Linear Regression procedure is presented forestimating the best exertion for the product proposals. The starting stage being the exertion expectation of programming is among the customer and business enterprise. Hence, the proposed technique orders the project based on the correlation coefficient evaluated. Then, the connection coefficient is utilized to show up at the strength of the relationship between project attributes for effort estimation with the help of the proposed technique. Thus, the proposed technique significantly earns upgraded order precision and thereby prediction of software project exertion gets increased in an efficient way.

VI. FUTURE RESEARCH

Future work can be acquired to perform software project effort estimation for additional increasing the presentation of proposed systems. Sooner rather than later, the usage of ensemble classification technique gives more grounded characterization results with least mistake and time utilization during the software project effort estimation. Future work centers around joining diverse enhancement methods to discover a way further developed numerical capacity for investigating present ventures with verifiable undertakings for giving the best effort prediction.

REFERENCES

1. A. Asongu, S., & Nwachukwu, J. C. (2018). Comparative human development thresholds for absolute and relative pro-poor mobile banking in developing countries. *Information Technology & People*, 31(1), 63-83.

2. Agarwal, A., Xie, B., Vovsha, I., Rambow, O., & Passonneau, R. (2011, June). Sentiment analysis of twitter data. In *Proceedings of the workshop on languages in social media* (pp. 30-38). Association for Computational Linguistics.
3. Barnes, S. J., & Corbitt, B. (2003). Mobile banking: concept and potential. *International Journal of Mobile Communications*, 1(3), 273-288.
4. Boire, R. (2014). Social media analytics. In *Data mining for managers* (pp. 181-184). Palgrave Macmillan, New York.
5. Dahlberg, T., Mallat, N., Ondrus, J., & Bachfischer, A. (2006). Mobile payment market and research-Past present and future. *Helsinki mobility roundtable*.
6. Dahlberg, T., Mallat, N., Ondrus, J., & Zmijewska, A. (2008). Past, present and future of mobile payments research: A literature review. *Electronic commerce research and applications*, 7(2), 165-181.
7. Fan, W., & Gordon, M. D. (2014). The power of social media analytics. *Communications of the ACM*, 57(6), 74-81.
8. He, W., Zha, S., & Li, L. (2013). Social media competitive analysis and text mining: A case study in the pizza industry. *International Journal of Information Management*, 33(3), 464-472.
9. Ho, T., & Do, P. (2018). Social Network Analysis Based on Topic Model with Temporal Factor. *International Journal of Knowledge and Systems Science (IJKSS)*, 9(1), 82-97.
10. Hu, M., & Liu, B. (2004, August). Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 168-177). ACM.
11. Kim, C., Mirusmonov, M., & Lee, I. (2010). An empirical examination of factors influencing the intention to use mobile payment. *Computers in Human Behavior*, 26(3), 310-322.
12. Krishnan, S. (2014). *The Power of Mobile Banking: How to Profit from the Revolution in Retail Financial Services*. John Wiley & Sons.
13. Mallat, N. (2007). Exploring consumer adoption of mobile payments— A qualitative study. *The Journal of Strategic Information Systems*, 16(4), 413-432.
14. NITI Aayog (2018), *Implementation of digital payment systems in India*, Retrieved at Nov 06, 2018 from <http://niti.gov.in/content/implementation-digital-payment-systems-india>.
15. Ondrus, J., & Pigneur, Y. (2006). Towards a holistic analysis of mobile payments: A multiple perspectives approach. *Electronic commerce research and applications*, 5(3), 246-257.
16. Pak, A., & Paroubek, P. (2010). Twitter as a corpus for sentiment analysis and opinion mining. In *LREc* (Vol. 10, No. 2010, pp. 1320-1326).
17. Rathore, A. K., Kar, A. K., & Ilavarasan, P. V. (2017). Social Media Analytics: Literature Review and Directions for Future Research. *Decision Analysis*, 14(4), 229-249.
18. Rosenthal, S., Farra, N., & Nakov, P. (2017). SemEval-2017 task 4: Sentiment analysis in Twitter. In *Proceedings of the 11th International Workshop on Semantic Evaluation (SemEval-2017)* (pp. 502-518).
19. Saif, H., He, Y., & Alani, H. (2012). Semantic sentiment analysis of twitter. In *International semantic web conference* (pp. 508-524). Springer, Berlin, Heidelberg.
20. Singh, P., Sawhney, R. S., & Kahlon, K. S. (2018). Sentiment analysis of demonetization of 500 & 1000 rupee banknotes by Indian government. *ICT Express*, 4(3), 124-129.
21. Soleymani, M., Garcia, D., Jou, B., Schuller, B., Chang, S. F., & Pantic, M. (2017). A survey of multimodal sentiment analysis. *Image and Vision Computing*, 65, 3-14.
22. The Economic Times (2018), *Mobile banking catching on fast: Here's the proof*. Retrieved at Oct 08, 2018 from <https://economictimes.indiatimes.com/wealth/>

