https://dx.doi.org/10.22161/ijaers/3.12.38 *ISSN: 2349-6495(P) | 2456-1908(O)*

Analysing the calibration curve of software Project work in Different stage Using Isotonic Regression Classifier

Dr. S.Veni[1], Aparna Srinivasan[2]

[1]Associate Professor and Head, Department of Computer Science, Karpagam Academy of Higher Education, Karpagam University, Coimbatore, Tamil Nadu, India

[2]Research Scholar, Department of Computer Science, Karpagam University, Coimbatore, Tamil Nadu, India

Abstract— *In this research work focuses the Software estimation time has been an vital role everyone in software industry. Software effort and cost estimation plays an important role in software development, often determining the success or failure of contract negotiation and project execution based on the estimation plan. Predicting the success of software projects based upon information related to estimation task and actual task is deemed to be one of the vital activities in software engineering research. In this research work, isotonic regression classifier can be implemented for analyzing the estimation quality based on estimation value and actual value to achieve better result from the model based upon information related to Estimation task and actual task in IT project management.*

Keywords— Estimation day, Actual day, Estimation effort, Actual Effort, Estimation Cost, Actual Cost, Isotonic Regression classifier for predicting the task.

I. INTRODUCTION

Calibration is one of the major applications of regression modelling. An Isotonic classifier regression is an idea of calibration [3], [5], [6], [7] is normally associated with measuring the device. Measuring the equipments has errors and calibration is a process by which the errors are determined by comparing the readings with real values. After the calibration, the error table can be used to predict true value from the value indicated. The relationship between an estimated value and an actual value for calibrating the estimation model can be equated and measure the regression line by the process of monitoring estimated task and actual task performance analyzing in difference stage of project development such as planning, designing, building stage. It enables organizations to improve the quality of management decision making by ensuring that reliable and secure information and data is available throughout the Software lifecycle.

II. DATA FOR RESEARCH

2.1 Objective

In this paper, a decision making database of machine learning is proposed, that a set of possible estimator are working together to estimate the project completion task in it. Further, predicting the success of software projects based upon information related to estimated task and actual task is deemed to be one of the vital activities in software engineering research.

- To proposed equation of isotonic regression classifier for analyzing the performance to identify a better result applying in software metric.
- In this research work focus to apply calibration of intuitive model for analyzing the estimation quality based on estimation value and actual value to achieve better result from the model based on isotonic regression classifier.

2.2 Data Collection in Planning stage, Designing Stage, Building stage

In this research work, the evaluation can be carried out in three stage of project development such as planning, designing, building stages of the project development. The collection of fifty projects of data can be imported into the data repository file for analyzing the predicted [2] true values in calibration curve. Similarly, the data of other two stage of project work can be collected and it can be analyzed based on Isotonic regression classifier [1] [3] [7].

https://dx.doi.org/10.22161/ijaers/3.12.38 *ISSN: 2349-6495(P) | 2456-1908(O)*

Planning Estimated day	Planning Actual day	Planning Effort Planned	Planning Actual Effort	Planning Planned Cost	Planning Actual Cost
22	22	437.7142857	437.7142857	20627.73714	20627.73714
22	22	328.2857143	328.2857143	15470.80286	27406.20286
15	16	1382.964	1382.964	49095.222	50372.157
15	15	1222.1975	1222.1975	43388.01125	45536.22275
8	9	365	365	12967.52	13114.6
18	18	101.4	101.4	4263.53	4361.53
18	18	169	169	7700.32	7777.32
11	11	259.995	259.995	10220.4	10220.4
19	19	363.7	363.7	18171.378	18171.378
12	12	363.7	363.7	18171.378	18171.378
16	16	636	636	22896	22896
15	15	173.33	173.33	6228	6228
4	5	244.196	244.196	10207.18	10207.18
5	6	66.1	66.1	3335.898	3335.898
19	19	182	182	7708.57	7708.57
12	12	34.67	34.67	883.64	917.64
4	5	34.67	34.67	1762.42	1830.24
4	4	100	100	3550	3550
12	12	128.3	128.3	6366.812	6366.812
15	15	446.8	446.8	20311.12	20311.12

Designing					
Designing Estimated day	Designing Actual day	Designing Est Effort	Designing Actual Effort	Designing Planned Cost	Designing Actual Cost
22	24	603.8411765	646.84	32434.54118	32434.54118
22	24	422.6888235	822.68	22704.17882	44757.18882
74	78	1382.964	1518.934	49095.222	53922.157
74	78	1222.1975	1382.7105	43388.01125	49086.22275
18	25	470.6	495.88	52399.18	53300.3
11	12	111.8	112	4854.53	4894.53
11	15	186.33	187	8489.83	8499.83
20	21	365.23	365.23	15247.1	15247.1
47	47	545.55	545.55	27257.067	27257.067
28	28	475.6	496.88	52400.18	53600.3
20	20	729	729	26244	26244
21	22	693	606	24948	21816
18	19	488.393	488.393	20414.37	20414.37
12	12	99.15	99.15	5003.85	5003.85
26	26	273	273	11562.85	11562.85
11	11	86.67	81	2209.09	2069.84
11	12	86.67	81	4406.05	4249.94
7	8	459	465	16294.5	16507.5
11	12	192.45	192.45	9550.218	9550.218
24	25	670.2	670.2	30466.68	30466.68
18	18	500.393	500.393	26414.37	26414.37

Building					
Estimated day	Actual day	Planned Cost	Actual Cost	Effort Planned	Actual Effort
38	50	20627.73714	20627.73714	603.8411765	646.84
40	55	15470.80286	27406.20286	422.6888235	822.68
42	43	49095.222	50372.157	1382.964	1518.934
45	50	43388.01125	45536.22275	1222.1975	1382.7105
13	14	12967.52	13114.6	470.6	495.88
24	25	4263.53	4361.53	111.8	112
25	29	7700.32	7777.32	186.33	187
30	35	10220.4	10220.4	365.23	365.23
39	52	18171.378	18171.378	545.55	545.55
90	150	18171.378	18171.378	385.23	395.23
39	49	22896	22896	729	729
60	75	6228	6228	693	606
25	35	10207.18	10207.18	488.393	488.393
11	13	3335.898	3335.898	99.15	99.15
41	41	7708.57	7708.57	273	273
22	22	883.64	917.64	86.67	81
22	22	1762.42	1830.24	86.67	81
16	16	3550	3550	459	465
11	12	6366.812	6366.812	192.45	192.45
39	39	20311.12	20311.12	670.2	670.2
24	25	6228	6228	470.6	495.88
26	26	10207.18	10207.18	111.8	112

Fig.1: Collection of data of project work of planning, designing and Building stages

III. MATERIALS AND METHODS

Isotonic regression belongs to the family of regression algorithms an equation represents in equation (1). Formally isotonic regression is a problem where given a finite set of real numbers $Y=y_1,y_2,...,y_n$ representing observed responses and $X=x_1,x_2,...,x_n$ the unknown response values to be fitted finding a function that minimizes with respect to complete order subject with

$$f(x) = \sum_{i=1}^{n} w_i \left(y_i - x_i\right)^2 \quad (1)$$

respect to complete order subject to $x_1 \leq x_2 \leq ... \leq x_n$ and $y_1 \leq y_2 \leq ... \leq y_n$ where w_i are positive weights. The resulting function is called isotonic regression and it is unique. It can be viewed as least squares problem under order restriction. Essentially isotonic regression is a monotonic function best fitting the original data points.

The estimated and actual values are plotted as scatter diagram and regression analysis is made. It provides estimates of values of the dependent variables from values of the independent variable. The device used to accomplish this estimation procedure is the regression line. This method is more general in that the only restriction is that the mapping function be isotonic.

An experiment demonstrate that feature selection and isotonic regression methods improve the accuracy of predictions for match performance of based on each stage data of project development stage, compared to regression methods alone.

https://dx.doi.org/10.22161/ijaers/3.12.38 *ISSN: 2349-6495(P) | 2456-1908(O)*

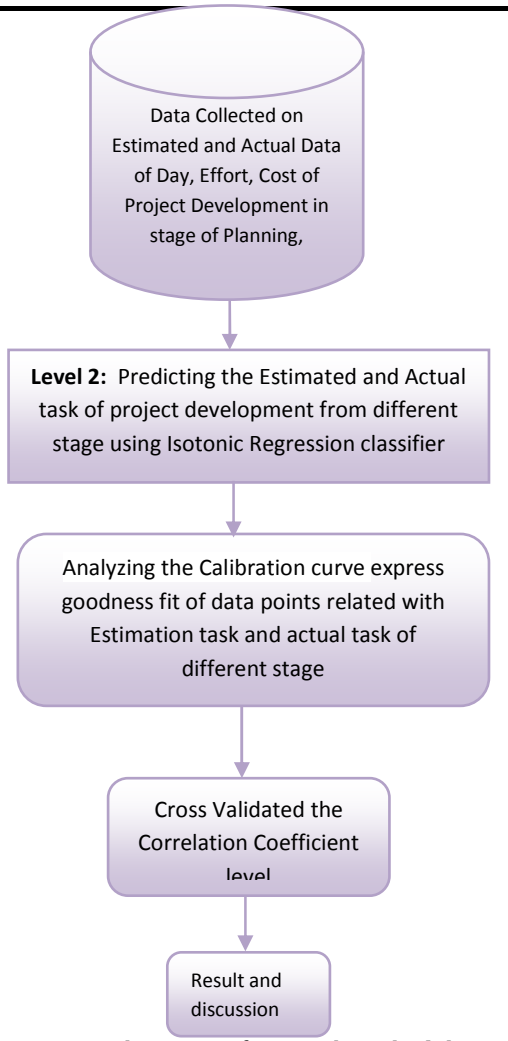

Fig. 1: Architecture of research methodology

Architecture of research methodology can be representing in fig. 1. For applying isotonic regression methods which improve the accuracy of predictions for match performance of based on each stage data of project development stage through calibration curve analysis.

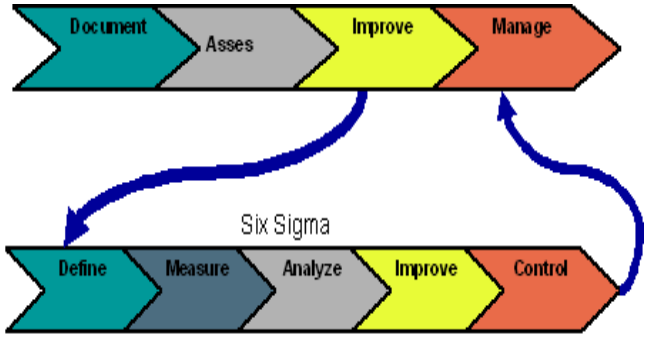

Fig. 2: Six sigma DMAIC Methodology

Define --> **M**easure --> **A**nalyze --> **I**mprove -->**C**ontrol

- **Define**: Define the problem or project goal that needs to be addressed.
- **Measure**: Measure the problem and process from which it was produced.
- **Analyze**: Analyze data and process to determine root causes of defects and opportunities.
- **Improve**: Improve the process by finding solutions to fix, diminish, and prevent future problems.
- **Control**: Implement, control, and sustain the improvements solutions to keep the process on the new course.

In this research work, methodology can be follow by DMAIC method of six sigma. In Define stage we can collect the project information which can analyse of the level of Business IT process. In measure stage, the deviation analysis can be carried out i.e., schedule variance, Effort Variance, Cost Variance respectively. In Analysis stage, an isotonic regression classifier of data mining technique can be used for predicting the level of project of Business IT in different stage. In improve stage, the error prediction can be analysing for rectification and follow to control stage for further improvement. The following fig. 2 represents the framework of proposed methodology, an objective of this metric is to reduce the schedule, effort and cost variation (deviation factor) by tracking it from beginning stage of the project through the end of the project task, thereby reducing time overruns. It can be follow by the methods of DMAIC (Define, Measure, **A**nalyze, Improve, and Control) in six sigma. It refers to a data-driven quality strategy for improving processes. This methodology is used to improve an existing business process.

In this research work, analyze stage can be focuses on the metrics to predicting the development of software project from different stage like planning, designing and building stage information related to estimated task and actual task respectively.

IV. RESULT AND DISCUSSION

4.1 Analysis of predicting effort and cost values of planning, designing and building stage

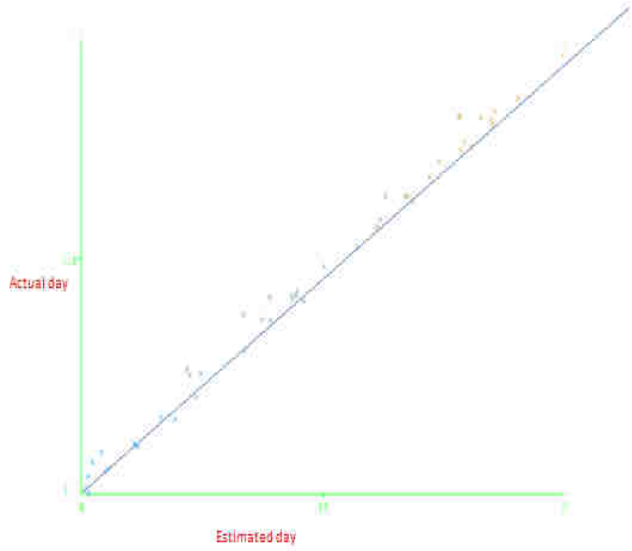

Fig.3: Calibration curve of Estimated and Actual day in planning stage

https://dx.doi.org/10.22161/ijaers/3.12.38 *ISSN: 2349-6495(P) | 2456-1908(O)*

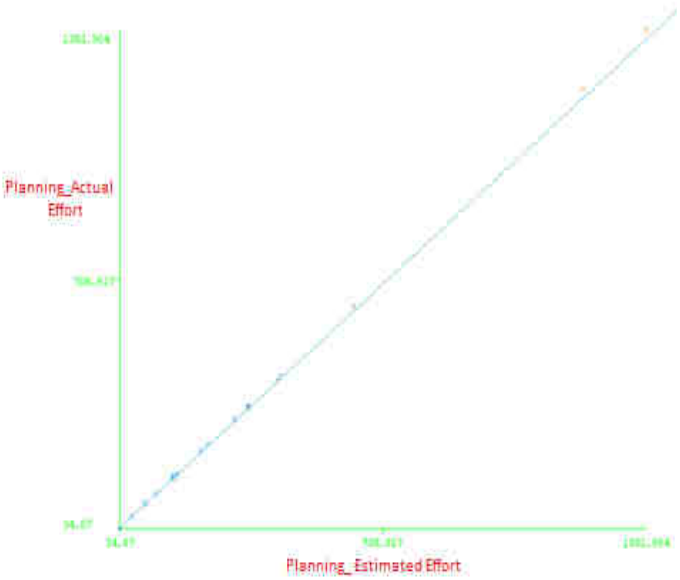

Fig.4: Calibration curve of Estimated and Actual effort in planning stage

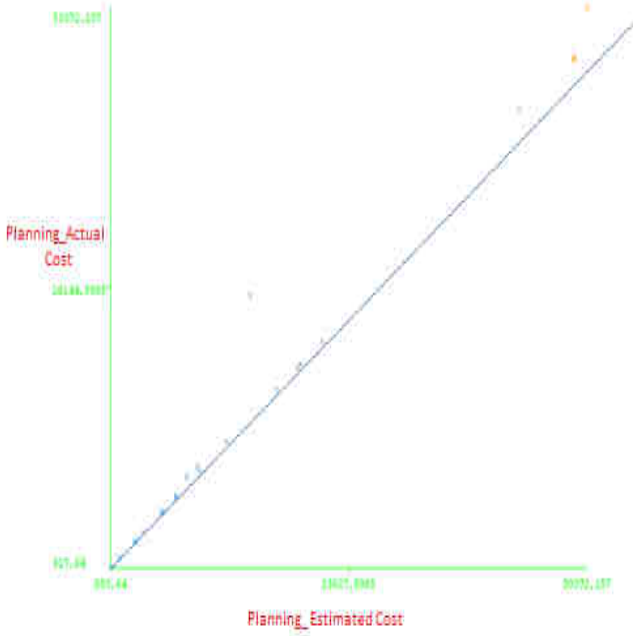

Fig.5: Calibration curve of Estimated and Actual Cost in planning stage

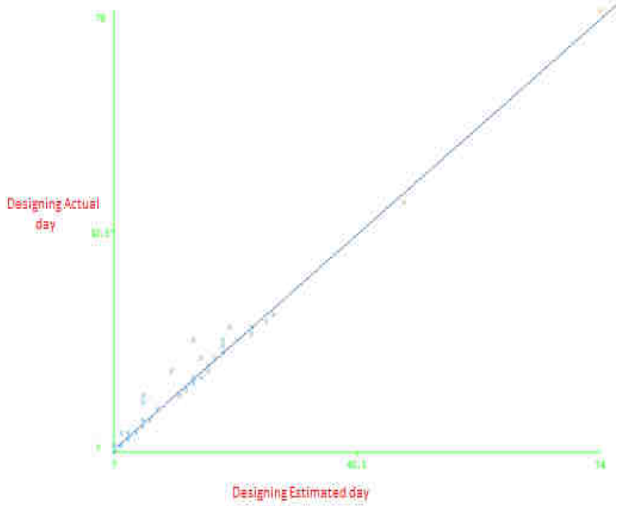

Fig.6: Calibration curve of Estimated and Actual day in Designing stage

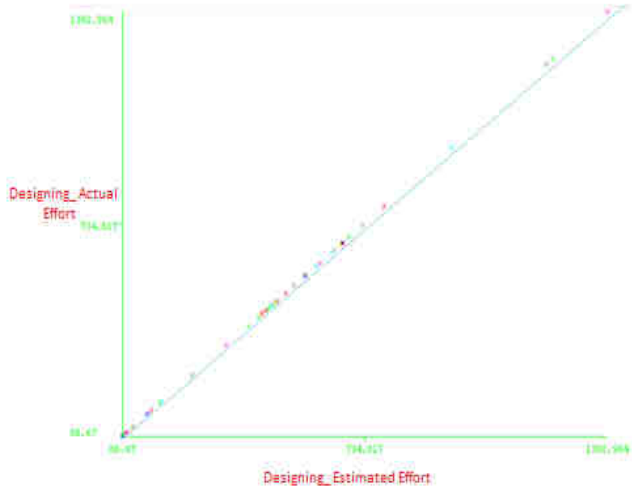

Fig.7: Calibration curve of Estimated and Actual Effort in Designing stage

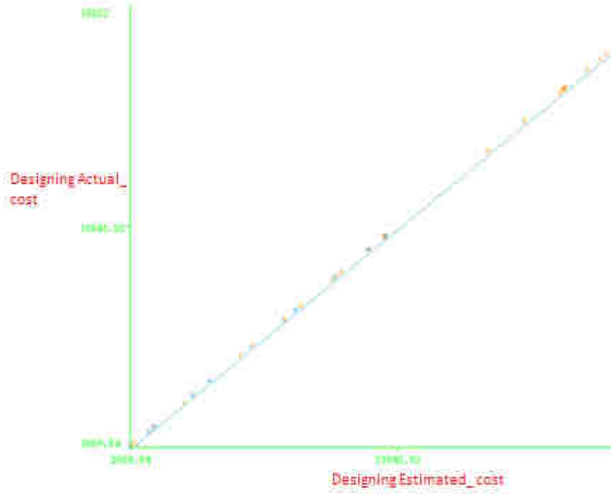

Fig.8: Calibration curve of Estimated and Actual Cost in Designing stage

https://dx.doi.org/10.22161/ijaers/3.12.38 *ISSN: 2349-6495(P) | 2456-1908(O)*

This work investigates the effectiveness of using computer-based machine learning isotonic regression methods to predict performance data for project development based on parameters collected on each stage of planning, building and designing task of estimated and actual values.

From the estimated and actual values are plotted as scatter diagram and regression analysis is made. It is found that isotonic regression fits well as shown in Fig. 1 to Fig. 8 for planning stage and designing stage. An Isotonic regression line shows a new calibration curve where the actual value is equal to the estimated value which express goodness of fit of the calibration curves can also indicate the estimation quality. The data points fitted positively which can data points above the graph.

Similarly, building stages can be good estimation, the actual value and predicted value agree very closely. In most cases a linear regression line captures this close relationship, in which case slope of the regression lines nearly equal to one.

The correlation coefficient indicates 0.9929 estimation capability of the classification in isotonic regression model, the coefficient values which express goodness of fit of the calibration curves can also indicate the estimation quality.

V. CONCLUSION

In this research work, it can be concluded that an estimated and actual value of planning, designing and building stage were plotted and shown the calibration curve of data point which has fitted in the plot. In this research planning stage of scheduling day, actual effort and actual cost of software development represents positively fitted which is a data point above the graph express goodness of fit of the calibration curves. Mostly the values are fitted from the plot analyzed as in positive where the data point is above the graph of each stage of the project work. Isotonic Regression is more powerful when there is sufficient data to prevent over fitting. Further it will examine the effectiveness of Isotonic Regression for calibrating the predictions made by different learning methods

REFERENCES

[1] Bibi, S., Tsoumakas, G., Stamelos, I. Vlahavas, I. 2008. “Regression via Classification applied on software defect estimation”, Elsiever, vol. 34, no. 3, pp. 2091-2101.

[2] D.T. Larose, “Discovery Knowledge In Data: an introduction to data mining”, Published by John Wiley & Sons, Inc., Hoboken, New Jersey, 2005.

[3] Chen Qingzhang, Fang Shuojin, Wang Wenfu, ”Development of the Decision Support System for Software Project Cost Estimation”, World Congress on Software Engineering, IEEE, 2009.

[4] Zadrozny, B., & Elkan, C. (2002). Transforming classifier scores into accurate multiclass probability estimates. *KDD* (pp. 694–699).

[5] R. L. Dykstra and T. Robertson. An algorithm for Isotonic Regression for two or more independent variables. The Annals of Statistics, 10(3):pp. 708–716, 1982.

[6] S. M Kakade, V. Kanade, O. Shamir, and A. Kalai. Efficient learning of generalized linear and single index models with Isotonic Regression. In NIPS. 2011.

[7] S. M. Mwagha, M. Muthoni, and P. Ochieg, “Comparison of nearest neighbor (ibk), regression by discretization and isotonic regression classification algorithms for precipitation classes prediction,” International Journal of Computer Applications, vol. 96, pp. 44-48, 2014.