



# ANALYTICAL SOLUTION TO MITIGATE THE RISK ARISING OUT OF TIME OVERRUN IN THE CONSTRUCTION PROJECTS VIA FORECASTING

Chandrabose M  
Senior Manager  
Centre of Excellence  
Advanced Analytics  
L&T Construction,  
Chennai, Tamil Nadu, India

Balaji B  
Lead  
Centre of Excellence  
Advanced Analytics  
L&T Construction,  
Chennai, Tamil Nadu, India

**Abstract**— Construction industry is getting affected by cost overruns, time overruns, poor budgeting & planning. Currently there is no analytical tool available which can act as a benchmarking system to measure the performance of a project based on the historical data of past projects. This is an effort to predict time overrun and cost overrun in projects for taking precautionary measures to minimize or reduce the impact.

**Keywords**—Machine Learning, Construction, Non-Linear Model, Auto-ARIMA, GAM Model, Cost Overrun, Time Overrun

## I. INTRODUCTION

Currently, time overrun and cost overrun are the two major concerns in construction industry. This is mainly due to lack of proper forecasting mechanism, because of which many projects end up in lower than projected gross margin or negative margin. Till the last phase of the project, management is unaware of this and end up spending more money on loss making projects rather than controlling the same in the initial stages itself. Concreting -Shuttering - Reinforcement Steel attributes to almost 36% of the total cost incurred in Construction and the percentage varies according to the nature of the work like 50% of the total cost in building and factories and 20% in water treatment unit. For the first phase we have considered CSR cost packages for arriving at an implementable solution to reduce the time overrun. Cost completion % will be arrived based on the actual completion

of the CSR activities based on the historical data for monitoring the progress of current construction projects. In recent times, a lot of study work happening in the field of monitoring and controlling the progress of construction projects, but majority of the studies arrived at conclusion based on primary data collected through structured questionnaires with site engineers and managers.

This project aims to develop a solution which proactively flags the cost package incurring delays resulting in timely corrective actions and to reduce time overrun and improve cost savings

## II. PROBLEM STATEMENT

Every Construction project is prone to risk, but it can be minimized with proper risk strategy.

If the Risk is ignored or not foreseen, business will face the below consequences

1. Cost Overrun
2. Time Overrun

Cost Overrun and Time overrun will lead to loss or decrease in profit, damage to reputation, and no repeated orders from customers.

**Time overrun** occurs when the project exceeds the committed timeline to the client. With digitization and IOT of things business could be able to control this to some extent but still time overrun remains as the major pain point to address



**Cost overrun** occurs when the project exceeds the budgeted cost and end up spending more which leads to erosion in the planned margin. According to global standards it is found out that every 9 out of 10 projects end up in cost overruns.

Time overrun and Cost overrun in construction industry occurs due to the following factors

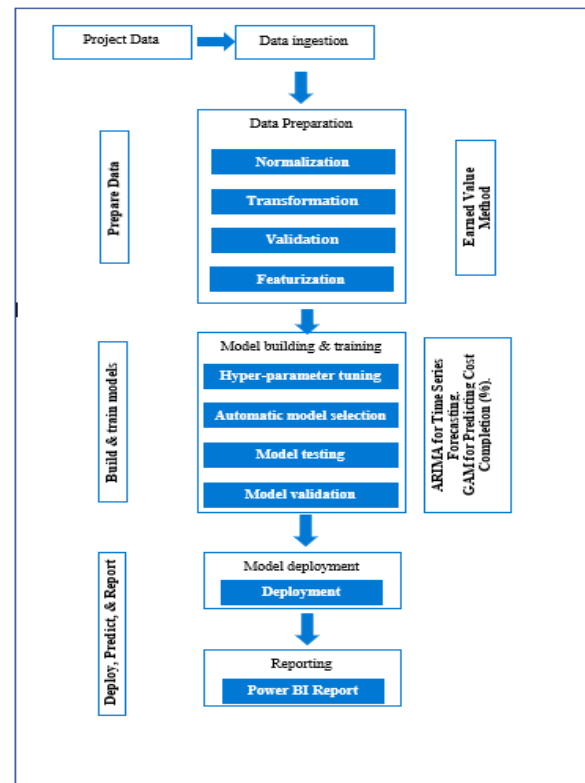
1. Change in design
2. Increase in raw material cost
3. Increase in labor cost
4. Environmental issues
5. Increase in interest rate
6. Delay in progress payment by clients
7. Political issues
8. Exchange rate fluctuations
9. Labor productivity
10. Non availability of skilled labors
11. Labor Shortage
12. Ineffective time management of sub-contractors
13. Poor cost estimation
14. Poor tendering documents
15. Poor material management
16. Unplanned Costs

To mitigate risk, business always keeps contingency cost as part of the budget, irrespective of that many projects ends up in either time overrun or cost overrun due to the above factors. There are two kinds of risks, controllable and uncontrollable. Uncontrollable risks can be mitigated or minimized. Business focus on controllable risks to save either time or cost. This can be achieved by reviewing each activity in the construction process and by having lead indicator which can highlight the high-risk projects at the lowest possible level for taking preventive action.

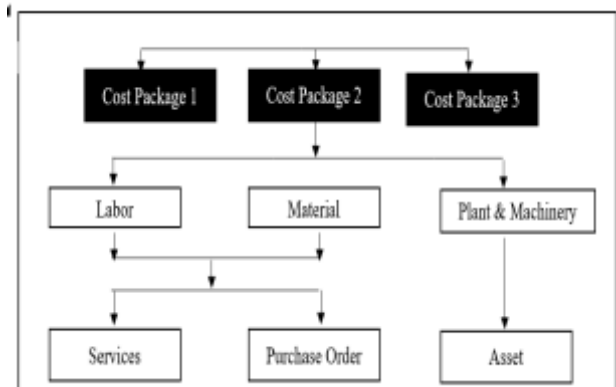
As per statistics from the Indian ministry of statistics and programme implementation, while analyzing projects worth Rs 150 crore or more, nearly 438 projects reported cost overrun and within the 1,670 projects which are monitored 563 projects were delayed. Original cost of implementation of 1670 projects was Rs. 21.66 Lakh Crore but the expected cost of completion is Rs. 25.96 Lakh Crore, which is Rs. 4.30 Lakh crore cost overrun. Out of the 563 delayed projects, 1-12 months delay observed in 100 projects, 13-24 months delay observed in 120 projects, 25-60 months delay observed in 260 projects and > 61 months delay observed in 127 projects. Average time overrun of 47 months observed in these 563 delayed projects

While the numbers reported above looks significant, but it is still under reported as the project agencies are not reporting revised cost estimates and project schedules for many projects. To mitigate the risk arriving out of time overrun and cost overrun, we have developed an analytical solution. Detailed approach is given below

### III. SOLUTION ARCHITECTURE



### IV. BUSINESS PROCESS FLOW



For this project we have considered cost package details from JCR and ACE. Cost completion of a project details from PMS, Job cost report (JCR) document is the status report submitted by a particular project site every 60 days. It contains the details of the executed work and the estimate for the revised work. It also contains the details of the deviation from the accepted cost estimate (ACE) document. Performance management system (PMS) is the data maintained and updated by finance team every month to monitor the financial performance of a particular site.



#### 4.1 Data Preparation

- JCR (Job Cost Report) Data is the underlying data for the ARIMA-based Forecasting Model.
- PMS (Performance Management System) is used to get the Cost Completion (%) of Projects.
- Cost Completion (%) should be less than 95% to be considered as an active operating job.
- Cost package code varies across years for concreting, shuttering and reinforcement and it needs to be mapped for analysis
- Data Cleaning to treat for missing and incorrect values.

#### Few Examples

- Quantity and value is available but rate is missing - Rate arrived based on Value divided by Quantity
- Negative values entered in the estimate to match the budgeted cost
- Cost Package with two different UOMs, converted to a common UOMs

#### 4.2 MODEL EXPLANATION – EVM

EVM (Earned Value Method) is used to derive features from JCR Data that can be used to evaluate Project Performance and Forecasting.

BCWP – Budgeted/Ideal Cost of Work Performed = Revised Rate x Work Done. ACWP - Actual Cost of Work Performed = Actual Rate x Work Done

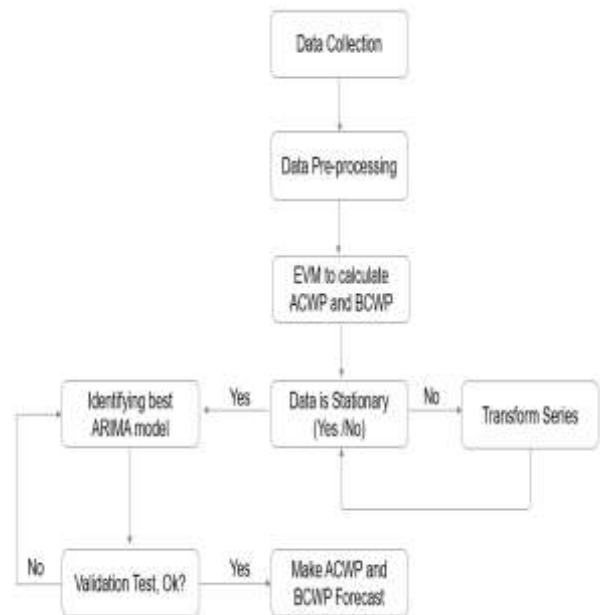
CPI - Cost Performance Index = BCWP/ACWP. CPI less than '1' means project is over budget.

Using Time Series Forecasting Models like ARIMA, BCWP and ACWP metrics can be forecasted to identify projects that are going to incur a cost overrun in near future.

#### 4.2.1 ARIMA MODEL -Explanation

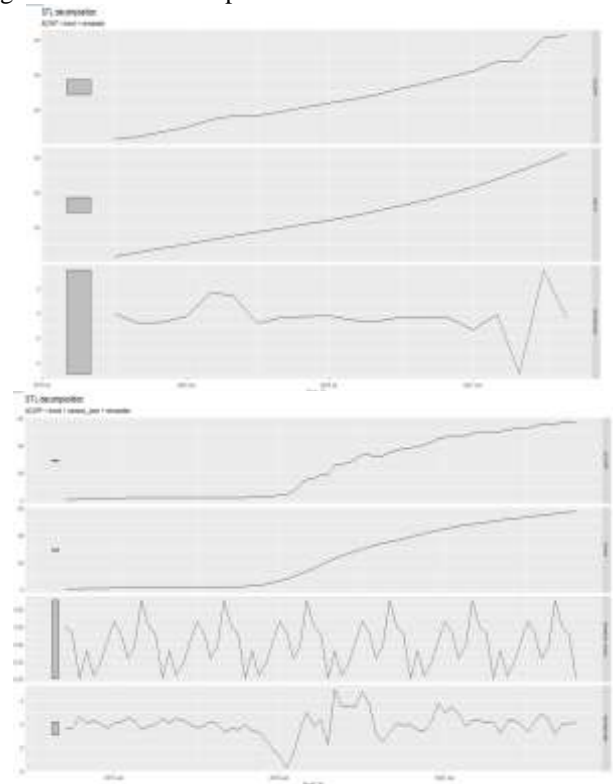
Using EVM Transformed JCR Data, ARIMA forecasts BCWP and ACWP and calculated Potential Cost Overrun is calculated by taking difference between them. It is calculated both at Project-level as well as at Cost Package-level. This will help in not only identify projects that are going to incur cost overrun in future. But, also to identify the cost packages that are contributing to cost overrun. Ultimately, helping in monitoring and controlling of costs

#### 4.2.2 ARIMA Model Flow



#### 4.2.3. Why Auto ARIMA is chosen

Figure 3A: STL decomposition





While observing the Seasonal and Trend decomposition for two different jobs, one job is having seasonality while the other doesn't have any seasonality. Auto ARIMA model takes care of the above by selecting the combination of values, where p is the number of autoregressive terms, d is the number of nonseasonal differences required for stationarity, and q is the number of lagged forecast errors in the equation

**4.3. GAM Model Explanation**

GAM is used to predict Cost Completion (%) of a project based on the Work Progress (%) of Concreting, Shuttering, and Reinforcement Activities.

GAM is trained on historical Data of projects in a Business Unit successfully completed on time.

Accuracy of GAM is dictated by the Weightage of CSR Activities in terms of Cost in that given Business Unit successfully completed on time.

**4.3.1 GAM equation**

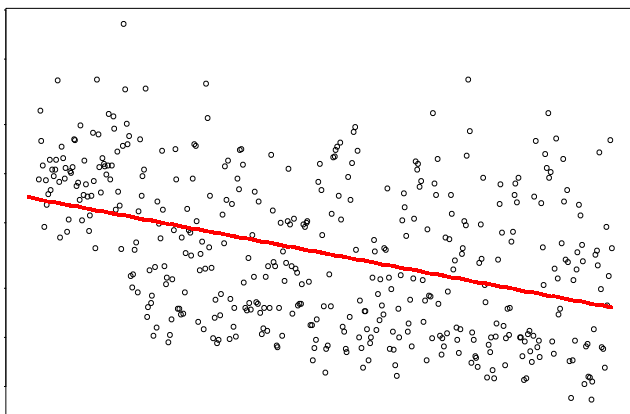
$$Y = a + f_1(x_1) + f_2(x_2) + \dots + f_n(x_n) + \epsilon$$

-a is an intercept and f are smooth functions. As smooths different types of functions can be used such as local linear regression (loess) or splines.

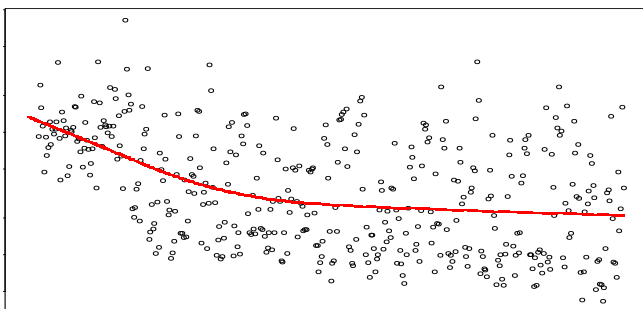
Splines are sums of weighted basis functions and have better mathematical properties and are most often used in GAM fitting.

**4.3.2 Why GAM Model**

**Linear Regression Fit**



**Additive Model Fit**

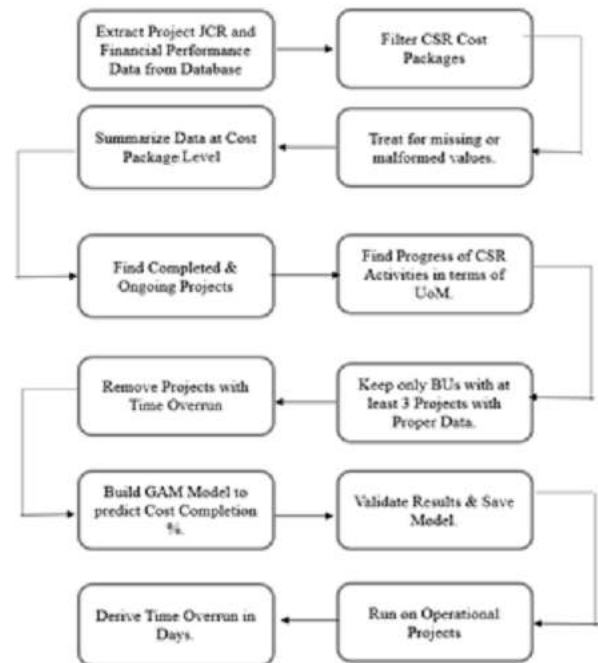


Using GAMs, assumption of linearity can be relaxed between predictors and response variable. Unlike Complex ML Models like Neural Networks, GAM allows us to do inferential statistics to understand and explain the underlying structure of models.

Unlike Linear Regression, GAM allows to make predictions of complex relationships like Neural Networks and Boosted Regression Trees.

It allows us to model these non-linear relationships through smooths or splines that can take a wide variety of shapes

**4.3.3 GAM MODEL FLOW**



**V. FINDINGS/ RESULTS**

More the weight age of CSR to the overall cost, more the accuracy of the model in predicting the cost completion %. For the projects considered for training, GAM model predicted the time overrun correctly with an accuracy level of 86%. Across BU's R-sq.(adj) varied from 0.882 to 0.494 based on the CSR weightage to the overall cost. Deviance explained varied from 89.4% to 51%.

**VI. CONCLUSION**

Model accuracy improved based on the weight age of CSR activities to the overall cost. This project is particularly suitable for projects that have CSR Activities as major cost activities. Through identification of key activities, Time Overrun and cost overrun for a project can be determined with higher accuracy for projects that don't have CSR as major cost activities.



This approach will help in data-driven decision making by identifying projects and related cost packages that are going to contribute to erosion of gross margin related to CSR Activities

#### VII. LIST OF SYMBOLS & ABBREVIATIONS USED

RSV	Realized Sales Value
IC	Independent Company
BU	Business Unit
ACE	Accepted Cost Estimate
JCR	Job Cost report
PMS	Performance Management System
EIP	Enterprise Information Portal
EPM	Enterprise Planning Management
ACWP	Actual Cost of Work Performed
BCWP	Budgeted/Ideal Cost of Work Performed
CPI	Cost Performance Index
EVM	Earned Value Method
LMP	Labor, Material and Cost
CSR	Concreting, Shuttering (Formwork) and Reinforcement
ARIMA	Autoregressive Integrated Moving Average
GAM	Generalized Additive Model
UOM	Unit of Measurement

#### VIII. KEY TERMS AND FORMULAS

**Timeline Progress (%)** is defined as the time elapsed till Performance Date since the beginning of the project.

Formula:  $\text{Timeline Progress (\%)} = (\text{Performance Date} - \text{Expected Start Date}) \times 100 / (\text{Expected End Date} - \text{Expected Start Date})$

**Cost Schedule Progress (%)** is defined as the cost that has already been spent out of the Revised Cost Budget till Latest JCR Date.

Formula:  $\text{Cost Schedule Progress (\%)} = (\text{Total Actual Cost} / \text{Total Revised Cost}) \times 100$

**Ideal Cost** is the Budgeted Cost of Work Performed as mentioned in the explanation of EVM.

Formula:  $\text{Revised Rate} \times \text{Work Done}$

**Actual Cost** is the Actual Cost of Work Performed as mentioned in the explanation of EVM. Formula:  $\text{Actual Rate} \times \text{Work Done}$

**Forecasted Ideal Cost** is given by the ARIMA Model based on the past trend of Ideal Cost of a given project.

**Forecasted Actual Cost** is given by the ARIMA Model based on the past trend of Actual Cost of a given projects

**Cost Overrun** is the difference between Ideal Cost and Actual Cost of Work Done.

**Potential Cost Overrun** is the difference between Forecasted Ideal Cost and Forecasted Actual Cost of Work Done.

**Cost Overrun Variance** is the difference between Cost Overrun and Potential Cost Overrun.

Potential Cost Overrun is calculated both at Project-level as well as at Cost Package-level.

#### IX. ACKNOWLEDGMENT

The authors wish to thank Mr. R Ganesan (Head - Corporate Centre, L&T Construction) for his continuous support and encouragement.

#### X. REFERENCE

- [1]. Ivan Damnjanovic, Kenneth Reinschmidt. (2020). Data Analytics for Engineering and Construction Project Risk Management, (pp 305-355)
- [2]. KAKA A.P. (1999). The development of a benchmark model that uses historical data for monitoring the progress of current construction projects,(pp 256-266)
- [3]. Shuzhuan Zheng, Rong Liu, Lijian Yang,
- [4]. Wolfgang K. Hardle. (2016). Statistical inference for generalized additive models: simultaneous confidence corridors and variable selection, (pp 256-266)
- [5]. Stone CJ. (1986). The dimensionality reduction principle for generalized additive models. *Ann Statist* 14. (pp 590–606)
- [6]. Shrestha, P.P., Burns, L.A., & Shields, D.R. (2013), Magnitude of construction cost and schedule Overruns in Public Work Projects. *Journal of Construction Engineering*, (pp 1-9 )
- [7]. Halloum, M.A., & Bajracharya, A.A. (2012). Cost and time overrun revisited: A study on the infrastructure construction projects in Abu Dhabi, UAE, The Third Conference on Construction in Developing Countries, Bangkok, Thailand (pp 93-100)
- [8]. A. A. Oladapo. (2007). A Quantitative Assessment of the Cost and Time Impact of Variation Orders on Construction Projects. *Journal of Engineering, Design and Technology*, vol. 5(1), (pp. 35-48).
- [9]. Ahmed Senoucia, Alaa Ismail, Neil Eldin. (2016). Time Delay and Cost Overrun in Qatari Public Construction Projects. (pp 368-375).
- [10]. Anandbabu Chandragiria, Syed Hamim Jeelani, Salim Akthar, Lingeswaran N. (2021). A study and identification of the time and cost overrun in the construction project. (pp 5426-5431).
- [11]. Vaibhav Y. Katre, Dr. D.M. Ghaitidak. (2016). Elements of Cost and Schedule overrun in Construction Projects *International Journal of Engineering Research and Development*. (PP.64-68).
- [12]. Harshita Ambre, Dhananjay Demse. (2019). Analysis of Cost Overrun in Construction Projects. (pp 5209-5212).
- [13]. Dalia Mohamed, Florida Srour, Wael Tabra, Tarek Zayed. (2009). A prediction model for construction project time contingency . (pp 736-745).
- [14]. Barrientos, Alexis & Ballesteros-Pérez, P. & Mora-Meliá, Daniel & Cerezo-Narváez, Alberto. (2022).



- Stability Study of Cost Prediction Methods with Earned Value Analysis. (pp 1-10)
- [15]. Anbari, F. (2003). Earned value project management method and extensions. IEEE
- [16]. Engineering Management Review, 32(3), (pp 97–97)
- [17]. Anysz, H., & Buczkowski, B. (2019). The association analysis for risk evaluation of significant delay occurrence in the completion date of construction project. *International Journal of Environmental Science and Technology*, 16(9), (pp 5369–5374).
- [18]. Batselier, J., & Vanhoucke, M. (2017). Improving project forecast accuracy by integrating earned value management with exponential smoothing and reference class forecasting. *International Journal of Project Management*, 35(1), (pp 28–43).