# Cost Overrun Analysis Using Artificial Neural Network in Residential Construction Projects

Ankita Yadav and Jaydeep Pipaliya



January 29, 2024

# Cost Overrun Analysis Using Artificial Neural Network in Residential Construction Projects

Ankita Omprakash Yadav
*M. Tech Scholar, Department of Civil Engineering*
*Parul Institute of Engineering and Technology, Parul University*
Vadodara, Gujarat, India
2203052160010@paruluniversity.ac.in

Jaydeep Pipaliya
*Assistant Professor, Department of Civil Engineering*
*Parul Institute of Engineering and Technology, Parul University*
Vadodara, Gujarat, India
jaydeep.pipaliya21306@paruluniversity.ac.in

***Abstract*— This study delves into the analysis of cost overruns in residential construction projects using Artificial Neural Networks (ANN). By incorporating factors identified through a stakeholder and expert-rated survey, the research aims to develop a predictive ANN model. The model utilizes project-specific combinations of these factors to predict cost overruns. Conducted in the context of Mumbai, this research contributes to the construction industry by offering an advanced tool for proactive cost management. The ANN model's adaptability and learning capabilities from historical data hold promise for improving accuracy in forecasting, enabling project managers to implement effective mitigation strategies. This study underscores the potential impact on project outcomes and the overall enhancement of cost management practices within the residential construction sector.**

***Keywords—Cost Overrun, Residential Construction Projects, Artificial Neural Networks, Predictive Modelling, Project Management***

## I. Introduction

The construction industry is characterized by inherent complexities, and one of the persistent challenges faced by project managers is the occurrence of cost overruns. In the realm of residential construction projects, understanding and mitigating the factors contributing to cost overruns are paramount for successful project outcomes. This study focuses on an innovative approach to address this issue by employing Artificial Neural Networks (ANN) for cost overrun analysis. With a growing emphasis on predictive modelling and advanced analytics in construction management, leveraging ANN provides an opportunity to discern intricate patterns and relationships within the data.

In order to pinpoint the crucial elements causing cost overruns in residential building projects, an extensive survey was conducted involving stakeholders, experts, and individuals closely associated with the construction industry. The survey yielded a set of key factors, including Project Size, Project Complexity, Scope Changes, Project Location, Project Management Quality, Design Changes, Labor Costs, Site Conditions, Material Cost Fluctuation, Weather Conditions. These factors form the basis for our ANN model, which seeks to learn from historical project data and predict cost overruns based on specific project characteristics.

Located in the dynamic urban landscape of Mumbai, the study's context is significant due to the city's unique construction challenges and diverse residential projects. The research endeavors to contribute a predictive tool that enhances cost management practices, enabling project managers to proactively identify and address potential issues. By harnessing the adaptive learning capabilities of ANN, this study aims to improve accuracy in forecasting cost overruns, fostering more effective decision-making in residential construction project management. The subsequent sections will delve into the methodology, data analysis, and findings, offering important information for industry professionals, researchers, and stakeholders engaged in the dynamic field of construction management.

## II. Literature Review

Swapnil P Wanjari; Gaurav Dobariya (2016) [1] identified the primary reasons behind cost overruns in Indian building projects as the rise in raw material prices, delays in the scheduled activities, and a absence of coordination among building teams. Using factor analysis to group these factors into three components within the comprehensive questionnaire: contractor control, client control, and project management components. These components offer valuable insights for various stakeholders involved in construction activities. Furthermore, the paper proposes frameworks based on discussions with a substantial number of construction professionals and experts.

Bichoy Samir Tamez Aziz, Mohamed Badawy et.al (2014) [2] has successfully recognized and, utilizing quantified relative importance indices, determined the ranking of influence for 28 elements that affect how much a home construction project costs in Egypt. The initial step in addressing this challenge involves the identification of factors influencing costs. If construction firms can adeptly control these pivotal factors and devise preventive strategies, it becomes possible not only to prevent cost overruns but also to enhance overall project profits. To provide construction companies with accurate insights into the key cost-influencing factors, a thorough analysis and specification of these factors are conducted from the perspective of construction firms.

Shemi S,N; Asok Kumar N (2020) [3] examined the importance of factors influencing the accuracy of cost estimates, identifying 40 reasons for cost overruns. Following this, statistical models were created using both neural network and multiple regression methods. Modeling regression was utilized to establish correlations among particular components and project costs, and the outcomes of the models developed using artificial neural networks (ANN) were statistically validated. The overall variation is characterized by the inclusion of all factors, amounting to 83%, making the task feasible from an informational standpoint, with only a 17% decrease in data. In the neural network models, the optimal engineering performance was achieved with a configuration of 10-13-1, comprising 10 inputs, thirteen hidden levels, and one output representing cost overrun, highlighting the most effective statistical model. When comparing the results obtained from regression and neural network models, the latter exhibits a more probabilistic model due to the inherent nature of neural networks.

V.B. Chandanshive; A.R. Kambekar (2019) [4] utilized a multilayer feedforward neural network model that undergoes training with a backpropagation algorithm. The purpose is to predict the costs associated with building construction. The trained neural network model's outcomes demonstrate its effectiveness in early-stage prediction of building construction project costs. This study contributes to construction management by providing insights into the complete financial budget, helping financiers and property owners manage their investments in the erratic construction sector and make decisions.

## III. Methodology

The comprehensive methodology utilized in this study is elucidated as follows. It encompasses the identification of elements impacting overruns in project costs for residential development and the implementation of an inquiry-based questionnaire. The collected data is then employed to construct neural network and regression algorithms to forecast the occurrence of cost overruns.

### A. *Determining the elements influencing cost overrun*

The first phase involved a review of past research papers to examine different factors that contribute to cost overruns. 35 factors were identified through extensive literature survey and the inputs from various individuals involved in the residential construction projects..

### B. *Preparation of Questionnaire Survey*

A questionnaire survey using a 5-point Likert scale was used to assess the crucial factors contributing to cost overruns. Participants were asked to rank the significance of various factors using a Likert scale, providing a quantitative measure of their impact on cost overruns. The survey aimed to gather valuable insights into the perceived significance of each factor, facilitating a comprehensive understanding of their influence on project costs. The scale ranges from 1 to 5 ranging from Not important to Very important.

### C. *Identification of Critical Factors*

The Relative Importance Index (RII) approach was used to examine the data gathered from the questionnaire survey. This analysis identifies the essential elements driving cost overruns. The Relative Importance Index (RII) is a statistical method used to assess and prioritize factors based on their perceived importance or impact. It is often employed in surveys or studies wherein participants are asked to rank the importance of certain elements on a numerical scale. The RII is calculated by assigning scores to the responses, and it provides a quantitative measure of the relative importance of each factor within the given context. This method allows for the systematic analysis and prioritization of factors, aiding in decision-making processes and resource allocation based on their perceived importance.

$$RII = \frac{Weighted\ Sum\ of\ Responses\ for\ a\ Factor}{Maximum\ Possible\ Weighted\ Sum\ of\ Responses}$$

### D. *Regression Analysis*

In the context of analyzing cost overrun in residential construction projects, regression analysis through SPSS software proves invaluable. The process entails entering project data into the SPSS spreadsheet, designating the dependent variable as cost overrun, and specifying relevant independent variables. The regression analysis is then executed through the SPSS interface, yielding crucial outputs such as coefficients, standard errors, and significance levels. The examination of these results, particularly the significance of coefficients and goodness-of-fit metrics like R-squared, aids in understanding the factors influencing cost overrun. It is imperative to assess assumptions and conduct additional analyses to ensure the robustness of the model. SPSS facilitates a user-friendly platform for this analysis, offering a comprehensive understanding of the influential variables and contributing factors to cost overruns in residential construction projects.

### E. *Artificial Neural Network*

When employing Artificial Neural Network (ANN) analysis for the investigation of cost overrun in residential construction projects, MATLAB software serves as a robust platform. The process involves the segmentation of the dataset into training and testing sets, facilitating the development and validation of the ANN model. Leveraging the functionalities of MATLAB, the neural network architecture, including the number of layers and nodes, is configured to optimize the process of learning the model. Through iterative adjustments, the training phase entails network's weights to minimize prediction errors. Following training, the model is tested using an impartial dataset to evaluate its ability to generalize. MATLAB's capabilities in handling complex computations and simulations make it an advantageous tool for implementing ANN models, enabling a comprehensive analysis of the intricate relationships and patterns influencing cost overrun in residential construction projects.

## IV. Data Analysis

General Information

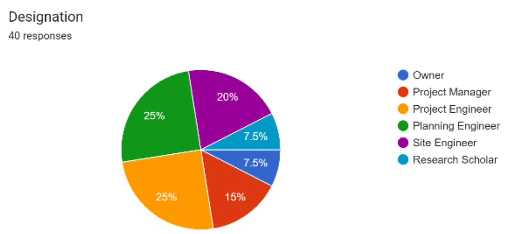


Fig. 1. Working Sector of Respondent.

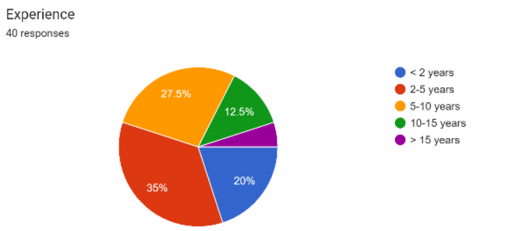


Fig. 2. Years of Experience of the Respondents

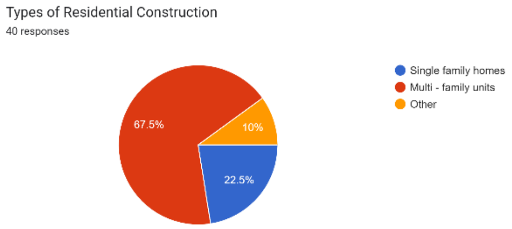


Fig. 3. Types of Residential Construction Projects

## A. *Identification of Critical Factors*

*Analysis through RII Method*

Using RII Method, top 10 critical factors are identified which majorly impact cost overrun in Residential Construction Projects.

TABLE I. RII RANK OF TOP TEN CRITICAL FACTORS

| Sr No | Factors | RII | Rank |
|---|---|---|---|
| 1 | Project Size | 0.89 | 1 |
| 2 | Project Complexity | 0.83 | 2 |
| 3 | Scope Changes | 0.83 | 2 |
| 4 | Project Location | 0.80 | 3 |
| 5 | Project Management Quality | 0.79 | 4 |
| 6 | Design Changes | 0.76 | 5 |
| 7 | Labor Costs | 0.72 | 6 |
| 8 | Site Conditions | 0.71 | 7 |
| 9 | Material Cost Fluctuation | 0.69 | 8 |
| 10 | Weather Conditions | 0.68 | 9 |

## B. *Regression Models*

Multiple regression analysis is an effective technique used to forecast the unobserved value of a variable by considering the known values of two or more variables.

TABLE II. MODEL SUMMARY

| Model | R | $R^2$ | Adjusted $R^2$ | Std. Error of Estimate |
|---|---|---|---|---|
| 1 | 0.957 | 0.916 | 0.882 | 1.273 |

The multiple regression model analyzed the relationship between various predictors (Weather Conditions, Design Changes, Project Location, Scope Changes, Project Size, Material Cost Fluctuation, Labor Costs) and the dependent variable, Cost Overrun (%). The model demonstrated a high level of overall fit, with an R Square of 0.916, indicating that roughly 91.6% of the variance in Cost Overrun (%) was explained by the predictors. The Adjusted R Square, considering the number of predictors, was 0.882. The model's standard error of the estimate was 1.273.

TABLE III. ANOVA STATISTICS

| ANOVA | | | |
|---|---|---|---|
| ***Model*** | ***Sum of Squares*** | ***Mean square*** | ***Sig.*** |
| Regression | 301.056 | 43.008 | <0.001 |
| Residual | 27.566 | 1.662 | |
| Total | 328.622 | | |

The ANOVA table for the multiple regression model reveals compelling evidence of its statistical significance in predicting Cost Overrun (%). The Regression section indicates that the predictors collectively contribute significantly, with a high Sum of Squares (301.056) and an F statistic of 26.523, yielding a p-value less than 0.001. This implies that approximately 91.6% of the variability in Cost Overrun (%) is explained by the predictors. The Residual section, representing unexplained variability, has a relatively low Sum of Squares (27.566), suggesting a good fit of the model. The Total Sum of Squares (328.622) underscores the overall variability in the dependent variable. These findings align with the earlier analysis, reinforcing that the multiple regression model, encompassing Weather Conditions, Design Changes, Project Location, Scope Changes, Project Size, Material Cost Fluctuation, and Labor Costs, is robust and statistically significant in its ability to predict Cost Overrun (%).

## C. *Artificial Neural Network Model*

The model is created using MATLAB

*Training the network*

Training a neural network involves a systematic process of preparing and teaching a model to make accurate predictions or classifications. This begins with defining the problem and collecting a labelled dataset, followed by data preprocessing and splitting into training, validation, and test sets. Choosing an appropriate neural network architecture, initializing the model with random weights, and defining a suitable loss function are critical steps. The training loop involves iteratively feeding batches of data, computing loss, and updating the model's weights using an optimization algorithm. Continuous validation ensures that the model generalizes well to new data, and hyperparameters may be adjusted accordingly. Testing on a separate dataset evaluates the final model performance. Fine-tuning and deployment follow, allowing the neural network to be utilized for making predictions on real-world data. The success of training relies on careful consideration of factors such as dataset quality, architecture selection, and hyperparameter tuning.

Seventy percent of the data was utilized in the training process for the network architecture. Levenberg-Marquardt algorithm was used and the performance was evaluated using Mean square error.

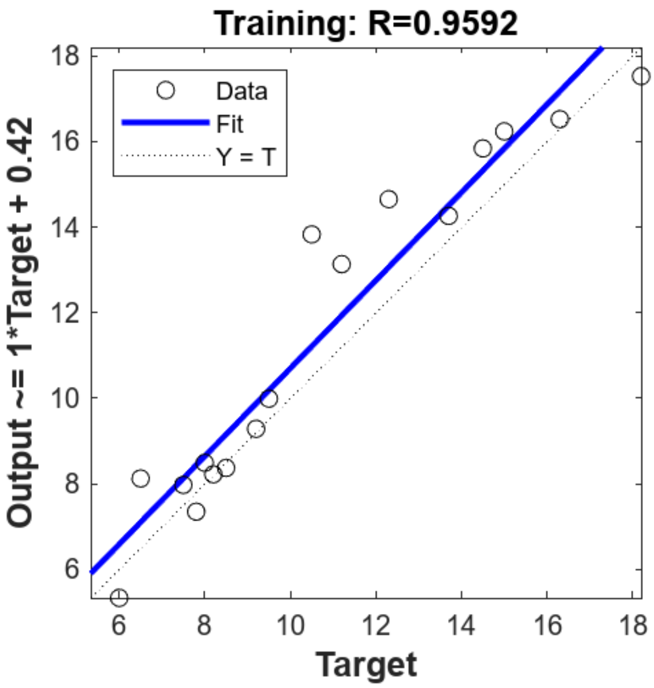


Fig. 4. Training Performance

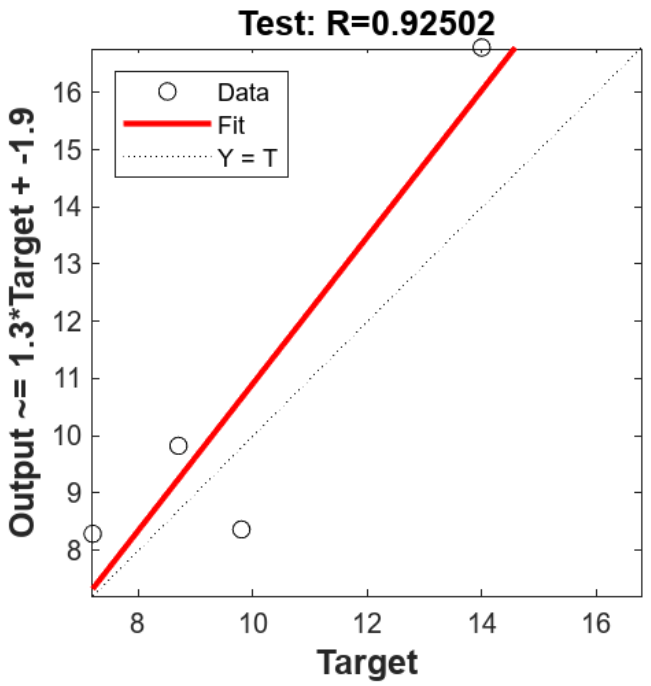


Fig. 5. Testing Performance

The R value, functioning as a correlation coefficient, serves to highlight the degree and direction of the relationship between solutions and objectives within a dataset. A value of 1 signifies a flawless positive linear relationship, where both variables rise in tandem. As R converges toward zero, it signals a weaker or non-existent linear connection. Regarding training data, an R value nearing 1 indicates a strong linear correlation between objectives and solutions. Specifically, in the provided instance, the output for cost overrun closely aligns with 0.95 times the target value, where 0.95 denotes the slope of the linear fit. This implies that as solutions undergo changes, the cost overrun is anticipated to increase by 0.95 times the target value, reinforcing the accuracy of the linear model's predictions in accordance with the answers provided during the training session.

*Testing the network*

Testing the network involves assessing the way in which the trained neural network model performs on a separate dataset that it has not encountered during the training process. This dataset, known as the test set, serves as an independent benchmark to analyze the model's generalization to fresh, untested data.

Fifteen percent of the data was utilized for testing the network and the network. Like training the input and output variable remain the same with ten layer of input sources and one layer of output source.

The test performance evaluated the result around 1.3 of the target value.

*Validation of the data*

Validation of data is a crucial step in the machine learning process, particularly during the training phase of a model. It involves assessing the performance of the model on a separate dataset from the one used for training, called the validation set. The purpose of validation is in order to adjust the hyperparameters of the model and avoid overfitting.

While training, the model continuously learns from the training set, and periodic evaluations on the validation set help determine how well the model generalizes to new, obscure data. By monitoring the model's functionality on the validation set, adjustments can be made to hyperparameters, such as learning rate or regularization strength, to enhance the model's capacity for generalization.

Fifteen percent of the data is utilized for the validation of the model.

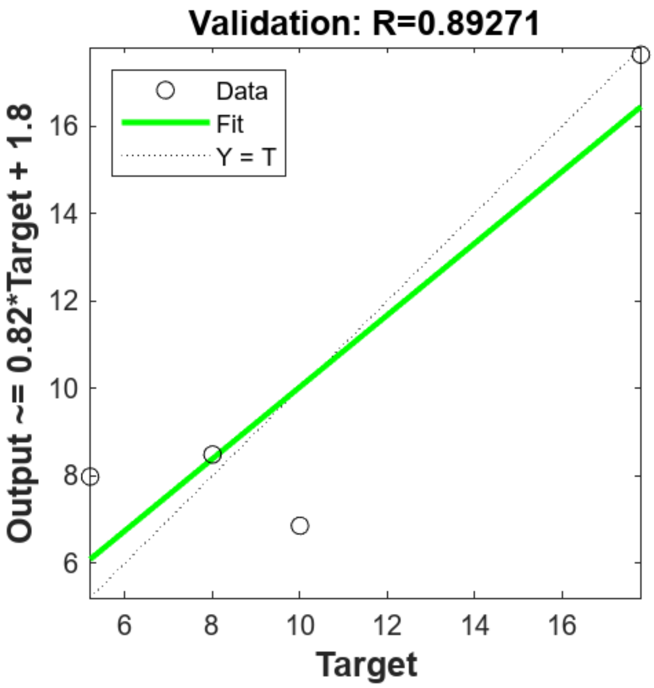

Fig. 6. Validation Performance

The obtained value is approximately equal to 0.82 times the target value, referring to the cost overrun derived from the model.

Cost Overrun Y= 0.82* target value + 1.8

## V. Result And Conclusion

In an extensive investigation into cost overrun within residential construction projects in Mumbai, a thorough field review was undertaken to evaluate the importance of factors influencing the accuracy of cost estimates. Through meticulous research, 35 distinct reasons contributing to cost overrun were identified. Statistical models were subsequently created using both neural network and multiple regression methodologies. Regression modelling played a crucial role in discerning the correlation between selected factors and project costs, while the results from the artificial neural network (ANN) model underwent thorough statistical validation. The collective spectrum of factors covered 88% of all variations, indicating the project's feasibility from an information standpoint, with only a 12% loss of data. The most efficient statistical model was the neural network model, which notably showed optimal performance with a configuration of 10-10-1, combining 10 inputs, 10 hidden layers, and one output signifying cost overrun.

This study mainly contributes to prediction of cost overrun in future residential projects of Mumbai. Recommendations for future work can involve including more dataset and more geographical location. This method can also be utilized for different calculation and different projects involved in Construction Projects.

## References


[1] Swapnil P Wanjari, Gaurav Dobariya, “Identifying factors causing cost overrun of the construction projects in India,” Indian Academy of Sciences, vol 41, pp. 679-693, 2016.

[2] Bichoy Samir Tamez Aziz, Mohamed Badawy Abd El-Megeed, and Ayman Hussein Khalil, “Critical Factors affecting Cost Variation in Residentail Construction Projects in Egypt,” 1st International Conference on Innovative Building Materials, 2014

[3] Shemi, S,N and Asok Kumar N, “Examination of cost overrun in Highway Projects using Artificial Neural Networks in Kerala,” International Journal of Innovative Science and Research Technology, vol 5(3), 2020.

[4] V.B. Chandanshive and A.R. Kambekar, “Estimation of Building Construction Cost using Artificial Neural Networks,” Journal of Soft Computing in Civil Engineering, June 2019.

[5] Serdar Durdyev, Syuhaida Ismail and Nooh Abu Bakar, “Factors causing cost overrun in construction of residential projects; case study of Turkey,” International Journal of Science and Management, vol 1(1), pp. 3-12, 2012.

[6] Poonam Raykar and Ghadge A.N. “Analyzing the critical factors influencing the time overrun and cost overrun in construction project,” International Journal of Engineering Research, vol 5(1), pp. 21-25, 2016.

[7] Kareem Mohammad Mahmoud Mostafa and Dr. Ayman Hamdy Nassar, “Using Artificial Neural Network to model Cost Overrun in Real Estate Projects,” International Journal of Research and Science Innovation, vol 5(10), pp. 72-86, 2018.

[8] Ibrahim Mahamid, “Cost Performance for Residential Building Projects,” Journal of Architecture, Planning and Construction Management, vol 11(1), pp. 64-79, 2021.

[9] Chhavi Gupta and Chitranjan Kumar, “Study of factors causing cost and time overrun in construction projects,” International Journal of Engineering Research and Technology, vol 9(10), 2020.

[10] Pranav K. Lende and Aradhana D. Rathod, “Study of factors affecting cost overrun in road construction project,” International Journal of Research in Engineering, Science and Management, vol 1(8), pp. 115-119, 2018.

[11] Thaseena T, Vishnu K, “Analysis of Cost Overrun in highway construction projects using Multiple Regression and Artificial Neural Network, vol 4(4), pp. 2332-2336, 2017.

[12] Yanfen Zhang and Haijun Mo, “Intelligent Building Construction Cost prediction based on BIM and Elman Neural Network,” Research Square, 2023.

[13] Shicheng Liu, Ruidong Chang, Jian Zuo, Ronald J.Webber, Feng Xiong and Na Dong, “Application of artificial neural network in construction management: current status and future directions,” Applied Sciences, October 2021.

[14] Seokheon Yun, “Performance analysis of construction cost prediction using neural network for multioutput regression,” Applied Sciences, September 2022.

[15] Erik Matel, Faridaddin Vahdatikhaki, Siavash Hosseinyalamdary, Thijs Evers and Hans Voordijk, “An artificial neural network approach for cost estimation of engineering services, International Journal of Construction Management, vol 22, pp. 1274-1287, 2021.

[16] Mohammed H. Aliand Abbas M. Abd, “ Using of an Artificial Neural Network with Particle Swarm Optimization Model in prediction of cost and delay in construction projects,” Diyala Journal of Engineering Sciences, vol 14, pp. 78-93, 2021.

[17] Ankita Gupta and Piyali Debnath, “Artificial Neural Network analysis for cost estimation of building projects in India,” 5th International Conference of Contemporary Affairs in Architecture and Urbanism, 2022.

[18] Amirhossein Balali, Alireza Valipour, Jurgita Antucheviciene and Jonas Saparauskas, “Improving the results of the earned value management technique using artificial neural network in construction projects,” Symmetry, vol 12(10), 2020.

[19] Michal JUSZCZYK, Krzysztof ZIMA and Wojciech LELEK, “ Forecasting of Sports fields construction costs aided by ensembles of neural network,” Journal of Civil Engineering and Management, 2019.

[20] Diksha Jadhav, B.A.Konnur and Sanjana Ptil, “Analysis of factors causing cost overruns in residential construction projects,” International Journal of Research in Engineering, Science and Management, vol 3(7), pp. 308-311, 2020.