Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Implementation of Forecasting Company Project Planning and Realization Using Linear Regression Method

Ni Kadek Kusuma Sari[1], I Made Subrata Sandhiyasa[2*], Yuri Prima Fittryani[3], Putu Wirayudi Aditama[4], Ida Bagus Ary Indra Iswara[5]

[1]Department of Informatics, Faculty of Technology and Informatics, Institute of Business and Technology Indonesia, Denpasar, Indonesia
[(2) (*)]Department of Informatics, Faculty of Technology and Informatics, Indonesian Institute of Business and Technology, Denpasar, Indonesia
[3]Department of Informatics, Faculty of Technology and Informatics, Institute of Business and Technology Indonesia, Denpasar, Indonesia
[4]Department of Informatics, Faculty of Technology and Informatics, Institute of Business and Technology Indonesia, Denpasar, Indonesia
[5]Department of Informatics, Faculty of Technology and Informatics, Institute of Business and Technology Indonesia, Denpasar, Indonesia
*Corresponding **author**: subrata.sandhiyasa@instiki.ac.id

ARTICLE INFO

Received: 22 Mar 2025

Revised: 26 May 2025

Accepted: 23 Jun 2025

ABSTRACT

Inaccuracies in planning and realizing project budgets are major challenges in construction project management, especially in high market dynamic environments such as PT Dream Island Development (PT.DID) in Canggu, Bali. This research aims to develop a forecasting system that is able to predict project realization value based on planning value and duration, using simple linear regression method. The research was conducted quantitatively with the CRISP-DM approach and analysis of historical project data in 2022-2024. The results show that the regression model produces a coefficient of determination (R^2) of 0.9996, which means that 99.96% of the variation in the realized value can be explained by the model. The average realized value of the project was recorded at 79.69% of the planning value, indicating a cost efficiency of about 20.31%. Project duration was also found to be negatively correlated to cost efficiency, with each additional week decreasing the realized value by ±Rp 144,354.88. The project progress model showed an R^2 of 0.564. Although the model showed high accuracy, there were violations of several classical regression assumptions. This research suggests the development of more adaptive and integrated prediction models, as well as the exploration of alternative prediction methods to improve accuracy in project decision-making.

Keywords: Simple linear regression, project management, budget realization prediction, project efficiency, CRISP-DM.

INTRODUCTION

In the context of a public sector organization's budget, the budget includes plans about how much money has been made and how much money will be obtained to fund the plan. Financial statements are the end product of an accounting process that aims to present information concerning the financial position for a number of users in decision making, both for internal and external parties. Accurate budgeting and financial forecasting is an important foundation for project success. Budgets not only reflect estimated costs and revenues, but also serve as a managerial tool to control resources, measure efficiency, and guide strategic decision-making .[1], [2]

PT Dream Island Development is a company engaged in contracting, therefore the challenge faced in the company is the difficulty of accurately estimating the time required to complete the project. This uncertainty often delays project completion and ultimately leads to significant cost overruns. Delays in project completion not only have a financial impact, but can also damage the reputation of the project and weaken the confidence of investors and potential buyers[3]–[5] . In addition, these delays can change the dynamics of the property market in Canggu, giving a significant competitive advantage to properties that are completed on time. Therefore, accurately estimating the project duration is critical to the success of the company's entire project. The second challenge of the project was the difficulty in accurately estimating the project cost[5], [6] . This uncertainty in cost estimation often leads to significant funding shortfalls, which can hinder the project execution process and even jeopardize profitability. The complexity of project cost estimation in Canggu was caused by several factors, including fluctuations in the price of building

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

materials, changes in government regulations regarding the development of tourist destinations and local labor market trends. The third challenge of the project was the difficulty in estimating the exact level of customer demand. Uncertainty in understanding and predicting fluctuations in market preferences and demand can result in overstocking of raw materials and finished goods. This situation can lead to various operational problems[7], [8] . When demand is overestimated, projects can risk having excess stock and unsold raw materials or units, thus wasting resources, increasing transportation costs, and decreasing asset value over time. On the other hand if demand is underestimated, the project may run out of raw materials or be unable to meet actual market demand, resulting in lost potential revenue and profits and a poor reputation for the development.

To overcome these challenges, a method is needed that can help in making more accurate and effective decisions. One method that can be used is the linear regression method. This method can be used to build a prediction model that can estimate the length of time for project completion, project costs, and customer demand levels more precisely[9] . With an accurate prediction model, developers can do better planning and budgeting, so as to minimize the risk of delays, cost overruns, and shortages or oversupply. In addition, linear regression models can also be used to identify the factors that have the most influence on these variables[10], [11] , so that efforts can be made to improve and improve project performance. The contribution and purpose of the study is to determine the comparison of the results of the financial statements between implementation and realization. As well as knowing the picture that will occur at the end of the project will be much different from planning or close to planning.

METHODS

This study uses a quantitative approach by applying the simple linear regression method to forecast project revenue and analyze the conformity between planning and realization of financial statements[12], [13] . The research was conducted at PT Dream Island Development (PT.DID) located in Canggu, Bali, during the period September 2022 to August 2024. Data collection techniques involved direct interviews, field observations, company documentation, and literature studies to ensure the completeness and validity of primary and secondary data.

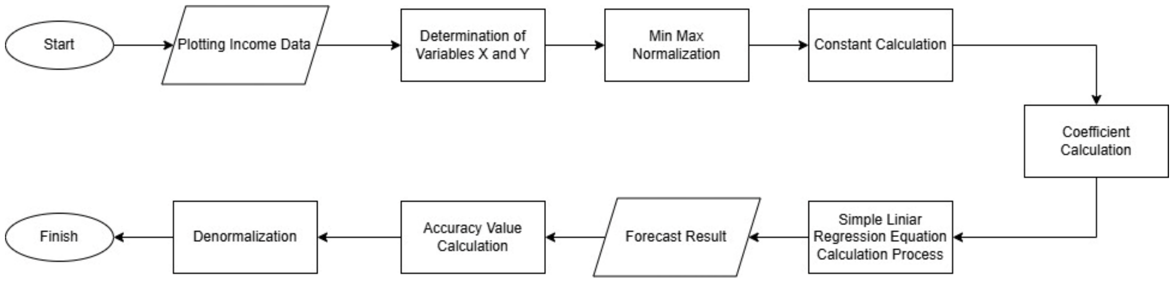

Figure: 1 Simple Linear Regression Calculation Flow

The simple linear regression calculation process in this study follows systematic steps as shown in Figure 1. The first step starts with plotting the revenue data, where historical data of project revenue is analyzed to visualize patterns and trends that will form the basis of modeling. Next, independent (X) and dependent (Y) variables that play an important role in forming the causal relationship to be predicted are determined. Once the variables are determined, the next step is data normalization using the Min-Max Normalization method. This step aims to change the scale of the data to a certain range (usually between 0 and 1), so as to minimize the influence of the dominance of variables with large scales and maintain model stability when mathematical calculations are performed. The process continues with the calculation of the constant (intercept) and coefficient (slope) which are the main components in the linear regression equation. These two values are obtained from the analysis of the normalized data distribution. After the constants and coefficients are obtained, the next step is the preparation and calculation of a simple linear regression equation[14], [15] , which takes the form:

$$\bar{Y} = a + bX \quad (1)$$

Where $\bar{Y}$= The dependent variable as the estimated/predicted variable. While X = Independent variable, the value of the known variable.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

After the model is formed, a forecasting process is carried out to predict the *Y* value based on the given *X* value. The prediction results are then evaluated using the accuracy value calculation method using RSME, MAPE and confusion matrix[16] . The last step is denormalization, which changes the prediction results from the normalization scale back to the original data scale. This is done so that the final results can be interpreted in the context of financial or actual revenue relevant to managerial decision-making at PT Dream Island Development. By following this flow, the simple linear regression model not only provides reliable historical data-based predictions, but also offers a transparent systematic approach in understanding the relationship between variables in the context of construction project management.

This research framework adopts the CRISP-DM (Cross Industry Standard Process for Data Mining) method which consists of six main stages, namely: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. At the Business Understanding stage[17], [18] , the focus is directed at understanding the context of project revenue based on historical data and company operational conditions. Furthermore, at the Data Understanding stage, researchers used monthly revenue data from February to July 2023. The independent variable (X) is time in months, while the dependent variable (Y) is the amount of revenue per month.

Figure: 2 CRISP-DM method

The Data Preparation stage includes the process of normalizing the data using the Min-Max Normalization method, as well as cleaning the data from inconsistencies that could potentially interfere with the accuracy of the model[19], [20] . The data that has been prepared is then used in the Modeling stage, which is a simple linear regression calculation process that includes determining X and Y variables, calculating constants (intercepts) and regression coefficients (slopes), compiling regression equations and making predictions. This process flow is visualized through the Simple Linear Regression Calculation Flow diagram, which illustrates the stages from data input, processing, to evaluation of prediction results. Furthermore, at the Evaluation stage, model performance is analyzed using two main metrics, namely Mean Absolute Percentage Error (MAPE) and Mean Square Error (MSE). MAPE is used to measure the level of deviation of predictions from actual data in percentage form. The tolerance limit used is <10% for good accuracy. If the evaluation results show low accuracy, adjustments are made at the Data Preparation stage. The final stage, Deployment, is the implementation of the predictive model into the company's financial analysis practices. Prediction results are used as the basis for more precise decision-making in financial planning and project management. To ensure the validity of the model, classical assumption tests were also conducted, including normality tests (using P-P plot and Kolmogorov-Smirnov), heteroscedasticity tests (using Park and Spearman methods), and autocorrelation tests (using Durbin-Watson statistics). Hypothesis testing was conducted through a t-test to assess the significance of the regression coefficients partially. In addition, the correlation coefficient (R) was analyzed to measure the strength of the relationship between variables, and the coefficient of determination (R^2) to evaluate how much the independent variables were able to explain the variation of the dependent variable. This approach provides a systematic, data-driven and replicable foundation for financial decision-making in construction-based projects, and demonstrates the real contribution of applying simple linear regression techniques in modern project management.

Data collection

This study uses several main variables in data analysis, namely: Planning Value: the value of the budget planned for each project in units of Rupiah (IDR); Realization Value: the actual value of expenditures incurred in the implementation of the project in units of Rupiah (IDR); Project Progress: the percentage of project completion measured weekly (in %); Project Duration: the length of time to complete the project measured in weeks; and Time

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

Point: a time marker divided into several milestones (Week 1, Week 25%, Week 50%, Week 75%, and Week 100%). These variables are the basis for conducting a linear regression analysis that will be used to predict the realized value based on the planning value and project progress.

Table 1: Company Income and Expenditure Data

Year	Total Revenue	Total Expenses
2022	IDR 34,601,000,000	IDR 27,236,000,000
2023	IDR 63,639,000,000	IDR 49,598,000,000
2024	IDR 174,404,000,000	IDR 136,753,000,000

Data preprocessing

Data preprocessing is an important stage in linear regression analysis to ensure data quality and suitability. In this study, project data at PT.DID was collected from internal documentation, including information such as planning and realization budget, weekly progress, and project duration. Initial identification was done of the main relevant variables, and the data structure was checked to ensure format consistency and data integrity. The data cleaning stage involved handling missing values by imputation (using mean or mode) and removal of incomplete data. Duplications were removed to avoid bias, and outliers were detected through box plot and IQR methods. Outliers due to errors were removed, while extreme outliers were handled with winsorizing techniques. Next, data selection was performed by selecting eight key projects that had complete data, and features with high correlation were eliminated to prevent multicollinearity. The data was then normalized using min-max scaling, converting project progress to a percentage of 0-100%. Budget values were normalized based on the ratio of realized to planned, which describes the budget status of the project. Further transformations were performed, including categorization of project duration (short, medium, long), calculation of average weekly progress, as well as S-curve analysis of cumulative progress. Deviations between realization and planning were calculated as indicators for root cause analysis. This process was complemented by validation of linear regression assumptions, including normality (Shapiro-Wilk or Kolmogorov-Smirnov), homoscedasticity, linearity, and multicollinearity (with VIF) tests. If assumptions are violated, data transformation or feature merging is performed. Finally, denormalization is used to return the results to the original scale for easier interpretation. The preprocessing results produced a high-quality, statistically valid dataset ready to be analyzed with linear regression to understand the relationship between PT.DID's project planning and realization.

Descriptive Statistics

Analysis of the project duration distribution shows that the majority of projects have a duration of around 40 weeks, as seen in the histogram graph of project duration distribution. From the graph, three main groups of project durations can be identified: a) Short-term projects (around 15-20 weeks): around 15-20 projects. b) Medium-term projects (around 40 weeks): around 50 projects. c) Long-term projects (around 50 weeks): around 30 projects.

This distribution shows that the company mostly handles projects with a duration of around 40 weeks, which can be categorized as medium-term projects. This distribution also shows that there are vacant projects in the duration range of 25-35 weeks, which indicates the company's preference in managing projects based on their duration. The comparison between the planning and realization values for each project shows an interesting pattern. From the comparison chart of planning and realized values per project, it can be seen that most projects have lower realized values than planning values. This shows that in general, the projects handled by PT Dream Island Development were successfully completed at a lower cost than the planned budget. However, there are several projects with relatively large budget values (around 6-7 billion rupiah) that have a significant difference between the planning and realization values. Other projects with a budget value of around 1-2 billion rupiah generally have a less significant difference. This shows that the larger the project value, the greater the potential deviation between planning and realization, which is an important finding for future project risk management.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

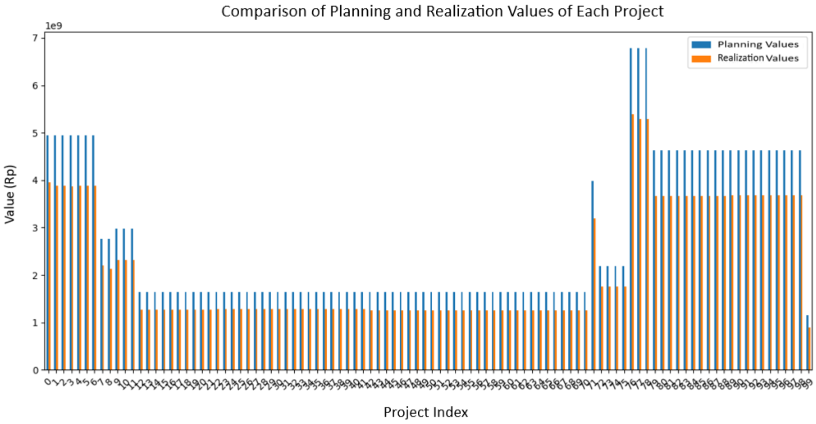

Figure: 3 The Value of Planning and Realizing Each Project

RESULTS AND DISCUSSION

Linear Regression Model Development

Linear regression modeling for the relationship between project progress and time was carried out using historical data on project implementation at PT.DID. This modeling aims to understand the pattern of project development over time and create a predictive model that can estimate project progress based on the duration that has been running. Analysis of historical data shows a linear relationship between time variables (in weeks) and project progress (in percentage). Based on the regression analysis conducted, the progress vs time model equation is obtained as follows

Progress = 0.1860 + 0.0187 ×

This equation shows that there is an initial value (intercept) of 0.1860 or 18.60%, which can be interpreted as the initial progress of the project. This indicates that the project at PT.DID has an initiation stage that has reached around 18.60% when the implementation period begins. Meanwhile, the coefficient value for the Week variable of 0.0187 indicates that for every additional 1 week, the project progress is estimated to increase by around 1.87%. Based on this model, it can be predicted that the project will reach full completion (progress = 100%) after about 43-44 weeks from the start of the implementation period.

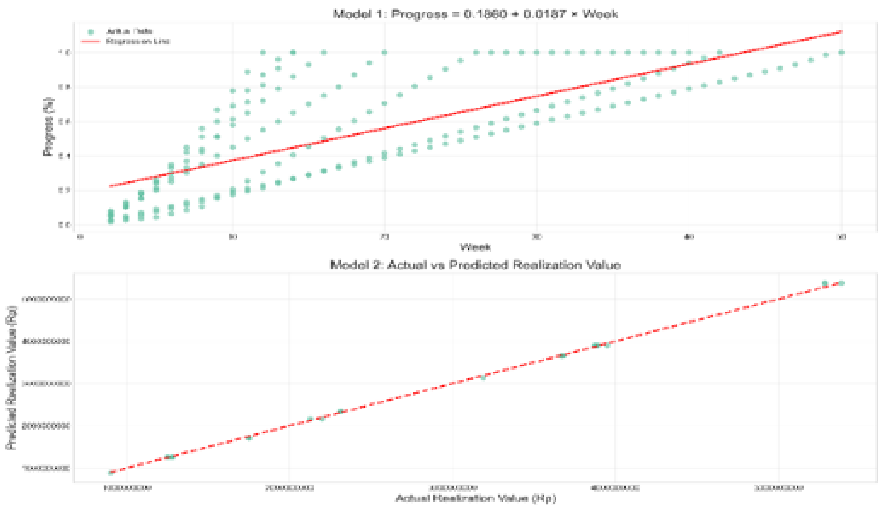

Figure: 4 Model Equation

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ **Research Article**

The visualization of the progress vs time regression model shown in Figure 4 shows the distribution of actual data compared to the regression line. It can be observed that most of the actual data is spread around the regression line, indicating that the linear model is representative enough to describe the relationship between time and project progress. However, there are some data points that deviate quite far from the regression line, especially in the early and late periods of the project, indicating that there are certain periods where the project progress fluctuates faster or slower than the model predicts.

Realized Value Regression Model

Modeling Project realization value modeling is performed to analyze the relationship between the realization value and the planning value and project duration. This approach provides insight into how the realized value is affected by the original plan and the time required to complete the project. Unlike the progress model that only involves one independent variable, the realized value model is developed as a multiple regression model that integrates the planning value and project duration variables. Based on the analysis of project data at PT.DID, a significant relationship between realized value and planning value and project duration was identified. The analysis results in the following realized value regression model:

Realized Value = -31,292,880.15 - 144,354.88 × Duration + 0.796891 × Planning Value

The model shows that the realized value has a negative base constant of IDR 31,292,880.15, which has practically no direct interpretation in the context of this project. The negative coefficient on the Duration variable (-144,354.88) indicates that every additional 1 week of project duration tends to decrease the realized value by about IDR 144,354.88. This finding suggests better efficiency on projects with shorter durations, which may be due to reduced overhead costs and resource optimization. Meanwhile, the coefficient for Planning Value of 0.796891 indicates that on average, the realized value of the project is about 79.69% of the planned value, indicating a cost efficiency of about 20.31% of the initial estimate.

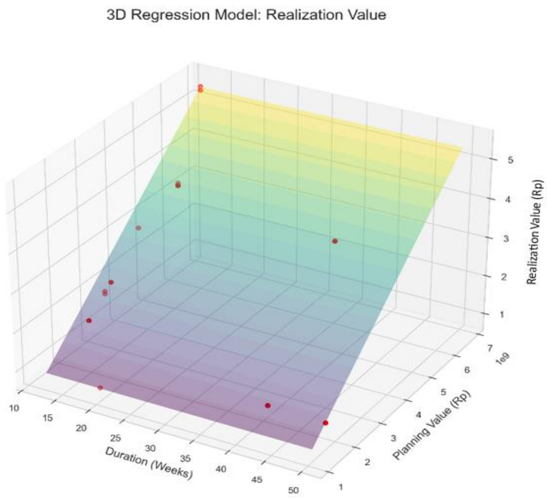

Figure: 5 Value Realization

The 3D regression model of realized value shown in Figure 5 visualizes the relationship between the three variables in the form of a three-dimensional surface. This visualization shows how the realized value (z-axis) changes with changes in project duration (x-axis) and planning value (y-axis). The actual data points (marked in red) scattered around the model surface show the model's fit with the actual data, although there are some points that have a certain deviation from the model surface.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/ Research Article

Regression Model Evaluation

The evaluation of the progress vs. time regression model was conducted to assess how well the model could describe the relationship between project implementation time and progress achieved. The coefficient of determination (R^2) is one of the main evaluation metrics used. Based on the analysis conducted, the progress vs time model has an R^2 value of 0.564, which indicates that about 56.4% of the variation in project progress can be explained by the time variable. This value indicates that the model has a moderate predictive ability, where more than half of the variation in progress can be explained by the time factor. To measure the level of prediction error of the model, the Root Mean Square Error (RMSE) metric was used. The progress vs time model yielded an RMSE value of 0.2079, which indicates that the average prediction error of the model is about 20.79% of progress. This RMSE value is quite significant, indicating that there is a considerable deviation between the predicted value and the actual value of project progress at some point in time

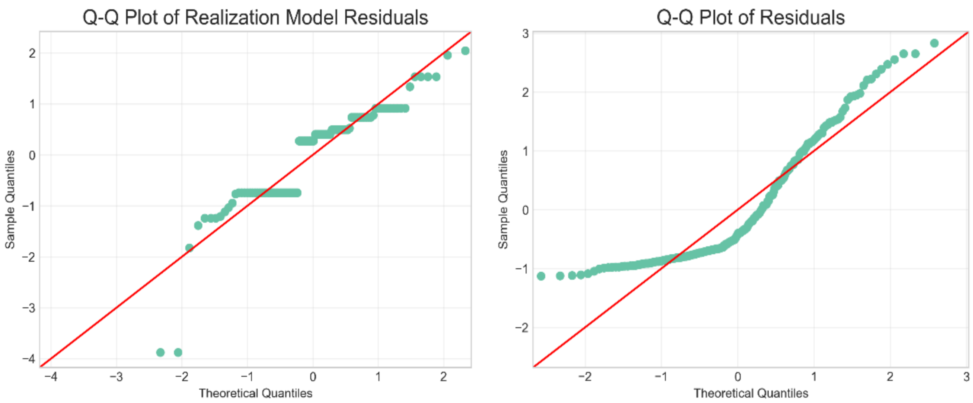

Figure: 6 Normality Test

Residual analysis is performed to identify the pattern of model prediction error. The Residual Q-Q Plot graph in Figure 6 shows the distribution of residuals that tend not to follow a perfectly normal distribution, with deviations especially in the tails of the distribution. Meanwhile, the plot of residuals against predicted values shows a certain pattern that indicates heteroscedasticity, where the variability of residuals is not constant across the range of predicted values. To evaluate the stability of the model, a 5-fold cross-validation was conducted. The cross-validation results yielded an average R^2 value of 0.5408, which is quite close to the R^2 value for the whole model (0.564). This consistency indicates that the model is quite stable and does not experience significant overfitting to the data used. Based on the overall evaluation, the progress vs time model can be considered to have moderate predictive ability. The model can provide a rough estimate of project progress based on time, but there are limitations in prediction accuracy that need to be considered in its use for decision making.

Model Evaluation Realization Value

The evaluation of the realized value regression model is carried out to assess how well the model can predict the project realization value based on the planning value and duration. Based on the analysis conducted, the realized value model has a very high R^2 value, which is 0.9996. This value indicates that 99.96% of the variation in realized value can be explained by the combination of planning value and project duration variables. This almost perfect R^2 value indicates that the model has a very good predictive ability in estimating the realized value based on the two independent variables.

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

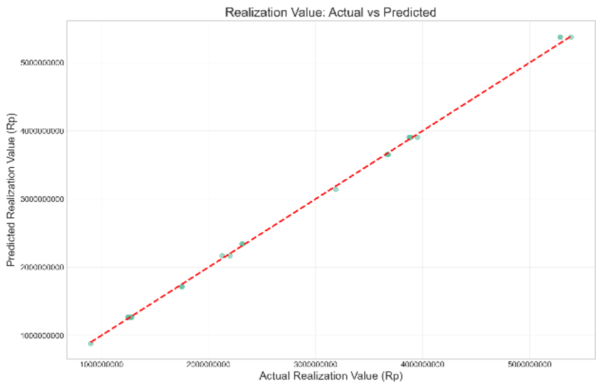

Figure: 7 Actual vs Prediction

The realized value prediction model shows high accuracy with an RMSE of IDR 25.2 million, which is relatively small compared to the project scale of IDR 100-500 billion. The prediction and actual comparison graph (Figure 7) shows the data points are close to the identity line, indicating accurate prediction results. The low MAPE value also indicates minimal prediction percentage error. Stability evaluation through 5-fold cross-validation resulted in an average R^2 of 0.9996, identical to the main model, confirming the stability and generalizability of the model. Overall, the model is effective at predicting project realization values based on project planning and duration data.

Correlation Analysis between Project Variables

Correlation analysis between project variables was conducted to understand the relationships and interrelationships between various factors affecting project planning and realization at PT.DID. The correlation matrix in Figure 7 displays the Pearson correlation coefficient between key project variables. This correlation matrix includes the variables Year, Planning Value, Realized Value, Difference, Percentage, and Duration (Week).

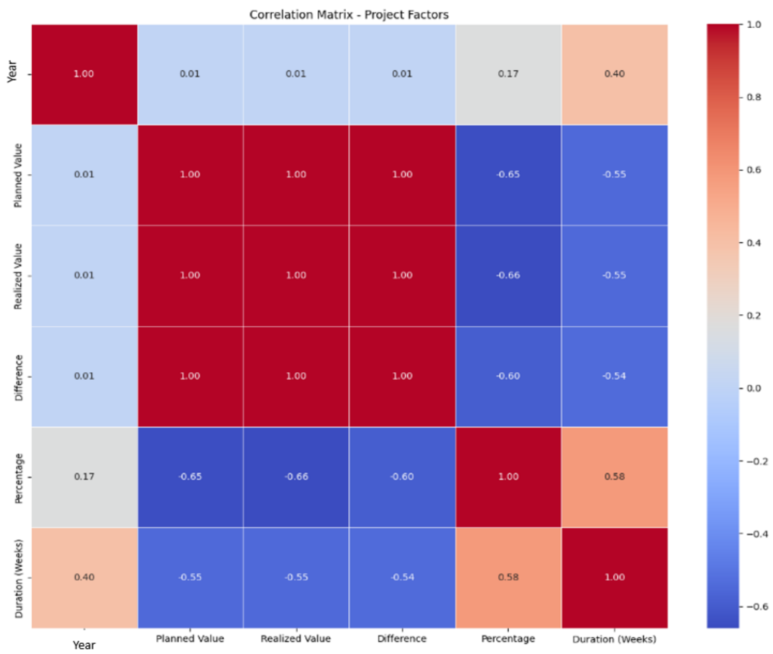

Figure: 8 Correlation Matrix

The results of the correlation matrix analysis show a perfect correlation (r = 1.00) between Planning Value, Realized Value, and Difference, indicating a very strong linear relationship. This is logical because mathematically, the realized

Journal of Information Systems Engineering and Management
2025, 10(4)
e-ISSN: 2468-4376
https://www.jisem-journal.com/
Research Article

value and difference are directly related to the planning. On the other hand, there is a fairly strong negative correlation between these three variables and the Percentage variable (r = -0.65 to -0.66), indicating that the larger the project budget, the lower the target achievement tends to be, possibly due to increased complexity. Meanwhile, the relationship between Year and Project duration showed a moderate positive correlation (r = 0.40), indicating that newer projects tend to last longer, possibly due to the increased scale or challenges of the project. The positive correlation (r = 0.58) between Percentage and Duration indicates that projects with longer duration tend to achieve higher targets, reflecting the importance of time allocation in achieving project success.

Correlation between Project Duration and Budget Difference

A more in-depth analysis of the relationship between project duration and budget variance is shown in the form of a scatter plot in Figure 8 This scatter plot visualizes how these two variables interact and whether any patterns or trends can be identified.

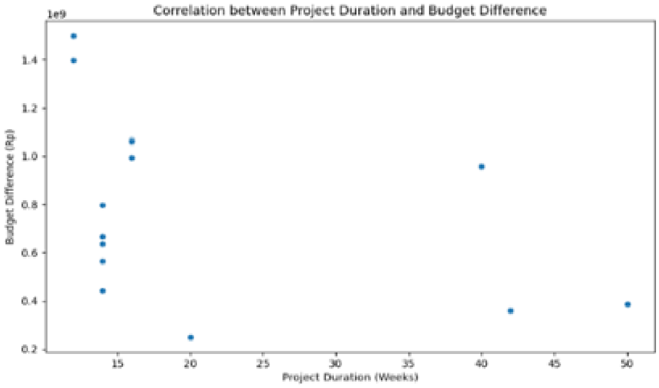

Figure: 9 Correlation between Project Duration and Budget Difference

From the scatter plot, it can be observed that there is considerable variation in the budget variance across various project durations. Some projects with short durations (around 10-15 weeks) show relatively high budget variances, reaching up to IDR 1.4 billion. In contrast, some projects with longer durations (40-50 weeks) show relatively lower budget variances of around IDR 400 million. The data distribution pattern on the scatter plot shows that there is no clear linear relationship between project duration and budget variance. This indicates that factors other than project duration, such as project complexity, type of work, or external factors, may have more influence on the budget variance. However, the general trend that can be observed is that the variability of the budget variance tends to be higher in projects with shorter durations, indicating a greater risk in budget estimation in these projects. The implication of this analysis for project management is that special attention needs to be paid to budget estimation for projects with short durations, as these projects exhibit higher variability in budget variances. A more conservative approach to budget estimation or implementation of special buffers for short duration projects can be effective strategies to manage financial risk.

Implications for Project Management at PT DID Company

The results of the linear regression analysis provide a number of practical implications for project management at DID. The finding of about 20% efficiency between planned and realized values suggests the need to evaluate the initial estimation system. Management can adopt a hybrid approach, i.e. maintaining current estimates while improving accuracy and providing a buffer for uncertainty. In addition, the negative relationship between duration and realized value encourages schedule optimization by targeting shorter duration and more intensive use of resources, while maintaining quality and avoiding increased risk. Project progress curve analysis emphasizes the importance of stricter monitoring, especially in the early and late phases of the project. The implementation of milestones and granular indicators can help early detection of potential deviations. Variations between projects also point to the need for a data-driven management system that can identify success or failure factors, thus supporting continuous improvement. However, there are limitations to the model developed. Linear regression assumes a linear relationship, while the relationship between project variables may be non-linear. The classical assumption tests also showed violations, such as abnormal residuals ($p < 0.05$), positive autocorrelation (DW = 0.267 and 0.809), and heteroscedasticity. This risks producing inefficient estimates and weak statistical inference. The model uses only a few predictor variables, potentially ignoring external factors such as economic or regulatory conditions, as well as internal factors such as technical complexity or project methodology. Finally, since the data is drawn from a project with a specific context (Canggu, Bali), generalization to other projects requires adjustments to local characteristics and timing of implementation.

CONCLUSION

This study concludes that a simple linear regression model can be effectively used to forecast the relationship between planning values and project financial realization at PT Dream Island Development. The developed model shows a very high prediction accuracy with a coefficient of determination (R^2) value of 0.9996, reflecting the ability of the model to explain almost all variations in the realized value based on planning data and project duration. The findings also showed that the average realization only reached 79.69% of the planning value, indicating a cost efficiency of 20.31%. In addition, project duration is shown to have a negative correlation with efficiency, where shorter projects tend to be more cost-effective. Although accurate, this model has limitations in fulfilling classical regression assumptions, such as residual normality and autocorrelation. For further development, integration of this forecasting system with other project information systems and the addition of real-time visualization features are recommended. Future research can expand coverage to different project types and consider relevant external variables. In addition, machine learning-based prediction model approaches and evaluation of additional dimensions such as project risk and quality are worth exploring to improve the accuracy and usefulness of the model in wider project management practice.

REFERENCES

[1] B. Haidabrus, J. Grabis, and S. Protsenko, "Agile project management based on data analysis for information management systems," *Des. simulation, Manuf. Innov. Exch.*, pp. 174-182, 2021, doi: doi.org/10.1007/978-3-030-77719-7_18.

[2] M. Schnegg and K. Möller, "Strategies for data analytics projects in business performance forecasting: a field study," *J. Manag. Control*, vol. 33, no. 2, pp. 241-271, 2022, doi: doi.org/10.1007/s00187-022-00338-7.

[3] R. J. Chapman, "Exploring the value of risk management for projects: improving capability through the deployment of a maturity model," *IEEE Eng. Manag. Rev.*, vol. 47, no. 1, pp. 126-143, 2019, doi: doi.org/10.1109/EMR.2019.2891494.

[4] R. Müller, N. Drouin, and S. Sankaran, "Modeling organizational project management," *Proj. Manag. J.*, vol. 50, no. 4, pp. 499-513, 2019, doi: doi.org/10.1177/8756972819847876.

[5] O. Zwikael and J. R. Smyrk, *Project management: A benefit realization approach*. springer, 2019. doi: doi.org/10.1016/j.promfg.2020.01.208.

[6] D. Ciric, B. Lalic, D. Gracanin, N. Tasic, M. Delic, and N. Medic, "Agile vs. Traditional approach in project management: Strategies, challenges and reasons to introduce agile," *Procedia Manuf.*, vol. 39, pp. 1407-1414, 2019, doi: doi.org/10.1016/j.promfg.2020.01.314.

[7] B.-J. Madauss, *Project Management: A comprehensive description of Theory and Practice*. Springer Nature, 2025.

[8] H. Herlina, "Utilization of Big Data for SWOT Analysis in Improving Business Sustainability of MSMEs," *TECHNOVATE J. Inf. Technol. Strateg. Innov. Manag.*, vol. 1, no. 2, pp. 89-95, 2024, doi: 10.52432/technovate.1.2.2024.89-95.

[9] F. M. Ottaviani and A. De Marco, "Multiple linear regression model for improved project cost forecasting," *Procedia Comput. Sci.*, vol. 196, pp. 808-815, 2022, doi: doi.org/10.1016/j.procs.2021.12.079.

[10] S. Ren, "Optimization of Enterprise Financial Management and Decision-Making Systems Based on Big Data," *J. Math.*, vol. 2022, no. 1, p. 1708506, 2022, doi: doi.org/10.1155/2022/1708506.

[11] K. Khun-anod, C. Limsawasd, and N. Athigakunagorn, "Predicting cost and schedule performance of green building projects based on preproject planning efforts using multiple linear regression analysis," *J. Archit. Eng.*, vol. 29, no. 3, p. 4023025, 2023, doi: doi.org/10.1061/JAEIED.AEENG-1424.

[12] M. Hirman, A. Benesova, F. Steiner, and J. Tupa, "Project management during the industry 4.0 implementation with risk factor analysis," *Procedia Manuf.*, vol. 38, pp. 1181-1188, 2019, doi: 10.1016/j.promfg.2020.01.208.

[13] Z. You and C. Wu, "A framework for data-driven informatization of the construction company," *Adv. Eng. Informatics*, vol. 39, pp. 269-277, 2019, doi: doi.org/10.1016/j.aei.2019.02.002.

[14] G. James, D. Witten, T. Hastie, R. Tibshirani, and J. Taylor, "Linear regression," in *An introduction to statistical learning: With applications in python*, Springer, 2023, pp. 69-134. doi: doi.org/10.1007/978-3-031-38747-0_3.

[15] L. Wang, Y. Guan, and X. Li, "Correlation and Simple Linear Regression," in *Textbook of Medical Statistics: For Medical Students*, Springer, 2024, pp. 157-170. doi: doi.org/10.1007/978-981-99-7390-3_11.

[16] K. D. Pradnyani, I. M. S. Sandhiyasa, and I. M. A. O. Gunawan, "Optimizing Double Exponential Smoothing for Sales Forecasting Using The Golden Section Method," *J. Galaksi*, vol. 1, no. 2, pp. 110-120, 2024, doi: 10.70103/galaksi.v1i2.21.

[17] Y. Pan and L. Zhang, "A BIM-data mining integrated digital twin framework for advanced project management," *Autom. Constr.*, vol. 124, p. 103564, 2021, doi: doi.org/10.1016/j.autcon.2021.103564.

[18] O. Azeroual, R. Nacheva, A. Nikiforova, and U. Störl, "A CRISP-DM and Predictive Analytics Framework for Enhanced Decision-Making in Research Information Management Systems," *Informatica*, vol. 49, no. 18, 2025, doi: doi.org/10.1007/978-3-031-84093-7_4.

[19] S. Jaggia, A. Kelly, K. Lertwachara, and L. Chen, "Applying the CRISP-DM framework for teaching business analytics," *Decis. Sci. J. Innov. Educ.*, vol. 18, no. 4, pp. 612-634, 2020, doi: doi.org/10.1111/dsji.12222.

[20] K. Shameti and B. Cico, "Comparison of Methodological Approaches: CRISP-DM vs OSEMN Methodology Using Linear Regression and Statistical Analysis," in *Balkan Conference in Informatics*, 2024, pp. 47-60. doi: doi.org/10.31449/inf.v49i18.5613.