SIMT ITS
Interdisciplinary School of Management and Technology
Institut Teknologi Sepuluh Nopember

JURNAL TEKNOBISNIS
2023 Vol. 9(2), 01-14

Received 19 Jun, 2023; Revised 31 Jul, 2023; Accepted 21 Aug, 2023 | DOI: 10.12962/j24609463.v9i2.1481

Analysis of Submitting Payments at The End of The Year in the PDAM Budget Management System Using Random Forest Classifier

Fahmi Agil Winata[1,2*], R. Mohamad Atok[3]

ABSTRACT

Submission of payments to providers as part of budget absorption in the budget management system at PDAM Surabaya is an important stage in carrying out the company's activity program. The classic problem is the tendency for payment applications to pile up at the end of the year or in the fourth quarter, resulting in several findings that have not been completed or canceled. Automatically there is also the risk of an increase in cash flow, an increase in the workload in the finance department and a lack of quality spending or company investment as well as vendor difficulties. This study aims to model the random forest classifier and analyze the tendency of accumulation of payment requests in the fourth quarter. The research process starts with data collection and processing using the Machine Learning method, in this case the development of the Random Forest Classifier model, data testing, and analysis of test results. The performance evaluation of the model shows very good results with near-perfect scores (1.0) for all evaluation metrics used: accuracy, precision, recall, and F1- scores. The AUC (Area Under the Curve) value on the testing data is also very high, although slightly lower than that of the training data. Overall, the evaluation results show that a model that has been trained using training data can very well generalize and predict new data in data testing. This indicates that the model has good performance and can carry out classifications with high accuracy.

KEYWORDS: Budget, Payment, PDAM, Machine Learning, Vendor

[1]Information Technology System Development, PDAM Surya Sembada, Surabaya, Indonesia
[2]Business Analytics, Interdisciplinary School of Management and Technology, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
[3]Department of Statistics, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia
*Corresponding author: fahmiawinata8@gmail.com

1. INTRODUCTION

Surabaya Regional Drinking Water Company (PDAM) is a business unit owned by the regional government engaged in clean water distribution services for the general public, with the vision of "Becoming a Modern Drinking Water Company". Its missions include ensuring transparent financial management for the welfare of the community, building a society that is wise in the use of water, providing efficient and sustainable drinking water, building a work environment that prioritizes integrity and achievement. Apart from serving as a public servant in terms of providing clean water services, PDAM must also carry out another important regional mission, namely as a driver of regional economic growth. Given the importance of the responsibilities carried out, the PDAM must work professionally, effectively, and efficiently in carrying out its business, including how to budget management so that it optimally realizes the company's programs and program owners, in this case the Surabaya City Government, and is responsible for financing routine & non-routine expenditures, as well as other supporting program activities.

The main problem in terms of budget scope is usually the discrepancy between budget absorption and its target. Likewise in this case the case of submitting payments to the PDAM budget management system. The classic problem that often occurs is a condition where payment requests are low at the beginning of the year (early quarter) and increase drastically at the end of the year (Quarter IV). Ideally, a budget can be absorbed or submitted for payment ideally every quarter. In other words, submission of good payments is carried out maximally in the early quarters so that in the final quarter stakeholders are not overwhelmed to absorb the budget. However, the current reality is that there are many phenomena where payment requests are not optimal according to the ideal limit per quarter. This pattern of requesting payments is not yet a reflection of quality disbursement of funds/spending because it is not done in a timely, effective, and efficient manner.

In addition, the accumulation of payment applications in the fourth quarter is supported by the fact that there is still a high accumulation of spending and capital which is not only contractual in nature, but also non-contractual which should be expedited because it requires a procurement process which tends to be simple through direct appointment. The fact that non-contractual spending, which is simpler to procure, also piled up at the end of this year, indicating a tendency for work units to wait and see in carrying out budget execution and activities to avoid the risk of uncertainty, as revealed by (Liebman & Mahoney, 2017) This also breaks the argument that the accumulation of spending / requesting payments at the end of the year is often only associated with obstacles to the tender process. From the top management side, they have warned to be careful about the trend of increasing cash flow in withdrawing funds. Other evidence of the impact of the accumulation is found in the payment application verification report where there are still remaining files that have not been processed or canceled at the end of 2022. Each work unit and Commitment Making Officer will also be preoccupied with submitting these payments, especially since employees in the finance department must work extra to verify or process these payment requests. The work unit is still carrying out the mechanism for withdrawing funds at the end of this year, even though the socialization of withdrawing funds as early as possible has been echoed due to the budget absorption target that must be achieved by the work unit. Meanwhile, the need for this research is also reinforced by the findings on the E-Company Performance system for the work unit (Warrants for Spending Money). With this happening, it can be said that the stakeholders have not been able to optimally run the company's programs which tend to start or finish late towards the end of the year.

Many factors influence the accumulation of budget absorption in the fourth quarter, as has been previously studied. The main factors are planning, implementation, human resources, and procurement of goods/services, which have been studied by Sanjaya (2018). Improving the quality of budget planning and execution has an impact on increasing and accelerating the budget absorption process so that the accumulation of budget absorption at the end of the year can be avoided (Basri, 2017). Meanwhile, Darma (2017) suggested that commitment to budget execution decreases as the level of echelonization increases. The long span of control and the complexity of the activities carried out are inhibiting factors for monitoring by superiors. A number of these studies focus on absorbing the budget at the end of the year, but none of these previous studies has made the scope of the Surabaya Regional Drinking Water Company (PDAM) and the case of requesting payments a specific research object and from the point of view of transactions requesting payments in the budget management system, especially with the implementation of the machine learning method, in this case the Random Forest Classifier.

The Random Forest algorithm, designed by J. Ross Quinlan, is named Random Forest because it is a descendant of the ID3 approach to constructing decision trees. Larose (2013) suggested Random Forest is an algorithm that is suitable for classification problems in Machine Learning and Data Mining. Random Forest maps the attributes of the class so that it can be used to find predictions for data that have not yet appeared. The decision tree itself is a "divide and conquer" approach in studying a problem from a set of independent data depicted in a tree chart. Yahya (2018) revealed that a decision tree is also a set of questions that are arranged systematically, where each question determines branching based on attribute values and stops at the leaves of the tree which are predictions of class variables.

Based on the explanation above, this prompted the researcher to propose a study entitled "Analysis of Submitting Payments at The End of The Year in the PDAM Budget Management System Using the Random Forest Classifier". Given the problem of submitting payments, this research tries to elaborate on the phenomenon of accumulation of payment requests from the perspective of payment submission transaction data on the Budget Management System website and is carried out to analyze what factors influence it. In addition, to formulate preventive and curative action strategies for stakeholders so that they can become recommendations or early warnings to overcome possible accumulations of payment requests and processes that have not been completed or canceled at the end of the year.

2. LITERATURE REVIEW

Literature reviews describe the literature relevant to my study and mostly references from credible journals. (Świecka et al., 2021) with a study entitled "Transaction factors' influence on the choice of payment by Polish consumers" reports that choosing a means of payment that is used daily to make life easier is a consumer financial decision that is influenced by several factors. The results represent an important step toward predicting consumer choice models. This advanced modeling approach has unparalleled advantages over traditional statistical methods. This methodology can measure the importance of each variable and effectively inform which ones should be saved for future modeling in a national perspective which is useful for the development of the payment system in Poland. (Boz et al., 2022) described the increasing use of the dollar and euro for bill payments, and the use of the euro as currency in parts of Africa. Countries that bill more in US dollars (euro) tend to experience greater pass-through of the US dollar (euro) exchange rate to import prices, and their trading volume is also more sensitive to exchange rate fluctuations. (Munoz et al., 2022) with a study entitled "Hierarchical classification for account code suggestion (ACS)" identifies as part of invoice processing, businesses are required to manually classify each line item on an invoice to a specific financial account. Best performance is achieved using BA-HMCN, a DAG structure designed using latent variables which act as induced parents of the target class. Compared to the basic multi-class model, the BA-HMCN was shown to achieve statistically significant improvements in both classification and recommendation evaluation measures. (Flynn & Li, 2023) identifies timely, compliant payments to suppliers as a significant problem in purchasing and supply management (PSM), but research on the determinants is still limited.

Here we use social responsibility, financial technology and digital buying literature, as well as institutional and stakeholder theory to explain variations in supplier payment timing. Supply Chain Finance (SCF) was associated with longer payment times during the pandemic while e-invoices had no effect. Overall, institutional pressures appear to contribute to faster payments but stakeholder-centric fintech and digitization do not. (Das et al., 2020) presented a distributed blockchain-based framework that requires no trust to automatic.

According to Ulfa (2017), the factors causing the accumulation of disbursement of funds at the end of the year include the quality of budget planning, human resources, procurement of work unit goods and services. Sudarwati (2017) identified planning and budget execution factors, PBJ factors and HR factors as inhibiting factors for budget realization. The same thing was revealed by Alfayuni (2021), namely the factors that influence the buildup of budget absorption in the City of Cirebon are the planning and execution of the budget, PBJ and HR factors. In addition to the four factors previously disclosed as factors that influence the accumulation of disbursement of funds at the end of the year, Zaenudinsyah (2016) added administrative factors and internal organizational factors. Heryanto (2012) adds that one cause of delays in withdrawing funds is the factor of replacing supply money. Chen & Xu (2017) determined the best model between the linear regression method and Random Forest in predicting used car prices. The data used in this study were obtained from several used car markets in Shanghai and national used car sales sites from October 2015 to 2016. Factors that influence used car prices include car novelty, economic value, and function. The two methods will be compared into 3 models, namely model 1 with a certain brand and year, model 2 with a certain series, and model 3 universal which uses the entire dataset.

The results showed that both methods were easily influenced by sample size. This is shown in model 1 the regression method is better than the Random Forest method which results are more stable, but the regression model is not recommended for predicting used car prices because the sample requirements are so high. Model 2 shows both methods have better predictive effects and are comparable as the sample size increases. Model 3 shows that the Random Forest model is five times better than the regression model with a predictive ability of 95.06%. This proves that the Random Forest method is the best method for dealing with complex models with many variables and samples compared to simple models with few variables.

3. METHODS

The type of data used in this research is secondary data. The data was obtained from PDAM Surabaya's budget management system database & website which has been published on the website www.ska.pdam-sby.go.id/. The data used is data for submitting payment s/ issuance of 13968 data from 2018 to 2022. The available data is data per item for submitting payments made in the budget management system consisting of data on contractual/SPK and non-contractual types, sub-directors in PDAM Surabaya, type of disbursement destination account, service code, type of invoice, value of SPMU (Order to Issue Money), month from date of PP (Procurement Request), month from date SPK (Work Agreement), month from the date the SPK is completed or the contract ends, month from the BAP (Inspection Minutes) date, month from the date of the SPMU register until the month the SPMU is issued. In total, 15 attributes are used to generate the model.

Modeling the system obtained payment application data, then data processing was carried out which began with data pre-processing which was this stage to eliminate some problems that could interfere with data processing. This is because a lot of data has an inconsistent format. Data pre-processing is carried out to handle missing values in the data, so that the raw data becomes data in a form that is easy to understand and in accordance with research needs. Through data pre- processing, it allows the mining process to run more effectively and efficiently. Because data that has gone through data pre-processing is data that has gone through several stages of cleaning. The stages in pre-processing include:

- Data Cleaning: The first step that needs to be done when preprocessing data is data cleaning. This means that the raw data that has been obtained needs to be re-selected. Then, delete or eliminate data that is incomplete, irrelevant, and inaccurate. By doing this stage, it will avoid misunderstandings when analyzing the data.
- Data Integration: Because data preprocessing will combine several data in a dataset, you must check the data coming from these various sources so that they have the same format.
- Data Transformation: The next process that must be carried out is data transformation. As explained above, data will be taken from various sources that may have different formats. We must equate all the data collected in order to simplify the process of data analysis.
- Data reduction. The last step that needs to be done is to reduce the amount of data (data reduction). The point is that certain conditions require reducing the sample of data taken, but with notes, it will not change the results of data analysis. There are three techniques that can be applied when performing data reduction, namely dimensionality reduction, numerosity reduction, and data compression.

Next, select and define the data to be the dependent and independent variables. After mapping the variables, a descriptive analysis was carried out to describe the state of the data used in the study. The results of the analysis are displayed in various chart forms. In the context of feature selection, Pearson correlation can be used to identify a linear relationship between each feature and the target variable. If a feature has a strong correlation with the target variable, then the feature is considered important in influencing the target variable and can be considered for inclusion in the model. The use of Pearson's correlation & P-Value as a feature selection method allows identification of features that have a high correlation & probabilities with the target variable, so that it can help select features that have a significant influence on the target variable and improve model performance. After that, the Gini Index can be used to determine important variables which can measure the extent to which a feature separates the target class in the dataset. The higher the Gini Index value of a feature, the better the feature is at separating the target class. The implementation uses the code `gini_index`, which are functions from the `skfeature` library used to calculate the Gini Index. These functions will provide a score for each feature in the dataset based on the method used, so that the score can be used to select the relevant variables.

At the stage of modeling the payment submission data using the Random Forest algorithm, it can be described as follows:

- Sharing training data & testing data and implementing Random Forest Classifier modeling on research data. The training data set is used to automatically generate models with machine learning algorithms that classify payment categories. The test data set is used to validate the resulting classification model.

The classification used in this study is of two classes, namely 1 = Non-Quarter IV and 2 = Quarter IV with the number of proportions used being 75% for training data, 25% for testing data.

- Performance Evaluation of modeling results and variables obtained from the Random Forest classification. Determining significant variables based on gain importance is useful for describing data understanding of more important variables in model building and determining which predictions will answer the objectives of this study, namely identifying significant variables that affect the buildup of payment requests in the fourth quarter and predicting the buildup of payment requests in the fourth round in the Budget Management System of PDAM Surya Sembada Surabaya City. Next, form a confusion matrix in which there are 2 classes being assessed, namely the positive class and the negative class. Then, calculating the performance of the Random Forest Classifier model based on accuracy, precision, recall, and F1-score and Area Under curve (AUC) values.

4. RESULTS

Dataset statistics show the number of variables is 15, the number of observations is 13786, there are no duplicate rows, but 23 missing cells are found because the data columns are not filled. The types of variables used in this study consisted of 13 categorical variables and 2 numerical variables. The categorical variables include Quarterly Y as the target variable which consists of the fourth quarter and non-fourth quarter. Meanwhile, the 12 categorical predictor variables consist of contractual/SPK and non-contractual data types, sub-directors in PDAM Surabaya, type of disbursement destination account, service code, type of bill, value of SPMU (Letter for Issue of Money), month from date of PP (Request for Procurement), month from date of SPK (Work Agreement), month from date of completion or expiration of contract SPK, month from date of BAP month from date of SPMU register to month of issuance of SPMU.

Feature Selection

Person correlation, or Pearson correlation, is a statistical measure used to measure the linear relationship between two variables. Pearson's correlations range from -1 to 1, with a value of 1 indicating a perfectly positive correlation, a value of -1 indicating a perfectly negative correlation, and a value of 0 indicating no linear correlation between the variables. In the context of feature selection, Pearson correlation can be used to identify a linear relationship between each feature and the target variable. If a feature has a strong correlation with the target variable, then the feature is considered important in influencing the target variable and can be considered for inclusion in the model. The use of Pearson's correlation as a feature selection method allows identification of features that have a high correlation with the target variable, so that it can help select features that have a significant influence on the target variable and improve model performance.

TABLE 1. Correlation and P-Value

	Correlation	P-Value
TYPE	0.254217062	3.86E-202
TYPE_BILL	0.061846518	3.71494E-13
SUBDIRECTORATE	0.06391715	5.94E-14
CODESERVICES	-0.064385212	3.89636E-14
VALUE_SPMU	-0.043583785	3.09391E-07
REVISED	0.123055228	1.25678E-47
MONTH_PP	0.215330333	3.0666E-144
MONTH_SPK	0.240227256	4.47E-180
MONTH_LASTSPK	0.221008781	4.9378E-152
MONTH_BAP	0.254217062	3.8631E-202
MONTH_REGISTER	0.336134018	0
MONTH_SPMU	0.112157384	8.20869E-40
PPREG	0.267354936	3.7284E-224
REGSPMU	0.037253118	1.22145E-05

P-values are known as probability values. It is defined as the probability of getting a result that is equal to or more extreme than the actual observation. The P-value is known as the marginal level of significance in hypothesis testing which represents the probability of occurrence of a given event. The P-value is used as an alternative point of rejection to provide the smallest significance where the null hypothesis will be rejected. If the P-value is small, then there is strong evidence to support the alternative hypothesis. The level of statistical significance is often expressed in terms of a p-value and a range between 0 and 1. The smaller the p-value, the stronger the evidence and the result must be statistically significant. Therefore, rejection of the null hypothesis is very likely to occur because the p-value becomes smaller. From table 4.1 it can be seen that the p-value is listed as less than 0.05, so it can be concluded that all independent variables or simultaneously have an effect jointly to the target variable Y (Quarterly). This is by following the 0.05 level as the cut-off value of the significance value. That is, if the probability value (significance) is below 0.05 then all independent variables affect the dependent variable and vice versa. It can be seen from the table, we can select features for your code service variable and your value_spmu is not included in the model.

Classification by Random Forest

In Random Forest, each decision tree is constructed independently using a random subset of the training data and a random subset of available features. Then, the final prediction is taken by voting or taking the average of the predictions produced by each tree. The stages of Random Forest Analysis in this case are the encoder label used is the sklearn.preprocessing. LabelEncoder library, the library is part of the scikit-learn (sklearn) library which is used to convert labels or category values into integers. LabelEncoder is used in the data preprocessing stage before performing analysis using the Random Forest algorithm or other machine learning algorithms. LabelEncoder works by associating each unique value in a categorical variable with an integer. For example, from the target variable (Y) "YQuarter" with the values "Quarter IV" and "Non Quarter IV", LabelEncoder will change the value "Quarter IV" to 1 and "Non Quarter IV" to 0. Next Split Datasets, the initial data is divided into two parts, namely training data and testing data. The training data is used to train the model, while the testing data is used to test the performance of the model that has been trained. Variable X is a feature matrix that contains all columns except the "YQuarterly" target column. The Y variable is a vector that contains the values from the target column " YQuarter ".

The division of data into training data and testing data is an important step in the development of machine learning models. The goal is to test model performance on data that has never been seen before. The size of the test data is determined by the value `test_size=0.25`. That is, data testing will have a proportion of 25% of the entire dataset. While the training data will have a proportion of 75%. The `stratify=y` argument is used to ensure that the division of the target class in the test data remains balanced or has the same proportion as the target class in the original dataset. This is useful if the dataset has an unbalanced target class, namely the proportion of the number of samples that are different in each class. By using `stratify=y`, the data sharing will maintain the same proportion of target classes between training data and testing data. By dividing the data into training data and testing data, you can train the model on the training data to adjust parameters and study patterns in the data, as well as test the model's performance on data testing to evaluate the extent to which the model can make accurate predictions on data that has never been seen before.

In the classification model, the Random Forest method is used using the Gini index as a criterion for separating features at each node in the decision tree. The Gini index is used to measure the extent to which features can properly separate the target class. The lower the Gini value, the better the feature is at separating target classes.In addition, in this case the random_state=42 parameter is used. This parameter is used to set the seeds used in randomization.

By setting the same seed, you can ensure that the model built will be consistent and reproducible with the same result every time the code is run.By using the Random Forest method, the model can learn patterns and relationships between features in the training data. This model can then be used to predict the target class based on features in the test data that have never been seen before. By evaluating the performance of the test data, you can find out how far the model is able to generalize and predict classes with good accuracy on new data.

Performance Evaluation

By using the Random Forest method, the model can learn patterns and relationships between features in the training data. This model can then be used to predict the target class based on features in the test data that have never been seen before. By evaluating the performance of the test data, you can find out how far the model is able to generalize and predict classes with good accuracy on new data.

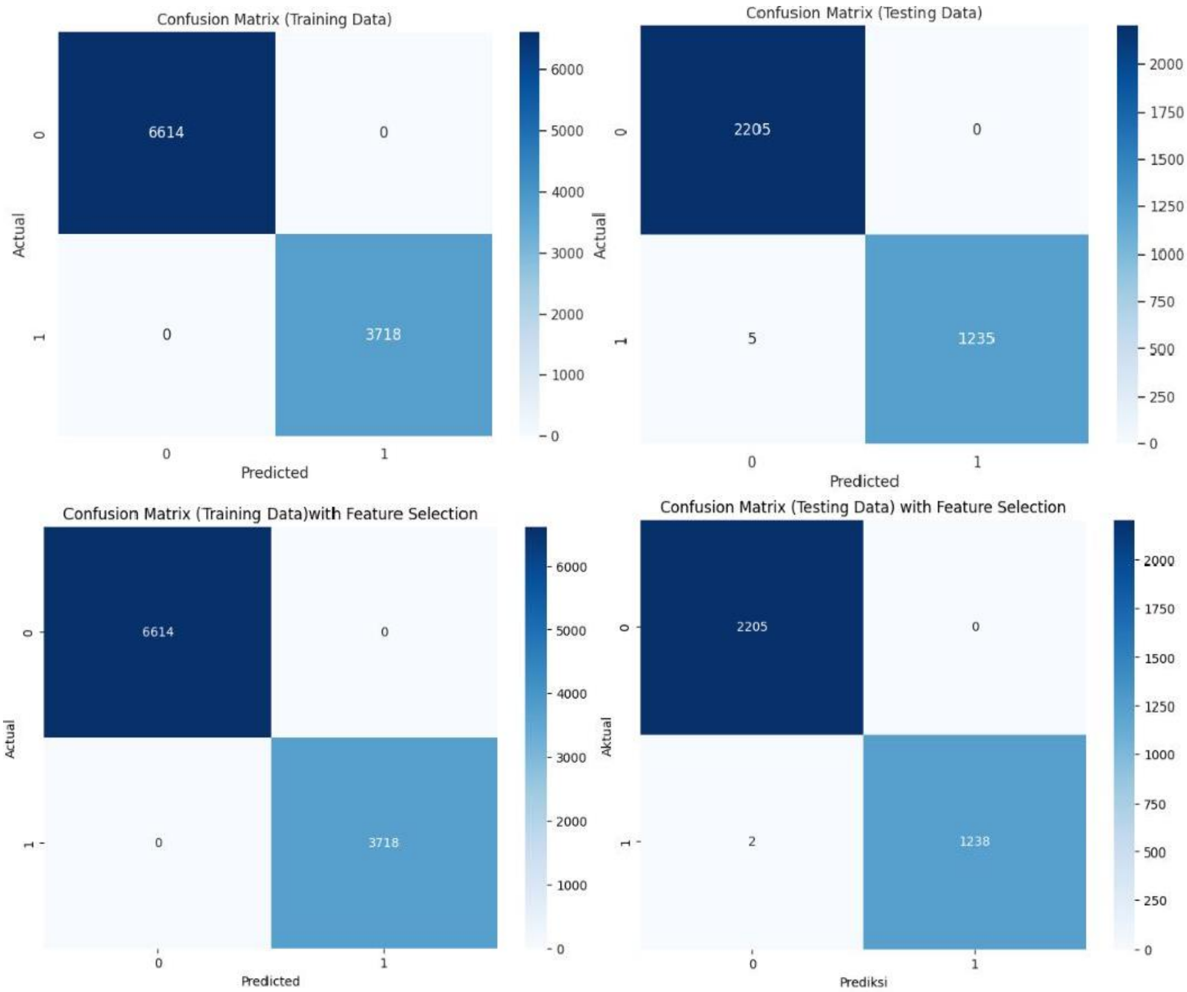

FIGURE 1. Confusion Matrix

In Figure 4.1 shows the confusion matrix of training data and testing data with feature selection and non-feature selection, there are 2 classes being assessed, namely the positive class and the negative class. The following is an explanation for each element in the confusion matrix:

- True Positive (TP): The number of observations that are correctly classified as positive class. In this case, there are 6614 observations of training data and 2205 observations of testing data that are correctly classified as positive class, well that with feature selection and non-feature selection.
- True Negative (TN): The number of observations that are correctly classified as negative class. In this case, there are 3718 observations of training data and 1235 observations of testing data that are correctly classified as negative class, all results are the same, well that with feature selection and non-feature selection.

- False Positives (FP): The number of observations that are incorrectly classified as positive class. In this case, none of the observations from training data and testing data were incorrectly classified as positive (FP = 0), well that with feature selection and non-feature selection.
- False Negative (FN): The number of observations that are incorrectly classified as negative class. In this case, none of the observations of training data were misclassified as negative (FN = 0). By non-feature selection, there are 5 observations of testing data that are incorrectly classified as negative class. But there are 2 observations of testing data that are incorrectly classified as negative class with feature selection.

Overall, the results of the confusion matrix from the training data show that the classification model has very good performance on the training data, with no misclassified observations. The results of the confusion matrix from testing data show that the classification model has a good performance, with most of the observations classified correctly. Even though there were some observations that were incorrectly classified as negative (FN = 5) but become better with feature selection (FN = 2), the number was still relatively small compared to the number of observations that were classified correctly.

```
Performance Evaluation for Training Data :
Training Accuracy: 1.0
Training Precision: 1.0
Training Recall: 1.0
Training F1-score: 1.0
Training AUC: 1.0

Performance Evaluation for Testing Data :
Testing Accuracy: 0.9985486211901307
Testing Precision: 0.9985519048525965
Testing Recall: 0.9985486211901307
Testing F1-score: 0.9985479769962343
Testing AUC: 0.997983870967742
```

```
Performance Evaluation (Data Training) with Feature Selection :
Training Accuracy: 1.0
Training Precision: 1.0
Training Recall: 1.0
Training F1-score: 1.0
Training AUC: 1.0

Performance Evaluation (Testing Data) with Feature Selection :
Testing Accuracy: 0.9994194484760522
Testing Precision: 0.9994199745762098
Testing Recall: 0.9994194484760522
Testing F1-score: 0.9994193457784194
Testing AUC: 0.9991935483870968
```

FIGURE 2. Performance Evaluation

Figure 4.2 shows performance evaluation of the model on the training data shows very good results with near-perfect scores (1.0) for all evaluation metrics used: accuracy, precision, recall, and F1-score. This shows that the model can perfectly predict all samples in the training data. Furthermore, the evaluation of model performance on data testing also produces very high scores and is close to perfect. With feature selection can improve performance for all evaluation metrics. Accuracy, precision, recall, and F1-score are very close, demonstrating consistency of model performance on data that has never been seen before. The AUC (Area Under the Curve) value on the testing data is also very high, although slightly lower than that of the training data. This shows that the model is still very good at distinguishing between positive and negative classes in data testing. Overall, the evaluation results show that a model that has been trained using training data can very well generalize and predict new data in data testing. This indicates that the model has good performance and is able to classify with high accuracy on data that has never been seen before.

5. CONCLUSIONS

The conclusions obtained from the results of the analysis and discussion are have been carried out are some variables exclude 2 variables (service code and SPMU value) by feature selection are proven to have an influence in determining accumulation of payment requests in the PDAM Budget Management System. Starting from the type of contractual / SPK as well as non- contractual, sub-directorates that in PDAM Surabaya, bill type, month from date of PP (Request Procurement), month from date of SPK (work agreement), month from SPK date of completion or expiration of the contract, month from the date of BAP, month from date of registration of SPMU to month the issuance of the SPMU, as well as the number of days between the PP-Register and the distance of days between the Register-SPMU. Evaluation of model performance on training data shows very good results classification with near perfect scores (1.0) for all metrics evaluation used: accuracy (accuracy), precision (precision), recall, F1-score and Area Under Curve (AUC). This shows that the model is perfectly capable of predicting all samples on the training data. Next, performance evaluation of the model on data testing also produces very high values and close to perfect. With feature selection can improve performance for all evaluation metrics. Accuracy, precision, recall, F1-score have value, which is very close, indicating consistency of model performance on data never seen before. PDAM Surabaya or other interested stakeholders can take preventive and curative actions to reduce the accumulation of payment submissions obtained based on the largest Gini index value with the first priority focus on the month of SPMU, followed by PPREG, REGSPMU, Month of Registration, Month of BAP, Final Month of SPK, Month of SPK, Month of PP, Subdirectorate, Type, Type of Bill, and the last Revision.

Work units in PDAM or other interested stakeholders can make preventive actions in the form of more policy formulation measurable, effective, and efficient in optimizing the realization of the submission payment of each stage starting from the PP date (Request Procurement), date of SPK (Work Order), date of BAP (News Examination Procedure), until the date of SPMU (Warrant Issuing Money). From the website application side, it can be applied use of notification features that help users or officials commitment makers are wiser to run step by step until complete on time & quality which can minimize processes that have not been completed or canceled let alone arrived affect the quality of the company's spending or investment. for cases submission of payments that are urgent and related to society can apply the Standard Operating Procedure (SOP). more focused and objective while still adjusting capacity existing budget & not much harm. All programs the company's activities are pursued in accordance with the Work Plan & Budget Companies without any significant & significant major changes be accounted for and strived for in the middle of the majority year work is almost done.

REFERENCES

Alves, V. M. (2018). Development of Web and Mobile Applications for Chemical Toxicity Prediction. Journal of the Brazilian Chemical Society, 29 (5), 982-984.

Boz, E., Casas, C., Georgiadis, G., Gopinath, G., Le Mezo, H., Mehl, A., & Nguyen, T. (2022). Patterns of invoicing currency in global trade: New evidence. Journal of International Economics, 136. https://doi.org/10.1016/j.jinteco.2022.103604

Ceh, M. M. (2018). Estimating the Performance of Random Forest Versus Multiple Regression for Predicting Prices of the Apartments. International. ISPRS International Journal of Geo- Information, 7 (5), 168.

Das, M., Luo, H., & Cheng, J. C. P. (2020). Securing interim payments in construction projects through a blockchain-based framework. Automation in Construction, 118. https://doi.org/10.1016/j.autcon.2020.103284

Flynn, A., & Li, Q. (2023). Determinants of supplier payment times before and during the pandemic: Empirical evidence from UK firms. Journal of Purchasing and Supply Management, 29(4). https://doi.org/10.1016/j.pursup.2023.100850

George, S., dan Sumathi, B. (2020). Grid Search Tuning of Hyperparameters in Random Forest Classifier for Customer Feedback Sentiment Prediction. International Journal of Computer Science and Applications, 11 (9), 173-178.

Hong, J. H. (2020). A House Price Valuation Based on The Random Forest Approach: The Mass Appraisal of Residential Property In South Korea. International Journal of Strategic Property Management, 24 (3), 1-13

Hong, J. H. (2020). A House Price Valuation Based on The Random Forest Approach: The Mass Appraisal of Residential Property In South Korea. International Journal of Strategic Property Management, 24 (3), 1-13

Liebman, J. B., & Mahoney, N. (2017). Do expiring budgets lead to wasteful year-end spending? Evidence from federal procurement. In American Economic Review (Vol. 107, Issue 11). https://doi.org/10.1257/aer.20131296

Munoz, J., Jalili, M., & Tafakori, L. (2022). Hierarchical classification for account code suggestion. Knowledge-Based Systems, 251. https://doi.org/10.1016/j.knosys.2022.109302

Sonmez, R., Ahmadisheykhsarmast, S., Gungor, A. (2022). BIM integrated smart contract for construction project progress payment administration. Automation in Construction, 139, 6.

Sun, Y. & Yang, Y. (2020). The impacts of climate change risks on financial performance of mining industry: Evidence from listed companies in China. Resources Policy, 69.

Świecka, B., Terefenko, P., & Paprotny, D. (2021). Transaction factors' influence on the choice of payment by Polish consumers. Journal of Retailing and Consumer Services, 58. https://doi.org/10.1016/j.jretconser.2020.102264

Trébucq, S. & Magnaghi, E. (2020). Using the EFQM excellence model for integrated reporting: A qualitative exploration and evaluation. Research in International Business and Finance, 42, 522- 531.

Ünal, E. & Shao, J. (2020). A taxonomy of circular economy implementation strategies for manufacturing firms: Analysis of 391 cradle-to-cradle products. Journal of Cleaner Production, 212, 754-765.

Vaio, A. D., Varriale, L. & Alvino, F. (2020). Key performance indicators for developing environmentally sustainable and energy efficient ports: Evidence from Italy. Energy Policy, 122, 229-240

Xia, D., Yu, Q., Gao, Q. & Cheng, G. (2017). Sustainable technology selection decision-making model for enterprise in supply chain: Based on a modified strategic balanced scorecard. Journal of Cleaner Production, 141, 1337-1348.

Xu, L., Zhang, Q., Wang, K. & Shi, X. (2020). Subsidies, loans, and companies' performance: evidence from China's photovoltaic industry. Applied Energy, 260, 221.

Yu, A., Shi, Y., You, J. & Zhu, J. (2020). Innovation performance evaluation for high-tech companies using a dynamic network data envelopment analysis approach. European Journal of Operational Research, 292, 199-212.

Zhao, H. & Li, N. (2020). Evaluating the performance of thermal power enterprises using sustainability balanced scorecard, fuzzy Delphic and hybrid multi-criteria decision making approaches for sustainability. Journal of Cleaner Production,108, 569-582.

How to cite this article:

Winata, F. A., & Atok, R. M. (2023). Analysis of Submitting Payments at The End of The Year in the PDAM Budget Management System Using Random Forest Classifier. *Jurnal Teknobisnis, 9*(2): 01-14. DOI: 10.12962/j24609463.v9i2.1481