Predicting Budget Absorption Categories Using Random Forest and Support Vector Machine Methods

Novardy, Kusumawati Ririen, Hariyadi M.A, Harini Sri, and Imamudin, M.

Abstract— **Budget classification plays a crucial role in planning, management, and budgeting, from implementation to accountability. We create budgets by considering various types of expenditures and funding sources. Each type of expenditure, such as employee salaries, goods, capital, grants, social assistance, subsidies, interest, and non-tax revenue (PNBP) or public service agencies (BLU), has its own set of rules and methods for tracking money. This study aims to demonstrate how budget classification, based on expenditure types and funding sources, is applied in the implementation of the Revenue Budget. This study aims to assess the classification performance of two models, namely the Random Forest Classifier (RFC) and Support Vector Machine (SVM), based on historical data and evaluate the performance of each model. Tests show that the Random Forest model consistently outperforms the SVM model for each data proportion, with a ratio of 90:10 to 60:40. The Random Forest model achieved its best performance at the 80:20 data split, with an accuracy score of 94 percent, a precision score of 94 percent, a recall score of 94 percent, and an F1 score of 87 percent. The average accuracy score of the SVM test results was 80 percent.**

Index Terms— **Budget, Classification, Random Forest, Support Vector Machine, Budget Implementation.**

I. INTRODUCTION

In the implementation of budgets in Ministries and Institutions (K/L), budget classification is very

Novardy is Magister Informatika of Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia (email 230605210009@student.uin-malang.ac.id);

Ririen Kusumawati is Magister Informatika of Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia (email ririen.kusumawati@ti.uin-malang.ac.id);

Muhammad Amin Hariyadi is Magister Informatika of Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia (email adyt2002@uin-malang.ac.id);

Sri Harini is Magister Informatika of Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia (email sriharini@mat.uin-malang.ac.id);

Muhammad Imamudin is Magister Informatika of Universitas Islam Negeri Maulana Malik Ibrahim Malang, Malang, Indonesia (email imamudin@ti.uin-malang.ac.id);

important because it is the basis for state financial management, from planning and budgeting to its implementation and accountability. In the implementation stage, the two most important ways to classify budgets are based on the type of expenditure and the source of funding. These two things are not just administrative technical difficulties; both also have a direct impact on how the budget is managed, how state assets are absorbed smoothly, how state assets are recorded, and how central government financial statements are presented.

The budget is prepared based on how the expenditure is made. With this method, the budget is prepared based on the type or purpose of the expenditure. According to the Ministry of Finance's Standard Account Chart (BAS), ministry/agency spending is divided into eight main categories: employee spending, goods, capital, grants, social assistance, subsidies, interest, and others. This classification is important because each category of expenditure has different accounting characteristics and methodologies. Spending on goods is used for things like food and supplies, while capital spending creates fixed assets that enrich the country. If goods spending and capital expenditure are mixed, the result will have a major impact on financial statements, especially government balance sheets, as assets may not be recorded or even misreported.

Different rules apply regarding how each funding source can be used and reported. SBSN funds must be used for productive activities or projects that create assets. On the other hand, when using PHLN funds, they must be used in accordance with the terms of the loan or grant agreement, which include the types of expenses allowed (permitted expenses), the steps to withdraw funds, and the reporting responsibilities. Classification by funding source, on the other hand, indicates the origin of the money. The State Budget of the Ministry of Finance (KF) collects money from various sources, including Pure Rupiah (RM), Non-Tax State Revenue (PNBP), Public Service Agency (BLU) revenues, State Sharia Securities (SBSN), Foreign Loans/Grants (PHLN), Domestic Loans/Grants (PDN), and direct grants.

When ministries and agencies prepare budgets, it is clear how important it is to separate the source of

income and the type of spending. For example, the Payment Order (SPM) sent to the State Treasury Service Office (KPPN) at the stage of withdrawing funds must be in accordance with the funding source code and the expenditure account code in the Budget Implementation Entry List (DIPA). Coding errors will make it difficult to comply with the budget, which can ultimately lead to SPM being rejected. Similarly, when State Property (BMN) is documented, expenses that should be reported as capital expenditures should not be included in the category of expenditure of goods. These mistakes can lead to incorrect asset reports and audit findings.

Funding a project from one source but executing it with funds from another often leads to problems. This could mean that projects that acquire funding from overseas loans or grants will have to pay for things that are not allowed, which would prevent the lender or grantor from processing reimbursement claims. As a result, the government lost twice: the project funds were non-refundable, and the financial reporting was considered wrong. When expenditures that should come from PNBP are charged to Pure Rupiah, such actions can disrupt the balance of K/L funds.

Many real-life examples show how important it is to categorize things appropriately. Ministries or institutions (KL/L) use funds from the State Treasury (SBSN) to make capital expenditures that improve their infrastructure. If these costs are incorrectly classified as goods expenditure, the SBSN project performance report will become useless, and infrastructure assets will not be recorded in the State Treasury balance (BMN).

This condition shows that the classification of budgets based on the type of expenditure and funding sources goes beyond basic technical administrative issues but is closely related to accountability, transparency, and effectiveness of the use of the state budget. Misclassification can have a major impact on the system, such as delays in budget realization, non-reporting of state assets, misrepresentation of financial statements, and making the public and foreign donors less confident in the state's ability to manage its finances.

Therefore, it is important to examine the categorization of the budget based on the type of expenditure and the source of financing in the implementation of the budget of the Ministries/Institutions (K/L). This study aims to find out the extent to which this categorization affects the effectiveness of budget implementation, the quality of financial accountability, and compliance with laws and regulations. This study can provide suggestions to improve budget governance, i[1]ncluding the implementation of systems, laws, and regulations and the development of the human resources capabilities of ministries/agencies.

The study expects that classifying the budget based on the type of expenditure and financing sources within the K/L budget framework will improve state financial management. This study uses machine learning methods, such as the Random Forest Classifier and Support Vector Machine (SVM), to make the classification process more accurate and appropriate.

Both methods were chosen because they can analyze complex and multidimensional data, such as budget data. Random Forest clarifies the elements that affect categorization, thereby increasing accountability and transparency. On the other hand, SVM is considered resilient in producing highly accurate classification models, thereby increasing effectiveness.

The study is a useful academic resource for comparing the performance of the two algorithms, and it is also very useful in real practice. The proposed categorization method aims to facilitate more transparent, accountable, and efficient state budget governance.

.

II. LITERATURE REVIEW

- **Budget Classification Concept**

Budget classification is a system of grouping state revenues and expenditures according to their characteristics, objectives, and sources of funds. According to the Government Finance Statistics (GFS) Manual published by the IMF, budget classification aims to ensure that government financial statements can be presented consistently, comparatively, and can be understood by various stakeholders, both domestic and international [2].

- **Classification by Shopping Type**

The type of expenditure in the K/L State Budget is an economic category that describes the nature of the use of the budget[2]. In accordance with budgeting regulations, the main types of expenditures used in budget documents are:

a. **Employee expenditure** is salary compensation along with allowances given to the State Civil Apparatus (ASN).

b. **Goods Expenditure,** which is operational and non-operational expenditure supporting K/L activities.

c. **Capital expenditure is** an expenditure in the acquisition of fixed assets and inventory goods with the benefit of more than one budget year.

d. **Social Assistance Expenditure** (Bansos) refers to the transfer of money or goods to the community, serving as a stimulus or form of social protection.

e. **Other shopping,** which cannot be categorized into the above posts but is still regulated according to the regulations.

- **Classification Based on Funding Source**

In addition to the type of expenditure, the management of the K/L State Budget also requires classification based on the source of funds, which describes the origin of financing an activity[3] Some of the main sources of funds are:

a. **Pure Rupiah (RM),** which is funds sourced from domestic revenues (tax and non-tax) without special use bonds.

b. **Non-Tax State Revenue (PNBP),** which is sourced from K/L services to the community, such as education fees, health services, or other services.

c. **Overseas Loans and Grants (PHLN),** which are used to support the financing of certain programs in accordance with agreements with international partners.

d. **Public Service Agency (BLU),** which has flexibility in managing revenue and expenditure based on its own financial patterns.

- **Implications in the Management of the State Budget of K/L**

The classification of the budget based on the type of expenditure and source of funds has strategic implications in the management of the State Budget of the Ministry of Finance [4]:

a. Planning and Budgeting, which provides a clear framework for the preparation of RKA-K/L and the determination of DIPA.

b. Budget implementation ensures that funds are used according to their classification, thereby preventing deviations in expenditure usage.

c. Fiscal Accountability, which supports the presentation of accurate, reliable, and compliant financial statements with Government Accounting Standards (SAP).

d. Performance Evaluation, which facilitates the assessment of program efficiency and effectiveness based on expenditure allocation and funding sources.

e. Audit and Supervision, which is what the Financial Audit Agency (BPK) and the Government Internal Supervisory Apparatus (APIP) use to check if the budget is being followed and if it is being done well

- **Classification Modeling**
Classification is finding models or functions that can describe and differentiate a class of data or concepts, with the aim that the model can be used to predict the unknown class of an object of observation [5].

- **Random Forest Classifier (RFC)**

Random Forest is a flexible and easy-to-use algorithm that produces excellent results most of the time. Random Forest consists of multiple trees, each of which grows through randomization and other processes that involve building several decision trees during training, ultimately selecting classes based on the mode or average predictions from these individual trees. This algorithm is a type of decision tree algorithm, which is an ensemble of decision trees used as a combined base classifier. This algorithm is used for classification and regression analysis [6][7]. The stages in the Random Forest algorithm are as follows:

a. Specify the number of trees (k) to be formed; the number of k is selected from the total feature m, where k is less than m.

b. A random sample was taken as many as N observations (with replacement) on the dataset foreach tree.

c. In each tree, a subset of predictors is also taken randomly. The variable mp represents the sum of the predictor variables.

d. The second and third processes are repeated until there are as many k trees as possible.

e. In classification, the Random Forest prediction results are derived from the majority vote of the classifications made by all trees.

- **Support Vector Machine (SVM)**
Support Vector Machine (SVM) is a new machine learning method used in making predictions, both in the case of classification and regression. SVM provides a statistically significant approach to pattern recognition [1]. SVM will look for the best support vector that separates two different classes using the largest margin. In the classification process, a Support Vector Machine (SVM) algorithm is used to recognize patterns in the labeled training data and then generate a machine learning model [8]. Support Vector Machine (SVM) is a technique developed by Vapnik in 1995 to make predictions, both in the case of classification and regression. This method is a machine learning method and enters the supervised learning class that can be used for the classification of both linear and non-linear data whose basic idea is to maximize the limits of the hyperplane [5].
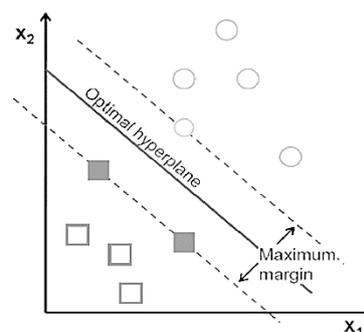

Fig. 1. Optimal Hyperplane dengan Maximum Margin

The calculation of the hyperplane margin is used to obtain the most optimal hyperplane line that separates the data into the two classes and finds the maximum point. In obtaining a hyperplane in SVM [8], you can use equation (1).

$$(w. xi) + b = 0 \tag{1}$$

The data point xi, which belongs to class -1, can be formulated as follows:

formulated as in equation (2).

$$(w. xi + b) \leq 1, yi = -1 \tag{2}$$

While the xi data belonging to the +1 class can be formulated as in equation (3).

$$(w. xi + b) \geq 1, yi = 1 \qquad (3).$$

- **Evaluation Model**

The classification model's performance calculation is based on testing both correct objects and false objects. The classification performance calculation used in this study is a confusion matrix which contains the calculation of actual predictable classification result[9].

a. Accuracy

Accuracy is a measure of the performance of a classification model and is the percentage of the amount of data correctly predicted out of the total data. The criteria for determining the classification are using the confusion matrix [10]which can be seen in Table 1 Confusion Matrix

Table 1. Confusion Matrix

Class	Classified as Positive	Classified as Negative
Positive	True Positive (TP)	False Positive (FP)
Negative	True Negative (TN)	False Negative (FN)

Table 1 shows the classification of the confusion matrix. True positive (TP) indicates that the classification model correctly labels the number of true positives. True Negative (TN) indicates that the classification model correctly labels the number of negative tuples. False positive (FP) indicates that the classification model incorrectly labels the number of negative positives. False Negative (FN) indicates that the classification model is labeling the wrong number of positive tupbles. Accuracy is a measure of the performance of a classification model and is the percentage of the amount of data correctly predicted out of the total data. The equation for calculating precision can be seen in Equation 1.

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (1)$$

Performance metrics consist of precision, recall and f1-score. Precision is the ratio of positive or degree of reliability, which is the proportion of a correctly labeled positive prediction to an overall positive prediction[10]. The equation for calculating precision can be seen in Equation 2

$$Precission = \frac{TP}{TP + FP} \qquad (2)$$

Recall is also known as true positive rate or sensitivity. Recall is also referred to as the degree of reliability of the model in correctly detecting positively labeled data. The equation for calculating the recall can be seen in equation 3.

$$Recall = \frac{TP}{TP+FN} \qquad (3)$$

F1-score summarizes all the results of precission and recall calculations by creating harmonic averages[10]. The equation for calculating F1-score can be seen in equation 4.

$$F1 - Score = \frac{2\ x\ Precission\ x\ Recall}{Precission + Recall} \qquad (4)$$

III. RESEARCH METHODOLOGY

In accordance with the formulation of the problem and the purpose of the research, the type of research chosen is quantitative research with a machine learning approach for classification analysis. The selection of this method is based on the purpose of the research, which is to describe, analyze, and evaluate the accuracy of budget classification and its impact on the accountability of state budget management in ministries/institutions.
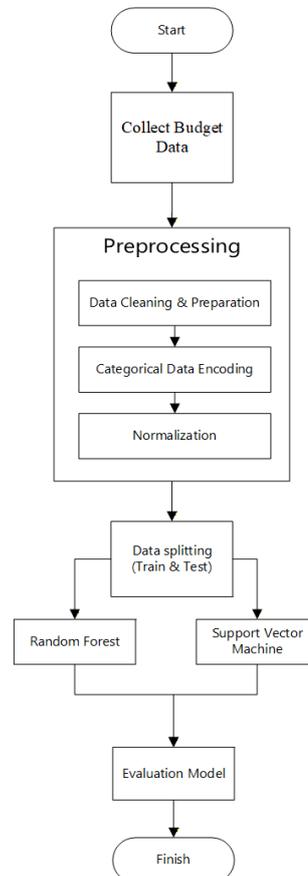

Fig. 1. Desain System

A. Data Collection

The data used in this study is collect official documents (DIPA, realization report, BAS, financial statements, audit results) and their revisions from the last few years as well as the Budget Realization Report (LRA) or detailed historical financial transaction data (up to the level of the expenditure treasurer/Payment Order SPM), which contains information: Account Code, Expenditure Description, Source of Funds, Ceiling, Realization, Program/Activity Code, and Organizational Unit. The dataset used is can be seen in Table 2 as follows:

Table 2. Data set

kode_ke giatan	uraian	nilai_anggaran	jenis_be lanja	sumber _dana	realisas i_angga ran	%_realisa si
101	Operasional dan layanan perkantoran Pendidikan	486,588,000	Belanja Barang	RM BOPTN	75,118,0 00	15.44%
102	Bimbingan Teknis Penulisan Karya Ilmiah Bagi Mahasiswa FITK	50,880,000	Belanja Barang	RM BOPTN	47,394,5 00	93.15%
103	Pengabdian Kepada Masyarakat	165,070,000	Belanja Barang	RM BOPTN	114,854, 000	69.58%
104	Evaluasi Kurikulum Program Studi di FITK	16,200,000	Belanja Barang	RM BOPTN	0	0.00%
105	Penjaminan Mutu FITK	40,390,000	Belanja Barang	RM BOPTN	0	0.00%
106	Pendampingan Publikasi Ilmiah FITK	82,000,000	Belanja Barang	RM BOPTN	990,000	1.21%
107	Peningkatan Mutu Riset dan Academic Writing	105,446,000	Belanja Barang	RM BOPTN	0	0.00%
108	Program Pembinaan Prestasi Mahasiswa	169,470,000	Belanja Barang	RM BOPTN	50,400,0 00	29.74%
109	Character Building Ormawa	35,200,000	Belanja Barang	RM BOPTN	0	0.00%
110	Kerjasama Seleksi dan Penempatan Lulusan	30,400,000	Belanja Barang	RM BOPTN	22,300,0 00	73.36%
111	Monev Pembelajaran/ Evaluasi Penilaian Pembelajaran	9,660,000	Belanja Barang	RM BOPTN	8,070,00 0	83.54%
112	Layanan Operasional BOPTN FITK	577,203,000	Belanja Barang	RM BOPTN	96,639,5 40	16.74%

After the data is prepared, the next step is to determine the budget implementation category based on the absorption ratio value. The category function (ratio) is created to group the ratio value into three levels, namely. Adding the Implementation Category Column (Target / Y) ram checks the existence of the 'ratio' column in the dataset.

- Use the category () function to classify budget absorption levels based on ratio values:
- High Absorption: ratio ≥ 90%
- Medium Absorption: ratio ≥ 60% and < 90%.
- Low Absorption: ratio < 60%.
- The classification results are added as a new 'implementation' column using the apply () method.

The above process yields the following data ratios:

- Activity 102 has a ratio of 93.15% → Serapan Tinggi
- Activities 103 and 110 have ratios of 69.58% and 73.38% → Serapan Sedang
- Other Activities with ratios below 60% → Serapan Rendah

The data results from adding the implementation category column are as follows:

Table 3. Budget Absorption Level Categorization Results

kode_kegiatan	uraian	nilai_anggaran	jenis_belanja	sumber_dana	realisasi_anggaran	%_realisasi	rasio	pelaksanaan
0	101	Operasional dan layanan perkantoran Pendidikan	486588000	Belanja Barang	RM BOPTN	75118000	0.15	15.44	Serapan Rendah
1	102	Bimbingan Teknis Penulisan Karya Ilmiah Bagi M...	50880000	Belanja Barang	RM BOPTN	47394500	0.93	93.15	Serapan Tinggi
2	103	Pengabdian Kepada Masyarakat	165070000	Belanja Barang	RM BOPTN	114854000	0.70	69.58	Serapan Sedang
3	104	Evaluasi Kurikulum Program Studi di FITK	16200000	Belanja Barang	RM BOPTN	0	0.00	0.00	Serapan Rendah
4	105	Penjaminan Mutu FITK	40390000	Belanja Barang	RM BOPTN	0	0.00	0.00	Serapan Rendah
5	106	Pendampingan Publikasi Ilmiah FITK	82000000	Belanja Barang	RM BOPTN	990000	0.01	1.21	Serapan Rendah
6	107	Peningkatan Mutu Riset dan Academic Writing	105448000	Belanja Barang	RM BOPTN	0	0.00	0.00	Serapan Rendah
7	108	Program Pembinaan Prestasi Mahasiswa	169470000	Belanja Barang	RM BOPTN	50400000	0.30	29.74	Serapan Rendah
8	109	Character Building Ormawa	35200000	Belanja Barang	RM BOPTN	0	0.00	0.00	Serapan Rendah
9	110	Kerjasama Seleksi dan Penempatan Lulusan	30400000	Belanja Barang	RM BOPTN	22300000	0.73	73.36	Serapan Sedang

B. Research Concept

The conceptual framework in this study aims to provide an overview of the thought process in analyzing budget data based on expenditure types and funding sources to predict budget implementation. To further clarify this concept, we have presented it in a flowchart, as shown in the as shown in the figure 2 below.
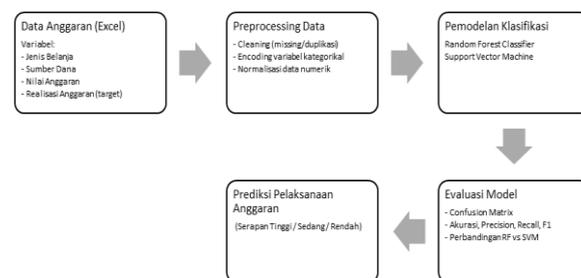

Fig. 2. Research Concept

The flow of the Concept Framework can be described as follows:

1. **Budget Data**
   The dataset used contains the following variables:
   - kode_kegiatan
   - uraian
   - nilai_anggaran
   - jenis_belanja
   - sumber_dana
   - realisasi_anggaran (target variable Y)
2. **Input Variable (X)**. Shopping Type, Source of Funds and Budget Value (if you want to add this feature)
3. **Preprocessing Data**
   - Get rid of duplicates and missing values in the data.
   - Encoding categorical data (type of spending, source of funds)
   - Normalizing or standardizing numerical data
4. **Classification Method**
   - Random Forest Classifier
   - Support Vector Machine
5. **Model Evaluation**
   - Confusion Matrix
   - Accuracy, Precision, Recall, and F1-Score
   - RFC vs. SVM performance comparison
6. **Budget Implementation Prediction (Y)**
   - The best models are used to predict Realisasi_Anggaran.

C. Random Forest Design

In the classification process used based on the Random Forest model with a training framework in budgeting based on expenditure categories and financing sources in the context of budget implementation, we will clarify this concept in the form of a flow diagram, as shown in the figure 3 below.
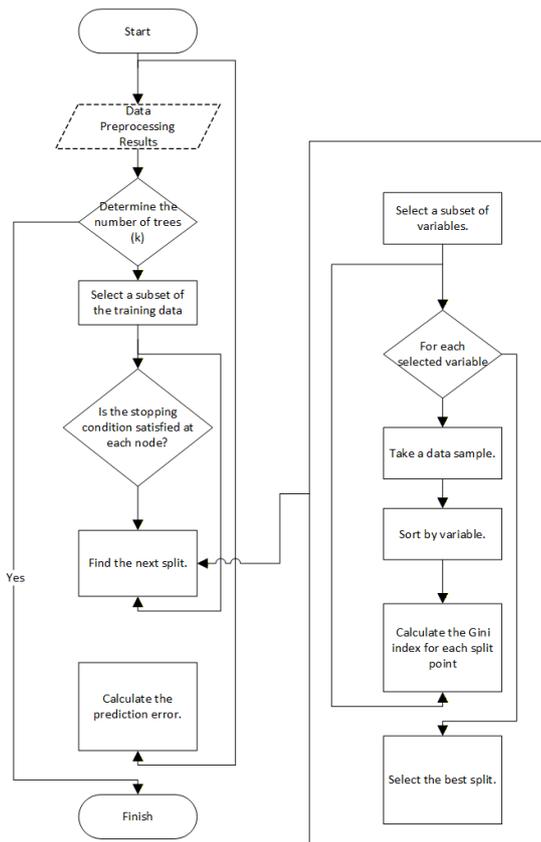
Fig. 3. Random Forest Model
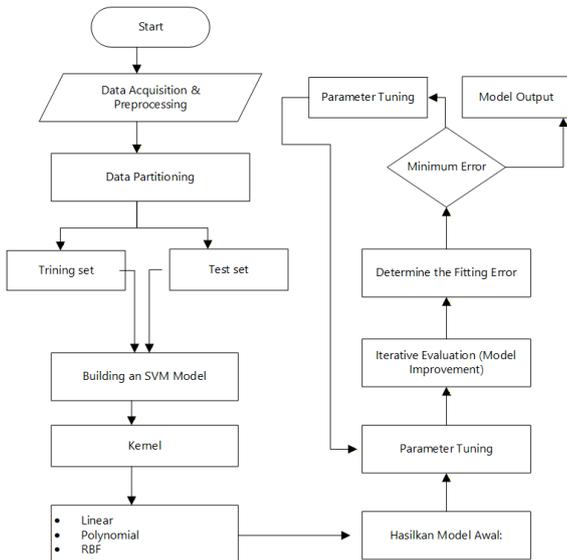
D. Support Vector Machine

Fig. 4. Random Forest Model

Figure. 4. shows a basic overview of how an SVM model is constructed. To create an SVM classifier, a kernel function and its values are required. The first step in creating an SVM model is collecting the dataset and performing some preparation. This includes handling missing values, storing categorical variables, and scaling numerical features. These steps are necessary to ensure that the data structure is suitable for training a machine learning model. Various kernel functions include polynomial functions, radial basis functions, and

linear classifiers. The training data is fed into the SVM algorithm, which attempts to find the best hyperplane in the feature space to separate as many classes as possible.

IV. RESULT AND DISCUSSION

- Random Forest Test Results

From the 5 trials that have been conducted, the average values for Accuracy, Precision, recall, and f1-score are listed in Table 4.

Table 4. Random Forest Test Results

Training	Testing	accuracy	precision	recall	f1-score
90	10	87%	79%	87%	0.82
85	15	87%	82%	87%	0.82
80	20	94%	93%	94%	0.93
70	30	90%	90%	90%	0.89
60	40	90%	90%	90%	0.90
Rata-Rata		**90%**	90%	87%	90%

After five classification tests using the Random Forest algorithm and varying Training and Testing data splits (from 90:10 to 60:40), the model consistently performed very well. The average Accuracy score was 90%, meaning the model correctly guessed the class 90% of the time. This success was supported by an average Precision score of 90% and an average F1 score of 90%, indicating a good combination of reducing False Positives and False Negatives. The model missed about 13% of true positives, as indicated by a slightly lower average Recall of 87%. However, the fact that the overall average measure was above 87% indicates that Random Forest is a good and reliable model for this classification task.

- Support vector Machine Test Results.

Table 5. Support vector Machine Test Results

Training	Testing	accuracy	precision	recall	f1-score
90	10	86%	79%	86%	82%
85	15	87&	75%	87%	81%
80	20	87%	76%	87%	81%
70	30	87%	75%	87%	80%
60	40	87%	75%	87%	80%
Rata-Rata		87%	87%	75%	87%

Table 2 shows the results of the Support Vector Machine (SVM) classification test five times with different training and testing data split ratios (from 90:10 to 60:40). The SVM model consistently demonstrated consistent results. The program was able to make accurate predictions 87% of the time, which is a very high rate. The average Precision score was 75%, the average Recall score was 87%, and the average F1 score was 80%. Although the Precision score was only 75%, this means there were more false positives in the

predictions. However, the total F1 score of 80% indicates that the SVM model achieved a good balance between Precision and Recall for this classification task.

Model Accuracy Comparison as explained in the following Table 6.

Table 6. Model Accuracy Comparison

Pengujian		Accuracy		precision		recall		f1-score	
Training	Testing	RF	SVM	RF	SVM	RF	SVM	RF	SVM
90	10	87%	86%	79%	79%	87%	86%	0.82	82%
85	15	87%	87%	82%	75%	87%	87%	0.82	81%
80	20	94%	87%	93%	76%	94%	87%	0.93	81%
70	30	90%	87%	90%	75%	90%	87%	0.89	80%
60	40	90%	87%	90%	75%	90%	87%	0.90	80%
Rata-Rata		90%	87%	90%	75%	87%	87%	90%	80%

Based on the test results in Table 3. using two main models—Random Forest Classifier (RFC) and Support Vector Machine (SVM), the classification tests to predict how the budget will be used based on the type of expenditure and funding source showed varying levels of success. The Random Forest model performed very well overall, with an average Recall of 85%, Accuracy of 90%, and Precision of 90%. This model also achieved an average F1 score of 90%. Based on these results, the Random Forest model is very good at detecting various types of budget applications in the real world. Furthermore, the SVM model performed well and consistently, with an Accuracy of 87% and recall of 87%, as well as a Precision of 75% and an F1 score of 80%. These figures indicate that both models can correctly predict how the budget will be used; Random Forest performed better in Accuracy and Precision, while SVM performed consistently across a wide range of test data ratios.

V. CONCLUTION

Based on the description in the previous chapter, several conclusions can be drawn as follows.

The application of budget classification based on expenditure type and funding source in the implementation of the State Budget (APBN) in Ministries/Institutions can be clearly identified through analysis of historical data and applicable regulations. The analysis shows that expenditure type and funding source have a significant relationship to the level of budget execution. Routine expenditures, such as operational and honorarium expenditures, tend to be more stable, while capital or maintenance expenditures are more variable depending on the procurement process and field conditions.

1. Measurement Model and Machine Learning Model Performance Analysis.

The Random Forest Classifier (RFC) demonstrated very high and superior performance, with an average accuracy of 90%, precision of 90%, recall of 87%, and an F1-score of 90%. The Support Vector Machine (SVM) demonstrated stable performance, but lower than the RFC, with an average accuracy of 87%, recall of 87%, precision of 75%, and an F1-score of 80%.

2. Best Classification.

The Random Forest Classifier (RFC) proved to be the best model for sorting K/L budget data based on a comparison of ranking measures. This is supported by the fact that the RFC had the highest average F1 score (90%) and the highest Precision (90%), indicating that it can find the best balance between reducing False Positives and False Negatives and providing more accurate predictions for budget execution.

RECOMENDATION

Based on these conclusions, some recommendations that can be submitted include, when implementing the ministry/agency budget, the following suggestions can be made to make it more transparent, accountable, efficient, and aligned with growth objectives:

1. Establish an Early Warning System: Use the Random Forest Classifier model as the primary engine in an early warning system that can quickly identify potential budget performance failures.

2. Focus on Prediction-Based Oversight: Advise lawmakers to focus their audit and oversight efforts on groups of expenditure types and funding sources that the model indicates are likely to be high-risk (based on the RFC Feature Importance analysis).

REFERENCES

[1] R. S. Tantika and A. Kudus, "Penggunaan Metode Support Vector Machine Klasifikasi Multiclass pada Data Pasien Penyakit Tiroid," *Bandung Conference Series: Statistics*, vol. 2, no. 2, 2022, doi: 10.29313/bcss.v2i2.3590.

[2] PMK 102 /PMK.02/2018, *Peraturan Menteri Keuangan RI Nomor 102/pmk.02/2018 tentang klasifikasi anggaran*. 2018. Accessed: Sep. 14, 2025. [Online]. Available: www.jdih.kemenkeu.go.id

[3] PMK 160/PMK.02/2012, *Lampiran i Peraturan menteri keuangan republik indonesia Nomor 160/pmk.02/2012 tentang petunjuk penyusunan dan pengesahan daftar isian pelaksanaan anggaran*. [Online]. Available: www.djpp.depkumham.go.id

[4] PMK 62 Tahun 2023, *Peraturan Menteri Keuanganan Nomor 62 Tahun 2023 tentang Perencanaan Anggaran, Pelaksanaan*

Anggaran, serta Akuntansi dan Pelaporan Keuangan. 2023.

[5] P. R. Sihombing and I. F. Yuliati, "Penerapan Metode Machine Learning dalam Klasifikasi Risiko Kejadian Berat Badan Lahir Rendah di Indonesia," *MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer*, vol. 20, no. 2, 2021, doi: 10.30812/matrik.v20i2.1174.

[6] N. Khasanah, R. Komarudin, N. Afni, Y. I. Maulana, and A. Salim, "Skin Cancer Classification Using Random Forest Algorithm," *SISFOTENIKA*, vol. 11, no. 2, 2021, doi: 10.30700/jst.v11i2.1122.

[7] A. F. Subarkah, R. Kusumawati, and M. Imamudin, "Comparison of Different Classification Techniques to Predict Student Graduation," *MATICS: Jurnal Ilmu Komputer dan Teknologi Informasi (Journal of Computer Science and Information Technology)*, vol. 15, no. 2, pp. 96–101, Nov. 2023, Accessed: Oct. 28, 2025. [Online]. Available: https://ejournal.uin-malang.ac.id/index.php/saintek/article/view/24095

[8] H. C. Husada and A. S. Paramita, "Analisis Sentimen Pada Maskapai Penerbangan di Platform Twitter Menggunakan Algoritma Support Vector Machine (SVM)," *Teknika*, vol. 10, no. 1, 2021, doi: 10.34148/teknika.v10i1.311.

[9] Louis M and Dian B, "Perbandingan Algoritma Knn, Decision Tree,Dan Random Forest Pada Data Imbalanced Class Untuk Klasifikasi Promosi," *Jurnal INSTEK (Informatika Sains dan Teknologi)*, vol. 7, no. 2, 2022, doi: 10.24252/instek.v7i2.31385.

[10] P. Dewi, P. Purwono, and S. Kurniawan Dwi, "Pemanfaatan Teknologi Machine Learning pada Klasifikasi Jenis Hipertensi Berdasarkan Fitur Pribadi," *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 11, no. 3, 2022, doi: 10.30591/smartcomp.v11i3.3721.