International Journal of Sciences: Basic and Applied Research (IJSBAR)

ISSN 2307-4531
(Print & Online)

http://gssrr.org/index.php?journal=JournalOfBasicAndApplied

Evaluation of Fund Usage and KJMU Potential Recipients Modeling using Classification Tree and EasyEnsemble

Vidya Ayuningtyas[a]*, Hari Wijayanto[b], I Made Sumertajaya[c]

[a,b,c]Department of Statistiscs, Faculty of Mathematics and Natural Sciences (FMIPA), IPB University, Bogor, 16680, Indonesia

[a]*Email: vidya_84@apps.ipb.ac.id*

[b]*Email: hari@apps.ipb.ac.id*

[c]*Email: imsjaya@apps.ipb.ac.id*

Abstract

(Kartu Jakarta Mahasiswa Unggul (KJMU) was one of Jakarta Provincial Government main program in education. This program aims to help students from poor family with excellent grades to continue higher education. This research was started by conducting survey to 354 recipient who entered college between 2016-2018 to see how they use KJMU funds. Even though they claim knowing how to manage their funds surprisingly education expenses come in fifth position out of 7 types of expenses and their incomes only covers 64.55% of their primary expenses. The main cause of these problems suspected because the candidates for KJMU recipients did not match the requirement. That is why this study continues by finding the right methods for classify the candidates. Since the recipient of KJMU is minority compares to majority people in Jakarta which born in 1997-2000 there is class imbalance issue in making classification model. If this issue not resolved well it will cause accuracy paradox where the prediction will tend toward majority class. This study compared CART (Classification Tree) and EasyEnsemble to find the most suitable model. Classification tree is known with its easy interpretation, high accuracy and fast but this method requires the balanced class that is why we add undersampling techniques into it. EasyEnsemble was designed for handling imbalanced class and it was combinations of UnderBagging and ADABoost. The results show that EasyEnsemble is the best method with the highest F1 score over 10, 50 dan 100 iteration modelling.

Keywords: KJMU; Survey; Classification Tree; EasyEnsemble.

---

* Corresponding author.

1. Introduction

Kartu Jakarta Mahasiswa Unggul (KJMU) which launched by Jakarta Provincial Government on June 20, 2016 was adapted from similar program called Conditional Cash Transfer (CCT). CCT proven to be effective at eliminating crisis effect in many develop countries [1]. This program aims to increase access and learning opportunities in State University (PTN) for students from poor family but has excellent grades. Recent research from [2] states that there is no conformity between organization and user where there is a lot of KJMU recipient did not match the requirements. Besides bureaucracy, technical issue is another reason for this problem. The field instrument which is the main determinant in detecting whether a student eligible or not as KJMU recipient can be easily manipulated by either student or the school. One way to detect the mismatches in the recipient criteria is to look at KJMU recipient fund usage. That is why this research starts with conducting survey among 354 of year 2016-2018 KJMU recipients using Stratified Random Sampling. Survey was held on May 2020 using online questionnaire to recipients which spread across 85 PTN around Indonesia. The expenditures will be based on the seventh components of Decent Living Needs (KHL). KHL was used by Provincial Government throughout Indonesia for counting the Provincial Minimum Wage each year. Then it will be analyzed using weighted average method as [3] used to calculate the cost of living for FMIPA IPB students. Grouping can help to identify respondents. The clustering method using the Two Step Cluster algorithm produces the best optimal cluster solution for data with mixed variables [4]. Further analysis for the characteristics of KJMU recipients from the survey data will begin by analyze using the Two Step Cluster method. The mismatches problem will be overcome by searching for the right method to classify the potential recipients of KJMU. Classification and Regression Tree (CART) is a non-parametric statistical method which also known as basic classifier in classification analysis. It has the advantages of ease in graphical interpretation, relatively high accuracy, fast computation and can be used in all kind of data (numeric and categoric) [5]. Since the response variable in this research is categoric, the Classification Tree method was used. One of the main requirements in classification using this method is that the amount of data in each class on the response variable must be evenly distributed. The population of KJMU recipients is 4542 people when compared to the population aged 19-22 years in DKI Jakarta Province which is 636.877 [6] will cause problems in making models with the Classification Tree. The model formed will have a high accuracy but low recall value because the prediction of the model will always lead to the major class. While the main concern in this research is minor class classification. Therefore imbalanced data is our main concern prior to modeling the data. The application of resampling method to CART can improve the result for minor class classification [7]. That is why the implementation of simple techniques in data collection namely undersampling techniques will be implemented prior to modeling using Classification Tree According to [8] one of the best methods for imbalanced data is to apply the ensemble method, a method that trains several basic classifiers from training data to make predictions and combine the results of their predictions to get the final prediction. The ensemble method divided into 3: boosting-based, bagging-based and hybrid-based. The hybrid-based is the combination of the boosting and bagging-based methods. [9] stated that EasyEnsemble is a hybrid method that has good performance and fast computation time. This is because this method uses a combination of UnderBagging for sampling the data and ADABoost as its basic classifier. This research will compare the basic classifier Classification Tree and EasyEnsemble method in making a classification model for KJMU potential recipients. The data used for modeling are data of DKI Jakarta

residents born in year 1997-2000 from the integrated data for the management program for the poor and needy. This data belongs to DKI Jakarta Provincial Social Service. The best model will be selected based on F1 score which is a harmonic mean between precision and recall.

2. Materials and Method

2.1 Data

The data used in this study are data related to individual data, family, residences and the use of funds from KJMU recipients. The data were obtained through 2 types of sources namely:

1. Primary Data

Primary data was collected through an online survey in May 2020 to 354 KJMU recipients from 2016 to 2018 using the Stratified Random Sampling method. The division of strata is carried out based on the scientific clump and the location of the PTN, see table 1.

Table 1: Allocation of sample survey respondents

Scientific clump	PTN location			Total
	Jabodetabek	Java Island (Non Jabodetabek)	Outside Java Island	
Mathematics and natural sciences	16	3	2	21
Plant science	8	3	2	13
Animal science	4	3	1	8
Medical and health science	18	3	2	23
Engineering science	27	8	5	40
Linguistics	9	4	1	14
Economics	38	10	2	50
Humaniora social science	36	11	4	51
Religion and philosophy	11	6	2	19
Arts, design and media	15	2	0	17
Education science	78	16	4	98
Total	260	69	25	354

The scientific clumps used refer to the division of the Ministry of Research, Technology and Higher Education's scientific clumps and the location of PTN refers to the Provincial Minimum Wage (UMP) in 2019. Determination of the number of samples in each stratum based on proportional allocation. Fund management patterns were explored through a questionnaire where the questions refer to the 7 components of the Decent Living Needs (KHL) according to [10], namely Food and Beverage, Clothing, Housing, Education, Health, Transportation, Recreation and Savings.

2. Secondary data

Secondary data that used in the formation of a classification model for prospective KJMU recipients were taken from:

a. 2016-2018 KJMU recipients data.

Source: UPT P4OP DKI Jakarta Provincial Education Office.

b. Integrated data on programs for handling the poor and needy.

Source: DKI Jakarta Provincial Social Service.

The two types of data above will be combined based on the ID number (NIK) to get a recipient class (minor class) and not a KJMU recipient (major class) as a response variable. Of the 8190 people born in 1997-2000 in the integrated data for the management of the poor and needy people used there were 734 KJMU recipients. Thus the response variables consisted of a major class of 7456 people (91,04 %) and a minor class of 734 people (8,96 %). The imbalanced ratio is 1 : 11,16. Meanwhile, the predictor variables used were adapted from the definition of micro-poverty according to the BKKBN, BPS and the Ministry of Social Affairs. The predictor variables used were categorical and divided into 5 main parts of measurement:

a. Individual data: gender (X1), number of household members (X2), school participation (X3), highest education level currently / ever attended (X4), highest diploma held (X5) and work status (X6).
b. Participation in social assistance programs: PBI Health insurance (X7) and PKH (X8).
c. Residence conditions: building tenure status (X9), land status (X10), widest floor type (X11), floor area (X12), widest wall type (X13), wall condition (X14), widest roof type (X15) and roof condition (X16).
d. Household resources: drinking water source (X17), how to get drinking water (X18), lighting source (X19), installed power (X20), fuel / main energy for cooking (X21), use of defecation facilities (X22), type of toilet (X23) and feces TPA (X24).
e. Ownership of family assets: gas cylinder (X25), refrigerator (X26), AC (X27), landline (X28), laptop (X29), motorbike (X30), car (X31), boat (X32), boat (X33) and immovable assets (X34).

2.2 Data Analysis Procedur

The stages of analysis carried out to answer the objectives in this study are the following:

1. Evaluating the fund management patterns of KJMU recipients.

Evaluation is carried out through exploration of primary data obtained from surveys conducted on KJMU recipients. Descriptive of the recipient characteristics is seen from the income, expenditure and financial management using the weighted average method. The survey results were then further analyzed using the Two Step Cluster method to find the optimal cluster. Descriptive statistics then used to further analysis.

2. Creating a classification model for potential KJMU recipients.

The steps taken at this stage are the following:

a. Perform data preprocessing stages, namely cleaning data (checking data that contains a lot of missing values), calculating the Cramer's V value to select which predictor variables have strong associations with response variables to be included in models.
b. The formation of training data and testing data, namely 80% of the total data becomes training data and 20% of the total data becomes testing data using the 10-fold cross validation method.
c. Creating a classification model for KJMU recipients using the Classification Tree method on the training data and validating the model against testing data with 10 replications. Calculate the average value of accuracy, precision, recall, F1 score and analyze the results of modeling imbalance data using Classification Tree.
d. Add the undersampling technique prior to the formation of a classification tree as in point c then modeling with 10, 50 and 100 repetitions. Calculate the average value of accuracy, precision, recall and F1 score from the Classification Tree using undersampling techniques.
e. Creating a KJMU recipient classification model using the EasyEnsemble method on training data and validating the model against testing data with 10, 50 and 100 replications. Calculating the average value of accuracy, precision, recall and F1 score from the EasyEnsemble classification model.

3. Determine the best classification model

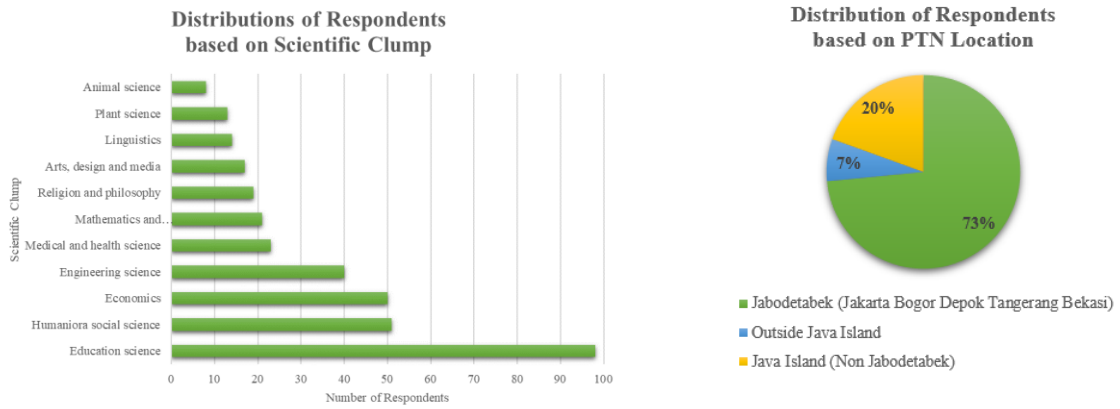
Identify the best classification model for KJMU recipients based on the highest F1 score of the two types of models. Interpret the level of importance of the predictor variables based on the results from the variable importance plot.

3. Results

3.1 KJMU Recipient Fund Management Pattern

3.1.1 Data Exploration

The data used to analyze the pattern of KJMU fund management is from a survey conducted on 354 KJMU recipients spread across Indonesia. From a total of 354 respondents, 33,33 % were male and 66,67 % female. Based on figure 1, 73 % of respondents studied at PTN located in Jabodetabek, 20% in Java Island (Non Jabodetabek) and 7 % outside Java Island. The top 3 favorite study programs for KJMU recipients are those included in the Education, Social Humanities and Economics clusters.



(a) (b)

Figure 1: Distribution of respondents based on scientific clump (a) and PTN location (b)

3.1.2 Income

The DKI Jakarta Provincial Government provides a fund of Rp. 9.000.000,00 for each semester to KJMU recipients. From this amount, after deducting the amount of Tuition Fee (UKT) in each PTN, the rest of fund can be used according to the provisions in [11]. On average, UKT paid by KJMU recipients is Rp. 3.216.571,33, thus the remaining Rp. 5.783.428,67 can be used as pocket money. Other sources of income that KJMU recipients may receive are pocket money from parents / siblings (57 %) and a side job (24 %). There are 4 income scenarios in each semester for KJMU recipients as in table 2. Based on this table, if only relying on KJMU funds, the recipients only have Rp. 5.783.428,67 per semester and will reach Rp. 10.857.091,10 if they had pocket money and side jobs.

Table 2: Income scenario

Scenario		Income
Pocket Money	Side Job	(Rupiah)
Yes	Yes	10,857,091.10
Yes	No	8,368,855.81
No	Yes	8,271,663.97
No	No	5,783,428.67

3.1.3 Expenditure

The indicators used in measuring KJMU recipient expenditure refer to KHL 7 components. Based on the survey results, the average expenditure per semester for each component is shown in table 3

Table 3: Expenditure based on KHL components

KHL	Details	
	Amount	Percentage
Food and beverage	3.728.813,56	28,25 %
Housing	2.962.500,00	22,44 %
Recreation and Savings	2.273.728,81	17,22 %
Clothing	1.786.723,16	13,53 %
Education	1.187.146,89	8,99 %
Transportation	1.081.355,93	8,19 %
Health	180.649,72	1,37 %
TOTAL	13.200.918,08	100,00 %

Based on table 3, it can be seen that even though KJMU is an education scholarship, expenditures for education are in the fifth place. Referring to [11], besides being used for UKT, the KJMU funds can also be used for personal support costs such as books, nutritious food, transportation and equipments. Therefore, the expenditures will be focused on the four main expenses of a student, namely food and drink, housing, educations and transportation. The main expenditure for KJMU recipient is Rp. 8.959.816,38, where 41,62 % of expenditures are spent on food and drink, 33,06 % on housing, 13,25 % on education and 12,07 % on transportation. If compared with the income scenario in table 2, KJMU recipients will be able to cover primary expenses only if they get pocket money and do side jobs.

3.1.4 Optimal Cluster Analysis with Two Step Cluster

A more detailed description of the fund management pattern of KJMU recipients is carried out through analysis using the Two Step Cluster method. The clustering is carried out based on the numerical variables of the remaining allowance for KJMU funds and categorical variables for main expenses. Data processing was carried out with the help of the SPSS version 21.

Table 4: Optimal number of clusters based on BIC

Number ofClusters	Bayesian Criterion (BIC)	Change in BIC	BIC Change Ratio	Distance Measurement Ratio
1	5013,989			
2	4620,609	-393,380	1,000	1,282
3	4343,375	-277,235	0,705	1,884
4	4259,508	-83,866	0,213	1,263
5	4221,197	-38,312	0,097	1,206
6	4212,547	-8,650	0,022	1,186
7	4226,405	13,858	-0,035	1,109
8	4252,161	25,756	-0,065	1,018

The maximum number of clusters is determined by looking at the BIC change ratio value that is closest to 0,04. Based on table 4, the maximum number of groups is 5 with the value of the ratio of change in BIC is 0,097. The

optimal number of cluster is obtained by finding the cluster with the maximum ratio of change in distance. Table 4 shows the ratio of the largest change in distance R (k1) = R (3) = 1,884 and R (k2) = R (2) = 1,282. The ratio of change in distance is R (k1) / R (k2) = 1,47. The value of the change in distance ratio is more than 1,15, thus the optimal number of cluster is 3.

3.1.5 Characteristics for Each Clusters

The distribution in each cluster group is as follows

Table 5: Optimal cluster membership distribution

Cluster	Amount	Percentage
1	136	38,4 %
2	120	33,9 %
3	98	27,7 %
Total	354	100,0 %

The following is an analysis of the clustering in each group based on cluster forming indicator.

1. The remaining KJMU fund pocket money

In cluster 1 the average remaining allowance for KJMU funds is Rp 6.180.937,06, cluster 2 is Rp 5.554.956,25 and cluster 3 is Rp 5.512.295,92. Cluster 1 is a group of KJMU recipients with the smallest average of UKT amount among other groups so that they have the largest income from the KJMU fund allowance.

2. Main expenditure

The 4 main types of expenses for KJMU recipients in each cluster is as follows.

Table 6: Main expenditures in each cluster

Main Expenditure	Cluster		
	1	2	3
Food and drink	3.282.352,94	4.560.000,00	3.330.612,24
Education	1.308.823,53	1.210.416,67	989.795,92
Transportation	1.530.656,93	596.296,30	655.102,04
Housing	4.277.777,78	2.618.181,82	3.065.217,39
Total	10.399.611,18	8.984.894,78	8.040.727,60

Based on table 6, it can be seen that cluster 1 has a much greater main expenditure than the other two groups. When compared with the income figures in point 1, it can be concluded that the greater the income then the greater the expenses. All groups have a major expense that far exceeds their income. The largest deficit was in cluster 1 of -Rp 4.219.214,12, group 2 of -Rp 3.429.938,53 and cluster 3 of -Rp 2.528.431,68. Group 1 as the group with the largest expenditure was represented by 3 categories of expenditure being in the highest position.

Further analysis will be carried out based on other research variables in the survey.

1. Allowances and side job

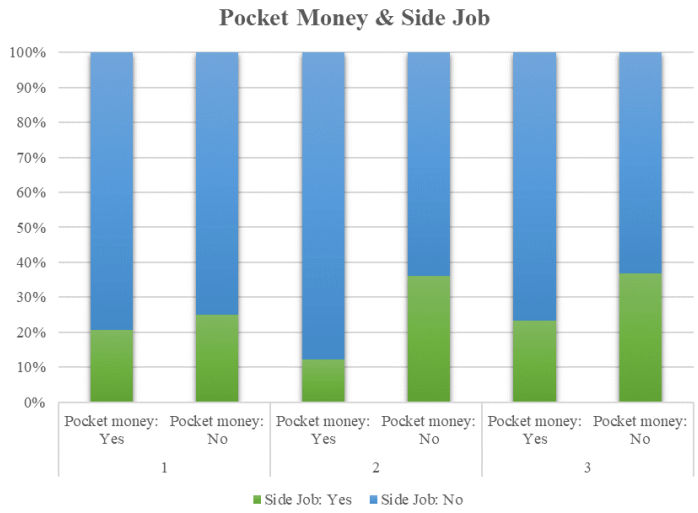


Figure 2: The proportion of KJMU recipients distribution

based on pocket money and a side job

Based on the survey results, 56,78 % of KJMU recipients have pocket money either from parents or relatives. If we look at figure 2, the characteristics of the majority of KJMU recipients in all groups are similar, have additional pocket money and do not have a side job. This characteristic is mostly found in cluster 2. KJMU recipients supposed to come from poor families so that generally parents are unable to send their children to higher education levels. This can be characterized by no allowance given to the child. The majority of KJMU recipients who do not have additional pocket money are in group 3. One of the characteristics of a student's independence is if they can meet their needs with their own resources. Independent KJMU recipients with the characteristic of not receiving additional pocket money but having a side job, mostly find in cluster 3.

2. Housing and ownership of private vehicle

In Figure 3, it can be seen that cluster 1 is filled by all KJMU recipients who do not have private vehicles and the majority live in their parents' houses. Cluster 2 is filled by the majority of recipients who live in a boarding house and do not have private vehicles. While cluster 3 is filled with recipients whose 98,98 % have motorcycles and the majority reside in their parents' houses.

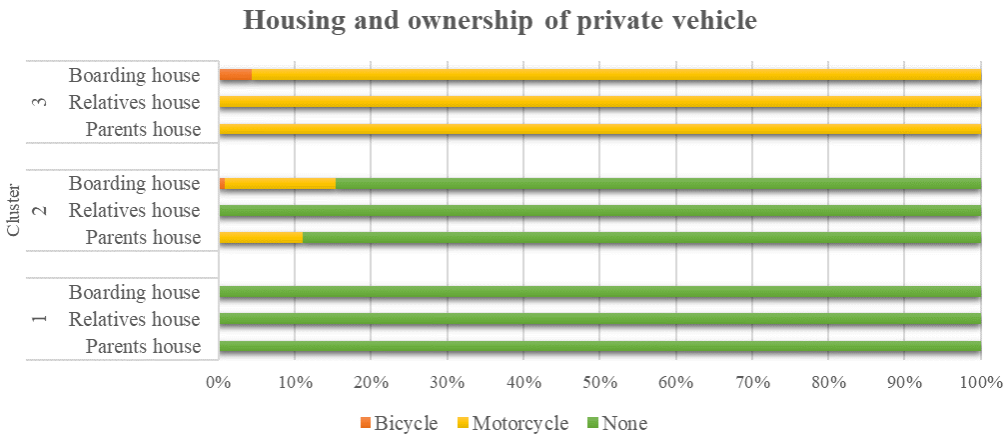


Figure 3: The proportion of KJMU recipients distribution based on

housing and ownership of private vehicle

3. Ownership of electronic devices

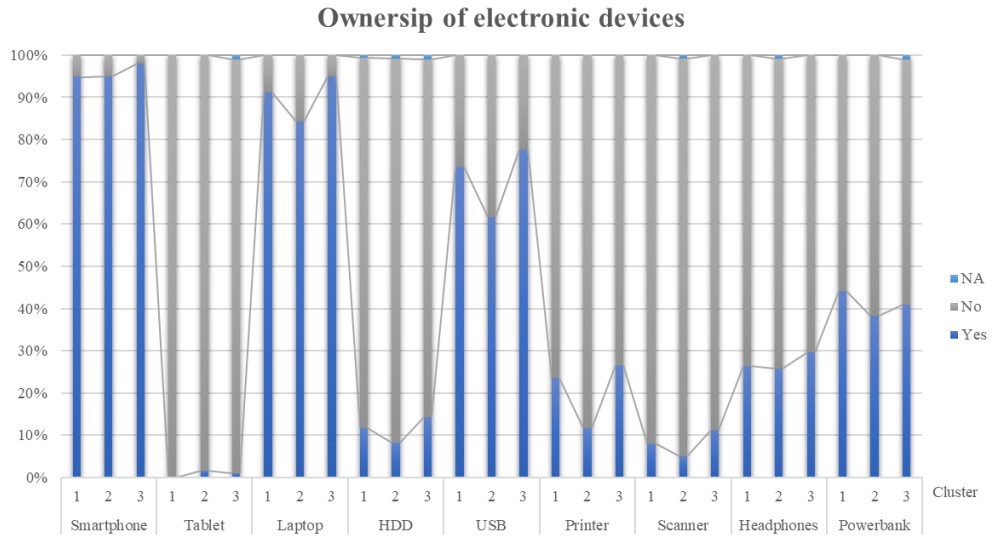


Figure 4: Proportion of ownership of electronic goods for KJMU recipients

KJMU recipients are millennial generation which characterized by high dependency on electronic devices to support their activities. Based on the ownership of 9 types of electronic devices that were asked in the survey (shown in figure 4) the same pattern was obtained. The KJMU recipients in cluster 3 owned more electronic devices than the recipients in other groups. Cluster 2 is a group containing KJMU recipients who have the smallest amounts of electronic devices. The top 3 electronic devices owned by KJMU recipients are smartphones, laptop and USB.

4. Percentage of expenditure per category

When KJMU recipients asked to describe their expenditures in terms of 3 types of expenditure (primary, secondary and tertiary) and savings. Results are obtained as shown in Figure 5. It can be seen that the three cluster have a similar pattern, the majority of KJMU recipients have primary expenditures in the range of 50% of the funds, 25% each for secondary expenses and savings and 0% for tertiary expenditures. Based on these results we can conclude that all KJMU recipients have a good understanding of good fund budgeting.

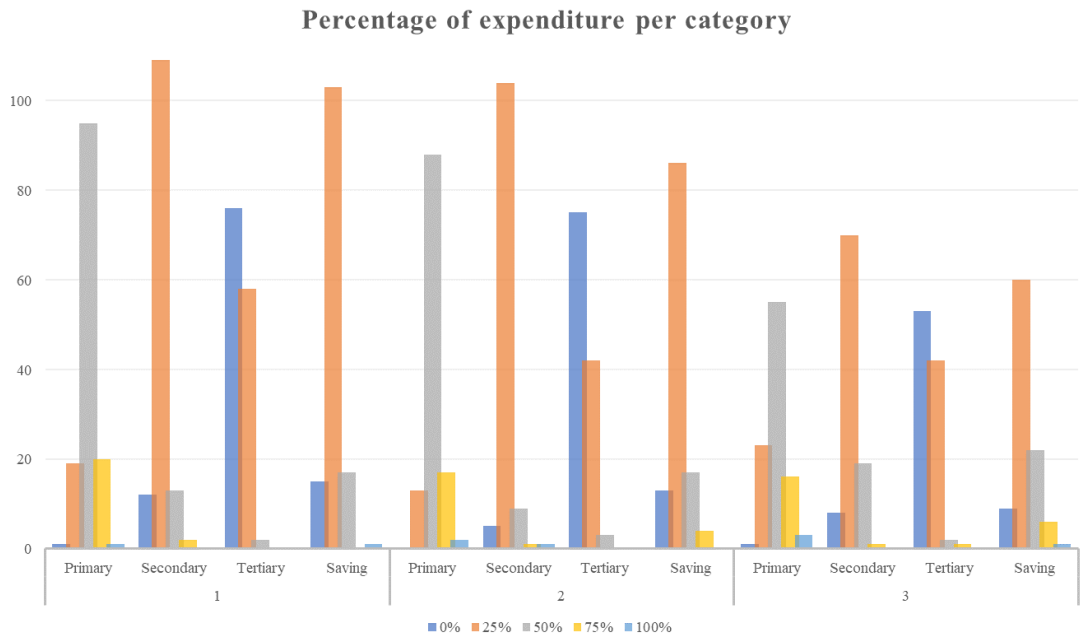


Figure 5: Expenditure per-category in each cluster

3.2 Classification Models for KJMU Recipient Candidates

3.2.1 Selection of Predictor Variables

The strength of the association between the 34 baseline predictor variables and response variables was measured using Cramer's V value.

Table 7: Strength of association between response and predictors variables

Association strength	Cramer's V range value	Predictor Variable
Very strong	0.301	X4
Strong	0.156	X29
Moderate	0.11-0.145	X1, X3
Weak	0.053-0.098	X5, X6, X7, X8, X11, X13, X17, X18, X20, X22, X26, X30

Interpretation of association strength based on [12]. 16 predictor variables that have associations from very

strong to weak with the response variables as shown in table 3 will be included in the modeling and the remaining 18 predictor variables that have no association with response variables or very weak associations are not included in the modeling.

3.2.2 Selection of the Best Classification Model

At the initial stage, the data used for modeling is divided into 2, 80% of the data into training dataset or data used to form the model and the remaining 20% used for testing dataset or data used for validating the model. When the Classification Tree as the basic classifier applied to research data it produces high accuracy 91.08% however if we investigated further the goodness values of the model are only 0.54% for recall, NA for precision and F1 score and the tree that formed is a stump (weak learner). This is called the accuracy paradox where the high accuracy value is due to all observations being classified by the model to major classes (not KJMU recipients). A better understanding of how classification tree works will be seen by applying the undersampling technique to data collection before modeling using the classification tree. In the undersampling technique, imbalanced data problems are solved by doing random sampling from major classes and select the same amount of data as many as the number of minor classes. One of the weaknesses of classification trees is that the trees that are formed depend on which training data that is used. To overcome this issue and prevent overfitting of the model, 10-fold cross validation method was applied to training and testing data collection techniques. The process of validating the model was carried out through modeling with 10, 50 and 100 repetitions as shown in table 8. Tree formation is strongly influenced by the complexity parameter (cp) value, the optimal tree size will be formed with a minimum cp value. The smallest value of minimum cp is obtained from the smallest x-val relative error (error in the cross validations process). For every iteration of the modeling process in table 8, the optimal cp value is used.

Table 8: Mean value of the measure of the goodness of the model with a 10-fold cross validation and 10, 50 and 100 replications of modeling

Measure of Goodness	Repetitions	Classification Models	
		Classification Tree + Undersampling	Easy Ensemble
Accuracy	10 50 100	75.48% 75.48% 75.48%	72.16% 72.26% 72.27%
Precision	10 50 100	20.21% 20.20% 20.20%	20.48% 20.55% 20.55%
Recall	10 50 100	58.72% 58.71% 58.72%	72.92% 72.95% 72.87%
F1 Score	10 50 100	30.06% 30.05% 30.04%	31.97% 32.05% 32.04%

The results of the classification tree modeling using undersampling technique show consistent results in the 3 types of modeling replications that were tried. Based on table 8, the accuracy are stable at 75,48 % indicating

the ability of the model is quite good in predicting a KJMU recipient, but on the precision and recall measures which are the main criteria for selecting the best model in this research EasyEnsemble is far superior. EasyEnsemble, which is a combination of 2 ensemble techniques (UnderBagging and ADABoost) is a method specifically designed to solve imbalanced data problems. Eventhough it has a lower average accuracy, EasyEnsemble excels at higher recall and precision values. The high recall value in the EasyEnsemble model indicates the model's ability to predict correctly the KJMU recipients from all KJMU recipients using this model is superior. In this research, the selection of the best model was based on the F1 score which is the harmonic value of precision and recall. Thus, the best classification model in determining KJMU recipient candidates is EasyEnsemble.

3.2.3 Characteristics of Variables Affecting the Best Classification Model

In the EasyEnsemble modeling, 16 predictor variables were used with the level of importance for each predictor variables is shown in figure 6. The 3 main variables that have the most important role in decision for a KJMU candidate recipient are the type of the widest wall (X13), how to get drinking water (X18) and the highest level of education that is currently / has been occupied (X4). While the variable that have the least effect in KJMU candidate recipients is gender (X1) with an importance level score close to 0.

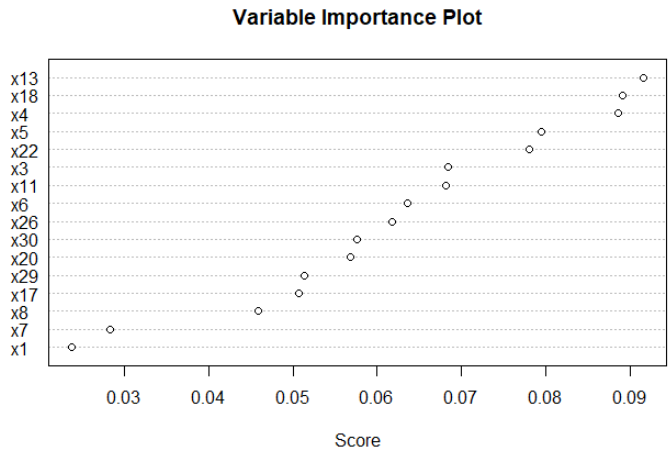


Figure 6: Level of importance for the predictor variables

4. Conclusion

Eventhough 57% of KJMU recipients get pocket money from parents or relatives and the majority of KJMU recipients already understand the theory of good financial management. Numerically, the expenses of KJMU recipients exceed the amount of income received. The income comes from the remaining KJMU funds after being deducted by the UKT on average only covering 64,5 % of the main expenses. They can only cover the main expenditure only if they have pocket money and doing side job. Further analysis for the detailed characteristics of the KJMU recipients is based on the 3 optimal cluster formed. The three cluster of KJMU recipients have the same financial pattern, if they only rely on pocket money from the remaining KJMU funds

then their primary needs cannot be covered. Cluster 3 is the most ideal KJMU recipient characterized by the smallest income and expenditure, majority do not have pocket money from parents or relatives but have a side job and the most have electronic devices as learning support. Even for the ideal KJMU recipient as in cluster 3, the KJMU fund does not cover their primary needs. There is a lack of funds by -Rp 2.528.431,68. The root of the problem from the large number of KJMU recipient participants that do not comply with the predetermined criteria is in the initial selection stage of KJMU recipient candidates. This can be overcome by making the best classification model for prospective KJMU recipients. Of the 2 types of classification models that were tried, namely the Classification Tree with the undersampling technique and EasyEnsemble, EasyEnsemble method which was specially designed for data handling with an unbalanced class is the best model in classifying KJMU recipients with a higher F1 score. The variables that have the most important role in the formation of the model are the widest wall type (X13), how to get drinking water (X18) and the highest level of education that is currently / has been occupied (X4).

5. Recommendations

Further research to find the optimal number of KJMU funds is needed.

Acknowledgement

This research is fully supported by Jakarta Provincial Government.

References

[1]. Rawlings LB and Rubio GM. “Evaluating the Impact of Conditional Cash Transfer Programs”. World Bank Policy Research Working Paper 3119, 2003.

[2]. Kusumajati. “Analisis Kebijakan Bantuan Biaya Peningkatan Mutu Pendidikan Bagi Mahasiswa Tidak Mampu”. Thesis, University of Indonesia, Indonesia, 2019.

[3]. Clinton E. “Survei Biaya Hidup Mahasiswa Berdasarkan Pengeluaran Konsumsi Makanan dan Non Makanan (Studi Kasus Mahasiswa Fakultas Matematika dan Ilmu Pengetahuan Alam)” Skripsi, IPB University, Indonesia, 2015.

[4]. Munthe AD. “Penggerombolan Desa/Kelurahan di Provinsi Nusa Tenggara Timur Menggunakan Algoritma Two Step Cluster (TSC) dan Algoritma K-Prototypes”. Thesis, IPB University, Indonesia, 2019.

[5]. Breiman L, Friedman JH and Olshen RA, Stone CJ. Classification and Regression Trees. New York : Chapman & Hall/CRC, 1984.

[6]. Dinas Kependudukan dan Pencatatan Sipil Provinsi DKI Jakarta. Buku Informasi Hasil Layanan Kependudukan 2018. Jakarta (ID): Disdukcapil, 2018.

[7]. Boonamnuay S, Kerdprasop N and Kerdprasop K. “Classification and Regression Tree with Resampling for Classifying Imbalanced Data”. International Journal of Machine Learning and Computing, vol 8, pp 336-340, 2018.

[8]. Liu XY and Zhou ZH. Imbalanced Learning: Foundations, Algorithms and Applications. First Edition.

John Wiley and Sons Inc, 2013.

[9]. Liu XY, Wu J, Zhou ZH. "Exploratory Under-Sampling for Class-Imbalance Learning". IEEE Transactions on Systems, Man and Cybernetics, vol 39, pp 539-550, 2009.

[10]. Kementerian Ketenagakerjaan Republik Indonesia. Peraturan Menteri Tenaga Kerja dan Transmigrasi Republik Indonesia Nomor 13 Tahun 2012 tentang Komponen dan Pelaksanaan Tahapan Pencapaian Kebutuhan Hidup Layak. Jakarta (ID): Kemenakertrans Republik Indonesia, 2012.

[11]. Pemerintah Provinsi DKI Jakarta. Peraturan Gubernur Provinsi DKI Jakarta Nomor 133 Tahun 2016 tentang Bantuan Biaya Peningkatan Mutu Pendidikan bagi Mahasiswa dari Keluarga Tidak Mampu. Jakarta (ID): Pemerintah Provinsi DKI Jakarta, 2016.

[12]. Akoglu H. "User's Guide to Correlation Coefficients". Turkish Journal of Emergency Medicine, vol 18, pp 91-93, 2018.