# Random forest algorithm for algorithm for prediction of high school science students acceptance snmptn based on students assesment report

U Pujianto [a,1,*], A R Taufani [b,2], J A Aziz [b,3]

[a] Departement of Electrical Engineering, Universitas Negeri Malang, Indonesia
[1] utomo.pujianto.ft@um.ac.id,*; [2] jordanabazyz@gmail.com;
* Corresponding Author

Received 25 February 2021; accepted 8 August 2021; published 13 September 2021

**ABSTRACT**

National Selection for State University (SNMPTN) is one of the selection lines for admission of new students in Indonesia to enter State Universities by invitation. Report card grades are one component of the assessment of admission of new students to enter state universities on this pathway. The difference in standards between universities in determining the admission of SNMPTN applicants, causing the need to predict based on several related factors. This research uses data mining techniques with Random forest algorithm. From the results of research that has been done, it was found that the Random Forest algorithm can be used to predict students who are accepted at SNMPTN based on report card grades, obtained from the results of the classification process with the student report card report survey dataset received by SNMPTN, This is indicated by the accuracy, precision, and recall values of 93%. Optimization of the random forest algorithm using the oversampling technique with the SMOTE method can improve the classifier's performance due to the imbalanced class problem

**KEYWORDS**
SNMPTN
Nilai Rapor
EDM
Random Forest
SMOTE



## 1. Introduction

National Selection for State University (SNMPTN) conducted by the Higher Education Entrance Examination Institute (LTMPT) under the auspices of the Ministry of Research and Higher Education is one form of student admission selection paths to enter state universities by invitation. aims to provide opportunities for high school / vocational / MA students who have the best achievements to pursue further education in state universities (PTN) and provide opportunities for PTN as the recipient of new students, to get prospective new students with qualified academic achievements.

General provisions of SNMPTN according to SNMPTN official website page www.snmptn.ac.id that SNMPTN is conducted on the basis of the results of a track record of academic achievement usingreport card grades. The report card grades used are semester one through semester five for applicants from high school / MA / SMK with three years study period, or for applicants from 4-year vocational school using report card grades semester one through semester seven. With detailed report cards on Mathematics, Indonesian, English, Chemistry, Physics, and Biology for students enrolling in the Department of Natural Sciences, then for the majors of Social Sciences are Mathematics, Indonesian, English, Sociology, Economics, and Geography. Then for students who register from Vocational School are Mathematics, Indonesian, English plus the value of expertise competency based on Vocational Theory and Practice taken during the learning period at Vocational School.

The criteria for selecting SNMPTN according to LTMPT are the academic track record of student report card grades, also the track record of applicant school performance, and the criteria set by each PTN. As for the specific provisions used by the Ten November Institute of Technology (ITS) contained in the official website of ITS *https://smits.its.ac.id* about ITS admission selection, on the SNMPTN pathway, ITS can only accept prospective new students coming from high schools / MA majoring in Natural Sciences and / or Vocational Schools with relevant majors in all study programs. Whereas the

Social Sciences majors can only choose the Business Management and Development Studies study programs [1]. Of course, the selection criteria for each state university are different, according to the Vice-Chancellor for Academic and Student Affairs for the 2012-2017 UGM period Prof. dr. Iwan. the criteria for evaluating students accepted through the SNMPTN at UGM are the first report card grades, then the ranking or accreditation of the school is coupled with achievements outside the student's academic grades and the track record of alumni who came from the student registrant's school while studying at UGM [2]. The same is also used by the University of Indonesia (UI) in determining the criteria for selecting new students, UI has the principle of instrument selection (not including factors of ethnicity, religion, RAS, class, gender, social status, and politics) and can be divided into two Student Acceptance Instruments New (PMB). The first for selection through non-test is student report card and school index. School index is a data of school quality which among other things is a function of the number of school graduates who enter UI, as well as its quality when selection enters UI and when becoming a UI student. Then the second for selection through the writtentest path using the selection script [3].

From the explanation above, it can be concluded that the report card scores determine the acceptance of SNMPTN, and also other considerations such as alumni and non-academic achievements are also considered whether or not to be accepted at SNMPTN, plus the selection stage according to LTMPT of student applicants selected at the first choice PTN based on order of selection of majors or study programs. So these four things are factors that are likely to be accepted at SNMPTN.

In previous studies relating to the problem of admission of new students on the SNMPTN pathway had been done by [4], where the selection of the best prospective students is based on criteria that have been determined with a decision support system using the ELECTRE method. This method is used to get the best alternative by eliminating alternatives that do not fit the criteria and can be applied to the decision making of the invitation through SNMPTN. The same thing was also done in making adecision support system using the Analytic Hierarchy Process (AHP) method, where this system can analyze the criteria in the selection of study programs, which can help prospective students in choosing study programs on the SNMPTN pathway [5]. Because this problem occurs in the world of Education, there is a domain of data mining methods that can be used to analyze data available at educational institutions or can be called EDM (Educational Data Mining). of course on this problem data mining techniques can be applied to do the classification. One of the data mining classification methods is Random Forest. Random Forest is a combined tree method derived from the development of the Classification and Regression Tree (CART) method, namely by applying the bootstrap aggregating method (Bagging) and Random Feature Selection)[6]. The Random Forest method has been applied to various problems in the fields of health, business, education, and others.

The Random Forest method was chosen because this method can give good results. This method uses voting from a number of Decision Trees that are built from existing training data, Random Forest is one of the classification algorithms with good accuracy. Random Forest is an ensemble method that consists of several decision trees as a classifier. The class produced from this classification process is taken from the most classes produced by the decision trees in the Random Forest. By voting on the available decision trees, the accuracy of Random Forest increases. This algorithm will group data or objects and can classify the data into several categories. Accuracy is a measure of the performance of the classification accuracy of a machine learning algorithm, but if the dataset used is imbalanced data then the measurement of accuracy is incorrect [7], imbalanced data is a condition where the distribution of data classes is imbalanced, the number of data classes (instances) is one less or more than the number of other data classes [8]. Classification of data with imbalanced classes is a major problem in the field of machine learning and data mining [9]. If working on imbalanced data, almost all classification algorithms will produce much higher accuracy for majority classes than minority classes [10]. Therefore, the researcher responds to the imbalanced class problem by doing resample onthe original data using the imbalanced class sampling technique that is oversampling with the Synthetic Minority Oversampling Technique (SMOTE) method.

As a solution to the SNMPTN problem previously explained, the data mining method, especially classification using the Random Forest algorithm, is expected to help predict the likelihood of students being accepted on the SNMPTN pathway and also to provide recommendations for majors to be accepted based on student report card grades and also supplemented by other supporting factors suchas alumni factors, academic and non-academic additional achievements, and priority of majors at the time of

SNMPTN registration. And also to determine the comparison of classifications using the Random Forest method before and after the implementation of the imbalanced class sampling technique, namely oversampling using the Synthetic Minority Oversampling Technique (SMOTE) method. The results of the best comparison are expected to be able to build a model that will later be used as a reference for the classification process based on the examples that have been studied.

## 2. Method

In this research, there are four stages, (1) Data Selection, (2) Preprocessing data, (3) Random Forest, (4) Evaluation. Show in Figure 1

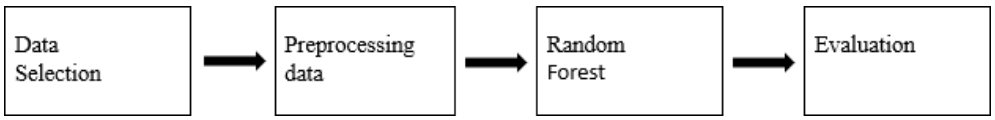


**Fig 1**. Research Stages

### 2.1. Dataset

The data used in this research process is student data that SNMPTN has received. Data is obtained from the campus halo website page www.halokampus.com and downloaded on 02/02/2019. Data contains student information received by PTN through SNMPTN. The data includes the academic track record of students used during SNMPTN registration such as, Student Data (name, origin, number of alumni, majors, types, school accreditation), UN scores, Report grades, and class rankings during semester 1-5 of the Natural Sciences and Non-majors -IPA, As well as priority choices and majors, PTN is accepted. Illustration of data and a list of attributes in the dataset are shown in Table 1 and Table 2

**Table 1.** Ilustration of Dataset

| Called Name | Major | Choice 1 | Choice 2 | Pass/Accpeted | MTK Semester 1 | … | Other Achiemenents |
|---|---|---|---|---|---|---|---|
| Nadya | IPA | Statiska | Mate-matika | Statiska | 86 – 90 | … | Piagam lomba MTQ |
| Salma | IPA | Mate-matika | Teknologi Pangan | Tekno-logi Pangan | 91 - 95 | … | Piagam Lomba PMR |

**Table 2.** List of Attributes in Datasets

| Number of Attributes | Explanation of Attributes | Data type |
|---|---|---|
| 18 | – Data Siswa (Nama siswa, Asal, Jurusan, Jenis,Akreditasi,Prestasi, Alumni Sekolah)<br>– Prestasi Tambahan dan Saran<br>– Prioritas Pilihan dan Jurusan,PTN diterima | String |
| 61 | – Nilai Rapor Semester 1-5 Jurusan IPA dan Non-IPA<br>– Jumlah Alumni<br>– Peringkat Class Semester 1-5 | Nominal /Categorikal |
| 1 | – Nilai UN | Numeric |

### 2.2. Preprocessing Data

Before carrying out the classification process in the dataset, the steps taken are to preprocessing the datafirst. Stages of data preprocessing that are done are (1) Data integratio n, (2) Data Cleaning (3) Data transformation of data, (4) Replace Missing Value, (5) Oversampling.

### 2.2.1. Data Integration

Data integration is the process of combining data from different sources into one new dataset. There are9 campus datasets in 2016 and 7 campus 2017 datasets combined into one new dataset with a total of 784 data instances with 80 attributes, the process of merging data is processed using the help of Microsoft Excel.

### 2.2.2. Data Cleaning

Furthermore, a data cleaning process aims to delete the data attributes that are not needed and make data selection or data attribute selection used. In this research, data selection is used to focus research on the prediction of the choice of majors received at (choice 1 or choice 2) on the SNMPTN based on the student report card grades, and other supporting factors such as the order of selection of majors or study programs (options 1 and 2 in one PTN), alumni factors (number of alumni received in the department), academic and non-academic additional achievements. The number after the dataset attribute selection process is 37 attributes and 290 data instances. List of atributes are show in Table 3.

**Table 3.** List of attributes after data selection processes

| Number of Attributes | Explanation of Attributes | Data type |
|---|---|---|
| 7 | – Data Siswa (Jurusan, jenis, Akreditasi Sekolah)<br>– Prestasi Tambahan<br>– Prioritas Pilihan (Pilihan 1 dan 2 dalam 1 PTN | String |
| 30 | – Jurusan diterima<br>– Alumni Sekolah, Nilai Rapor Semester 1-5 | Nominal /Categorikal |

### 2.2.3. Transformation Data

Data Transformation data is used to reduce many classes in the dataset using the concept hierarchy generation for nominal data, where attributes can be generalized to higher-level concepts. At this stage, data transformation is also used to convert all attributes to nominal/categorical.

At this stage the grouping focuses on the priority attributes of majors/study programs in the SNMPTN by applying the concept hierarchy generation to nominal data that refers to the knowledge clusters set by the Ministry of Research and Technology published on the *simlitabmas.ristekdikti.go.id* site in the attachment letter A about the knowledge cluster, obtained 7 nominal / categorical data from the results of transformation, namely Mathematics and Natural Sciences, Engineering, Health Sciences,Animal Sciences, Plant Sciences, Economics, and Social Sciences. Examples of transformation results are shown in Table IV. Because the research aims to predict the choice of majors accepted at (choice 1or choice 2) in SNMPTN, the attributes of the choice of majors accepted are changed based on the conditions accepted in option 1 or choice 2. The transformation results are shown in Table V. Data transformation was also carried out on additional achievement attributes. It aims to group students whoregister SNMPTN with or without attaching additional academic or non-academic achievements. The results of the transformation are shown in Table 4 Table 5 and Table 6.

**Table 4.** Transformation data in 'prioritas pilihan' attributes

| Original data | Result of transformation | Data type |
|---|---|---|
| Kebidanan | Ilmu Kesehatan | |
| Teknik Elektro | Ilmu Teknik | Nominal /Categorical |
| Statistika | MIPA | |
| Teknologi pertanian | Ilmu Tanaman | |

**Table 5.** Transformation data in 'Pilihan diterima' attributes

| Pilihan 1 | Pilihan 2 | Pilihan diterima | Result of transformation | Data type |
|---|---|---|---|---|
| MIPA | Ilmu Kesehatan | MIPA | Pilihan 1 | |
| MIPA | Ilmu Teknik | Ilmu Teknik | Pilihan 2 | Nominal /Categorical |
| Ilmu Teknik | MIPA | MIPA | Pilihan 2 | |
| Ilmu Teknik | Ilmu Tanaman | Ilmu Teknik | Pilihan 1 | |

**Table 6.** Transformation data in 'prestasi tambahan' attributes

| Original data | Result of transformation | Tipe Data |
|---|---|---|
| Semifinalis NLC 2015, | Ada | |
| Olimpiade Teknik Kimia ITS 100 Nasional, | Ada | Nominal /Categorical |
| (No data) | Tidak Ada | |
| - | Tidak Ada | |

### 2.2.4. Replace Missing Value

This process purpose to fill the data with empty value with relevant input. Blank data on the report cardgrades semester 1-5 will be replaced according to the distribution of values. Because the data is nominal type, the data will be replaced by the data that most often appears (Mode) from the empty data attribute itself. The results are show in Table 7

**Table 7.** Example of replace missing valur process in 'NIlai rapor' attributes

| mtk 1 | mtk 2 | mtk 3 | mtk 4 | mtk 5 |
|---|---|---|---|---|
| 86-90 | 81-85 | 81-85 | 81-85 | 81-85 |
| 86-90 | 81-85 | 76-80 | 86-90 | 86-90 |
| 81-85 | 91-95 | 76-80 | 91-95 | 91-95 |
| 81-85 | 71-75 | 91-95 | 71-75 | 71-75 |

### 2.2.5. Oversampling

Oversampling is the process of making or duplicating certain minority data so that the distribution of class members in the dataset becomes balanced. In this study, oversampling was performed using the Synthetic Minority Oversampling Technique (SMOTE) algorithm.

SMOTE is an oversampling method that uses a data synthesis approach or replication data from minority data. SMOTE works by finding k nearest neighbors for each data in the minority class, then synthetic data will be made as much as the percentage desired between minority data and k nearest neighbors that have been determined.

### 2.3. Random Forest Algorithm

The Random Forest method is the development of the Classification and Regression Tree (CART) method, namely by applying the bootstrap aggregating method (Bagging) and the Random Feature Selection [6].

Bootstrap aggregating (Bagging) is a technique that can be used to form a bootstrap sample whereeach decision tree is constructed using bootstrap samples from attribute candidate data to be shared at each node derived from a set of random attributes from the selected data result and stored [11]. Bagging and Boosting are relatively new ensemble methods but have become popular [12]. RandomForest goes through a randomization process that is done not only on sample data but also on taking independent variables so that the raised classification trees will have different sizes and shapes. Random Forest is a classification consisting of several decision trees that are constructed using random vectors [11].

Random Forest is the development of a decision tree using several decision trees where each decision tree has been trained using individual samples and each attribute is broken down in a tree chosen between a subset attribute that is random and in the classification process, the individual is based on the votes of the most votes in a collection tree population [13] .when selection determines overall classification, bad trees will make predictions that are good predictors that will appear in response [14]. The Random Forest operator produces a set of random trees, the class generated from the classification process is chosen from the most classes (mode) generated by the random tree that exists. Many trees are grown in the Random Forest method so that a forest is formed to be analyzed [6].

There are several phases in the formation of a classification model with random forests. these phasesare:

- Boostrap phase:

  built n training data by taking n samples from the the tarining data.

- Random feature selection phase:

  Create a new dataset from a random sample of attributes

- The phase of decision building

  Built a decision tree from the training data that has been made before

- Repeat steps 1 to 3 k times

### 2.4. Evaluation

After the algorithm is applied, performance testing is done using the cross-validation method with 10 fold settings. Next, the results of testing using cross-validation will be presented in the form of a confusion matrix. Confusion-matrix is a media that can be used to evaluate and present classification results. The results of confusion matrix are show in Table 8

**Table 8.** Confusion Matrix

| ***Correct Classification*** | **Result of Classification** | |
|---|---|---|
| | + | - |
| + | TP | FN |
| - | FP | TN |

True Positive (TP) Is data that has "True" value both in the classification results or in the original label class. While True Negative (TN) is a data that has a value of "false" both in the classification results and in the original label class. False Negative (FN) is data that is actually labeled as "True" but in Classification, Results is "False" and the last is False Positive (FP), False Positive (FP) is data that onthe label is actually "false" but on the classification results are "true". To evaluate the algorithm using aconfusion matrix can be done in 3 kinds of ways, namely (1) Accuracy, (2) Precision, (3) True PositiveRate (Recall).

#### 2.4.1. Accuracy

Test the accuracy of the classification results displayed by an algorithm. Accuracy can be calculated using equations 1:

$$Accuracy = \frac{TP+TN}{TP+FP+FN+TN} \text{ x } 100\% \quad (1)$$

#### 2.4.2. Precision

Testing on the algorithm in determining the true classification that matches the actual classification. The Precision of Precision can be calculated by the equation 2:

$$Precision = \frac{TP}{TP+FP} \quad (2)$$

### 2.4.3. True Positif Rate (Recall)

Calculate the proportion ratio classification (TP) on all data that are positive on the actual label. TPRate can be calculated using the equation 3 :

$$TP\ Rate = \frac{TP}{TP+FN} \quad (3)$$

## 3. Results and Discussion

### 3.1. Evaluation of Classification

Following are the results of research conducted using the Random Forest algorithm in the student survey dataset received SNMPTN obtained from *www. halokampus.com*

### 3.1.1. Classification Using Random Forest Algorithm without oversampling

Random Forest Algorithm test results on a dataset of a total of 290 data, with a test method in the formof 10 Folds Cross-Validation is shown using the confusion matrix, as presented in Table 9.

**Table 9.** Confusion matrix random forest algorithm without oversampling

| Class | Prediction class | |
|---|---|---|
| | **Choice 1** | **Choice 2** |
| Choice 1 | 240 | 2 |
| Choice 2 | 48 | 0 |

From the results of the confusion matrix shown by Table 1. Then the calculation of its accuracy

$$\text{Accuracy} = \frac{240+0}{240+2+48} \text{ x } 100\%$$

$$\text{Accuracy} = \frac{401}{418} 100\% = 83\%$$

Obtained an accuracy of 83%, The results of Precision and Recall of each class are shown in Table10.

Results of Precision and recall every class random forest algorithm without oversampling

**Table 10.** Results of Precision and recall every class random forest algorithm without

| Class | Accuracy | Precision | *Recall* |
|---|---|---|---|
| Choice 1 | | 83% | 99% |
| | 83% | | |
| Choice 2 | | 0% | 0% |

### 3.1.2. Classification Using Random Forest Algorith with Oversampling

Random Forest Algorithm test results on a dataset of 484 total data, with a test method in the form of 10 Folds Cross-Validation shown using the confusion matrix in Table 11.

**Table 11.** Confusion matrix random forest algorithm with oversampling

| Class | Prediction class | |
|---|---|---|
| | Choice 1 | Choice 2 |
| **Choice 1** | 235 | 7 |
| **Choice 2** | 25 | 216 |

From the results of the confusion matrix shown in Table 11. Then the calculation of accuracy isas follows:

$$\text{Accuracy} = \frac{240+0}{240+2+48} \text{ x } 100\%$$

$$\text{Accuracy} = \frac{401}{418} 100\% = 83\%$$

It obtained an accuracy of 93%. The results of Precision and Recall of each class are shown in Table12.

Table 12. Results of Precision and recall every class random forest algorithm with oversampling

**Table 12.** Results of Precision and recall every class random forest algorithm with

| Class | Accuracy | Precision | *Recall* |
|---|---|---|---|
| **Choice 1** | 93% | 90% | 97% |
| **Choice 2** | | 96% | 89% |

### 3.1.3. Comparison of All Evaluation Results

All of the Random-Forest algorithm evaluation results in this dataset are shown in Table 13. The results of Precision and Recall displayed are the results of the average weight or average Precision, and Recall of the whole class.

**Table 13.** Random Forest Evaluation Results

| Oversampling | Accuracy | Precision | Recall |
|---|---|---|---|
| No | 83% | 69,5% | 83% |
| **Yes** | **93%** | **93,5%** | **93%** |

## 3.2. Discussion

After testing, it can be concluded that the classification experiment using scenario 2 has the highest accuracy value compared to the first scenario accompanied by unequal distribution of data between classes, thus affecting the classifier to classify data, especially in minority classes. The accuracy obtained shows that the Random Forest method is good in classifying Option 1 class, but it cannot classify Option 2 class properly. This indicates the result of biased accuracy or called the accuracy paradox. If this happens, it is better to use other measures of classification accuracy other than accuracy such as Precision and Recall.

As a result of the imbalance in the prediction, the researcher will handle the imbalanced classby using the imbalanced class sampling technique with the SMOTE method. From the results obtained, Precision and Recall owned by the scenario without oversampling show relatively low results compared to Recall results on scenarios that use oversampling. the results of testing the level ofaccuracy, by 83%, 69.5% percent precision, and 83% Recall. While the results of testing the dataset that has gone through the stages of preprocessing and classification using the Random Forest algorithm with oversampling obtained the results of each test value of accuracy, Precision, and Recall is 93%. Based on the above results it can be said that the classification of majors received on this dataset by using the Random Forest algorithm

without oversampling shows unfavorable results but by doing oversampling from the test results obtained accuracy, Precision and Recall are better than without oversampling

## 4. Conclusion

From the research results that have been done, it can be concluded that the Random Forest method can be used to predict the choice of majors accepted at SNMPTN based on the report card grades. It was obtained from the results of classifying majors received at the National Selection for State University (SNMPTN) with a survey dataset of student report scores received by SNMPTN coupled with other supporting factors such as prioritizing majors, school accreditation, alumni factors, and additional academic and non-academic achievements. This is indicated by the results of an average accuracy, precision, and Recall of 93%. The use of oversampling with the SMOTE method in the minority class is very influential on the success of this classification process due to the imbalanced classproblem. Further research can be developed by expanding the dataset as added by other factors considered acceptable in the National Selection for State University (SNMPTN) or other preprocessingprocesses, so We hoped it would improve accuracy, precision and higher Recall

## References

[1] "Seleksi Kemitraan dan Mandiri Program Sarjana - ITS Admisi." https://www.its.ac.id/admission/en/undergraduate/seleksi-kemitraan-dan-mandiri/ (accessed Dec. 08, 2020).

[2] "Gunakan 7 Kriteria Penilaian, UGM Terima 3.316 Mahasiswa Baru Jalur SNMPTN | Universitas Gadjah Mada." https://www.ugm.ac.id/id/berita/9000-gunakan-7-kriteria-penilaian-ugm-terima-3-316-mahasiswa-baru-jalur-snmptn (accessed Dec. 08, 2020).

[3] "International Online Application for Undergraduate and Postgraduate Admissions - International Admission - Universitas Indonesia." https://admission.ui.ac.id/ (accessed Dec. 08, 2020).

[4] F. Setiawan, F. Indriani, and M. Aziz, "IMPLEMENTASI METODE ELECTRE PADA SISTEM PENDUKUNG KEPUTUSAN SNMPTN JALUR UNDANGAN," *KLIK - Kumpul. J. ILMU Komput.*, vol. 2, no. 2, pp. 197–120, Apr. 2016, doi: 10.20527/KLIK.V2I2.29.

[5] S. E. Putri, R. Sovia, and A. F. Hadi, "Implementasi Sistem Pendukung Keputusan Untuk Pemilihan Jurusan Pada Jalur SNMPTN Dengan Metode Analitycal Hierarchy Process(Studi Kasus: SMAN 3 Batusangkar)," *Semin. Nas. Teknol. Komput. Sains*, vol. 1, no. 1, Feb. 2019, Accessed: Dec. 08, 2020. [Online]. Available: https://prosiding.seminar-id.com/index.php/sainteks/article/view/231.

[6] L. Breiman, "Random Forests," *Mach. Learn. 2001 451*, vol. 45, no. 1, pp. 5–32, Oct. 2001, doi: 10.1023/A:1010933404324.

[7] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, "SMOTE: Synthetic Minority Over-sampling Technique," *J. Artif. Intell. Res.*, vol. 16, pp. 321–357, Jun. 2002, doi: 10.1613/JAIR.953.

[8] A. Ali, S. Shamsuddin, and A. Ralescu, "Classification with class imbalance problem: A review," *undefined*, 2015.

[9] R. Kothandan, "Handling class imbalance problem in miRNA dataset associated with cancer," *undefined*, vol. 11, no. 1, pp. 6–10, Jan. 2015, doi: 10.6026/97320630011006.

[10] Q. Gu, X. M. Wang, Z. Wu, B. Ning, and C. S. Xin, "An improved SMOTE algorithm based on genetic algorithm for imbalanced data classification," *J. Digit. Inf. Manag.*, vol. 14, no. 2, pp. 92–103, 2016.

[11] A. Byna and S. SARI MULIA Banjarmasin, "ANALISIS PERBANDINGAN ALGORITMA C.45, RANDOM FOREST DENGAN CHAID DECISION TREE UNTUK KLASIFIKASI TINGKAT KECEMASAN IBU HAMIL," *SEMNASTEKNOMEDIA ONLINE*, vol. 5, no. 1, pp. 2-1–103, Feb. 2017, Accessed: Dec. 08, 2020. [Online]. Available: https://ojs.amikom.ac.id/index.php/semnasteknomedia/article/view/1771.

[12] B. Widjanarko Otok, J. Statistika, F. Matematika dan Ilmu Pengetahuan Alam, and I. Teknologi Sepuluh November Surabaya Alamat, "RANDOM FOREST DAN MULTIVARIATE ADAPTIVE REGRESSION SPLINE (MARS) BINARY RESPONSE UNTUK KLASIFIKASI PENDERITA HIV/AIDS DI SURABAYA," *J. Stat. Univ. Muhammadiyah Semarang*, vol. 3, no. 1, 2015, doi: 10.26714/JSUNIMUS.3.1.2015.

[13] I. M. Budi Adnyana, "Prediksi Lama Studi Mahasiswa Dengan Metode Random Forest (Studi Kasus : Stikom Bali)," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 8, no. 3, pp. 201–208, 2016, doi:

10.22303/csrid.8.3.2016.201-208.

[14] Y. Sulistyo Nugroho and dan Nova Emiliyawati, "Sistem Klasifikasi Variabel Tingkat Penerimaan Konsumen Terhadap Mobil Menggunakan Metode Random Forest," *J. Tek. Elektro*, vol. 9, no. 1, pp. 24–29, 2017, doi: 10.15294/JTE.V9I1.10452.