JITK (JURNAL ILMU PENGETAHUAN DAN TEKNOLOGI KOMPUTER)

VOL. 6. NO. 1 AUGUST 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i1.1405

CLASSIFICATION OF THE PROSPECTS FOR CITY TREES LIFE EXPECTANCY USING NAIVE BAYES METHOD

Muhammad Rifqi Firdaus[1*]; Abdul Latif[2]; Ipin Sugiyarto[3]; Windu Gata[4]

Computer Science Study Program[1]
STMIK Nusa Mandiri
Nusamandiri.ac.id
14002384@nusamandiri.ac.id[1*]; 14002369@nusamandiri.ac.id[2]; ipin.isy@nusamandiri.ac.id[3]
windu@nusamandiri.ac.id[4]

(*) Corresponding Author

Abstract— Besides the city is a large and extensive residential area. as a center for the activities of its citizens, both from economic, cultural, and development activities. Development in the city leads to the physical development of the city with the many facilities and infrastructure in the city, making activities in the city cause some pollution problems. To overcome this problem, the government often creates green open space in the middle of the city. Planting shade trees will help to balance the problem of pollution due to development. Trees can reduce temperatures, in addition to absorbing air and climate pollution. trees can help save energy. Naive Bayes is a classification with probability and statistical methods, namely predicting future opportunities based on experience based on the assumption of simplification that attribute values are conditionally free if given output value. Data processing with Naive Bayes produces a Precision value of 0.840%, a recall value of 0.848%, and an AUC of 0.873%. These results indicate that the results are included in the excellent category.

Keywords: City Trees, Classification, Naive Bayes

Abstrak— *Kota merupakan kawasan permukiman penduduk yang besar dan luas sebagai pusat kegiatan warganya, baik dari kegiatan ekonomi, budaya maupun pembangunan. Pembangunan di kota mengarah pada perkembangan fisik kota dengan banyaknya sarana dan prasarana yang ada di kota, menjadikan aktivitas di kota menimbulkan beberapa masalah polusi. Untuk mengatasi masalah tersebut, pemerintah kerap membuat ruang terbuka hijau di tengah-tengah kota. Dengan ditanami pohon-pohon rindang, akan membantu menyeimbangkan masalah polusi akibat pembangunan. Pohon mampu menurunkan suhu, selain menyerap polusi udara dan iklim. Selain itu, pohon bisa membantu menghemat energi. Naive Bayes merupakan metode gabungan probabilitas dan statistik untuk pengklasifikasian, yang digunakan untuk memprediksi peluang di masa depan berdasarkan pengalaman dimasa sebelumnya. Penghitungan tersebut, berdasarkan pada asumsi penyederhanaan yang menyatakan bahwa nilai atribut akan saling bebas jika diberikan nilai output. Pengolahan data dengan Naive Bayes menghasilkan nilai Precision 0,840 %, nilai recall sebesar 0,848 %, dan AUC sebesar 0,873%. Hasil tersebut menunjukkan bahwa hasil tersebut masuk dalam kategori baik sekali.*

***Kata Kunci**: pohon kota, klasifiksi, naive bayes.*

INTRODUCTION

The city is a large and extensive residential area. as a center for the activities of its citizens, both from economic, cultural, and development activities. Development in the city leads to the physical development of the city with the many facilities and infrastructure in the city, making activities in the city cause several pollution problems[1]. To overcome this problem, the government often creates green open space in the middle of the city. Planting shade trees will help to balance the problem of pollution due to development. Trees can reduce temperatures, in addition to absorbing air and climate pollution.A, trees can help save energy [2].

Like humans, trees also have an age limit for life. Some tree species even have the potential to be developed. The condition of trees in urban areas must also be in good condition so that their functions can truly be felt. Considering the importance of trees in urban areas, tree health must be considered to prevent accidents caused by trees along the road. Maintenance of trees in a green open space will also affect life in the city. More and more untreated trees will reduce green open space and its facilities. The reduced green space results in an increase in local temperatures in the city. The loss of growth period faced by trees in the young age class becomes very prone to damage [3][4][5]. Replanting tree seedlings as a substitute must be made a selective selection which is also calculated from the expected level of life of the tree [6].

Referring to a series of research results, the initial growth of trees will be more with a tight spacing [7], with further consideration that will give better tree growth. Research is also carried out in determining the quality of wood. Where the results are, the quality of the wood can maximize the use of value and product life. Therefore the intensity of cutting down trees will be reduced and more environmentally friendly[8]. Research time has also been done periodically. As a result, each tree has a vulnerable harvest time between the vulnerable 35 years and once 45 years[9]. Research has also been conducted on tree species that are more effective at absorbing dust in the Green Open Space (RTH) area. The study was carried out along the median (the pathway that extends at the green space or can be referred to as the greenhouse entrance). The trees studied were a cape, jackfruit, pillar, and mahogany trees. As a result, higher dust absorption is carried out by jackfruit and cape trees[10]. The identification of the condition of tree damage has been done by research with forest health monitoring. The results were found 12 types of tree damage that occurred at the study site, namely Taman Abdul Raya Raya Rachman Park Lampung. Among them, Brum damage to branches, responses/gummosis, Konk, broken stems, growing stems, color-changing leaves, damaged shoots, termite nests, cancer, broken branches, and open wounds. The biggest types of damage are open wounds by 46% and the location of many tree damage occurs at the root and lower trunk by 29% [11].

Research on road user perceptions of the aesthetics and functionality of trees to meet green open space in urban areas. This research was conducted in two green lines in Malang, namely Ijen Green Lane and Veteran Green Lane. The result was that the Veterans green belt produced a high score of 59.97% from the Ijen green belt of 57.48%. Each vegetation in the two green lines has a good relationship value in terms of leaf color, flower color, the shape of the tree, tree texture, theme unity, accent, dominance, and balance[12]. Research using the Naive Bayes method was once conducted to predict the use of large amounts of household electricity usage. The data used were 60 data on household electricity usage. The results obtained an accuracy value of 78.33%, wherefrom the 60 data, there were 47 data on household electricity usage that were successfully classified correctly[13]. Financing approval decisions on Islamic cooperatives have been carried out the researched by comparing the algorithms used in the classification method namely Naive Bayes with an accuracy of 77.29% and a Decision Tree of 89.02%. The accuracy value generated by the Decision Tree is greater than Naive Bayes[14]. Research is also conducted to classify the opportunity (Opportunity) to buy or not to increase the achievement of targets as marketing. The method used to compare the two methods of classification C4.5 and Naive Bayes. The result is that the accuracy value of Naive Bayes is greater that is 75.67% and C4.5 has an accuracy of 66.67%. From these results, it can be seen that Naive Bayes is more effectively used to determine the classification of the possibility of buying or not from potential customers[15].

Research on the classification of tree life expectancies in urban green space has never been done. The author enters 6 attributes that will be used. The choice of the Naive Bayes algorithm was chosen because it produced a good accuracy value. Seeing from research using the same algorithm, it also produces a large accuracy value. Classification is the process of finding a model that can distinguish and describe classes in data[16]. The model was formed based on training data analysis. Derivative models can be represented in several forms[17]. Simply put, classification is the process of grouping data that is known by the number and name of the group.

Naive Bayes is a classification with probability and statistical methods, namely predicting future opportunities based on experience[18]. Naive Bayes is based on the simplified assumption that attribute values are conditionally free if output values are given. In this theory combined with Naive which means in attribute with the nature of the independent (independent) [19]. Naive Bayes can work quickly so it can handle attributes that are discrete and sustainable. Naive Bayes also has very good performance in handling data in real life and can make very good decisions[20].

Based on the description above, the Naive Bayes algorithm's data mining approach can determine the expected life span of trees in urban areas. The data obtained were sourced from data.sa.gov.au[21]. The data contains the prospect of tree life span in an area in South Australia (South Australia), with 6 categories including Longitude, Latitude, Tree Species, Trunk Circumference, Tree Health, Tree Structure, Tree Height, and Tree Quality life. The author will use the Data Mining

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

VOL. 6. NO. 1 AUGUST 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i1.1405

Classification technique by using the Naive Bayes algorithm to determine the existence of tree age in cities. It is expected that from this research, the life span of trees can be seen from the mature, immature, and veteran levels. Data to be processed using Rapidminer 9.3.0 Tools.

MATERIALS AND METHODS

The research method carried out to analyze the problem above uses the Data Mining technique with the Naive Bayes algorithm, as shown from Figure 1 below:

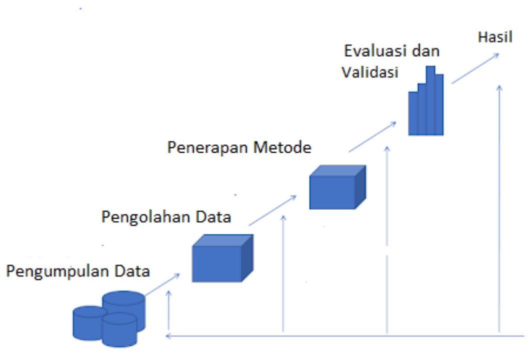


Figure 1. Flowchart

The stages of research conducted by the author in conducting this research are as follows:
Data Collection

This research data collection obtained from the South Australian government website about the expected life span of trees in urban areas. The number of datasets used was 855.

Data Processing

Preliminary data obtained in the form of tree life expectancy data in the South Australia region. After processing the data again, the attributes used in testing include Tree Species, Trunk Circumference, Tree Health, Tree Structure, Tree Height and Tree Quality life with the label attribute used is Tree Quality Life.

Application of the Model

The proposed model for the classification of tree life expectancy in this study is the Naive Bayes algorithm using RapidMiner 9.0 software for the analysis and testing of models.

Evaluation and Validation

Evaluation is used to observe and analyze the work of Naive Bayes on RapidMiner. Validation is done to measure the predicted results.

RESULTS AND DISCUSSION

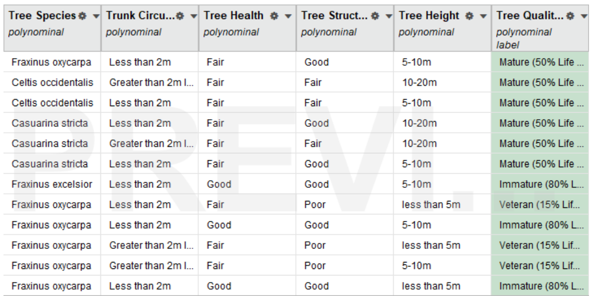
In this discussion the data collection stage is carried out first, then it will be analyzed to obtaining data that is truly appropriate, followed by data processing with classification data mining techniques with Naive Bayes using Tools RapidMiner 9.0 to get the value of accuracy and data modeling.

At the data collection stage, there is a problem with the pattern of tree life expectancy. The data used comes from the Australian government website with a total dataset of 855, including parameters Tree Species, Trunk Circumference, Tree Health, Tree Structure, Tree Height, and Tree Quality life.

Table 1. Data attributes and information

Nama Attribute	Information
Trunk Circumference	*(>2 m, 2m, <2m)*
Tree Health	*(Good, Fair, Poor)*
Tree Structure	*(Good, Fair, Poor)*
Tree Height	*(>5 m, 5-10 m, 10-20 m)*
Tree Quality Life	*(Mature 50% life, Immature 80%, Veteran 10%)*

Then the data is tested using the RapidMiner 9.0 Tools.



Tree Species polynominal	Trunk Circu... polynominal	Tree Health polynominal	Tree Struct... polynominal	Tree Height polynominal	Tree Qualit... polynominal label
Fraxinus oxycarpa	Less than 2m	Fair	Good	5-10m	Mature (50% Life ...
Celtis occidentalis	Greater than 2m I...	Fair	Fair	10-20m	Mature (50% Life ...
Celtis occidentalis	Less than 2m	Fair	Fair	5-10m	Mature (50% Life ...
Casuarina stricta	Less than 2m	Fair	Good	10-20m	Mature (50% Life ...
Casuarina stricta	Greater than 2m I...	Fair	Fair	10-20m	Mature (50% Life ...
Casuarina stricta	Less than 2m	Fair	Good	5-10m	Mature (50% Life ...
Fraxinus excelsior	Less than 2m	Good	Good	5-10m	Immature (80% L...
Fraxinus oxycarpa	Less than 2m	Fair	Poor	less than 5m	Veteran (15% Lif...
Fraxinus oxycarpa	Less than 2m	Good	Good	5-10m	Immature (80% L...
Fraxinus oxycarpa	Greater than 2m I...	Fair	Poor	less than 5m	Veteran (15% Lif...
Fraxinus oxycarpa	Greater than 2m I...	Fair	Poor	5-10m	Veteran (15% Lif...
Fraxinus oxycarpa	Less than 2m	Good	Good	less than 5m	Immature (80% L...

Figure 2. Input process to RapidMiner

Figure 2 explains the data input process to RapidMiner to apply the algorithm model to be used.

This process is the process of giving data patterns to a machine. Later the machine will read the pattern from the data entered and the results as an accurate new pattern reference in modeling the results. In this stage the sequence of tests carried out is the same as the testing of other models on RapidMiner.

Accredited Rank 3 (Sinta 3) based on the Decree of the Dirjen Penguatan RisBang Kemenristekdikti No. 28/E/KPT/2019, September 26, 2019. Published by PPPM STMIK Nusa Mandiri

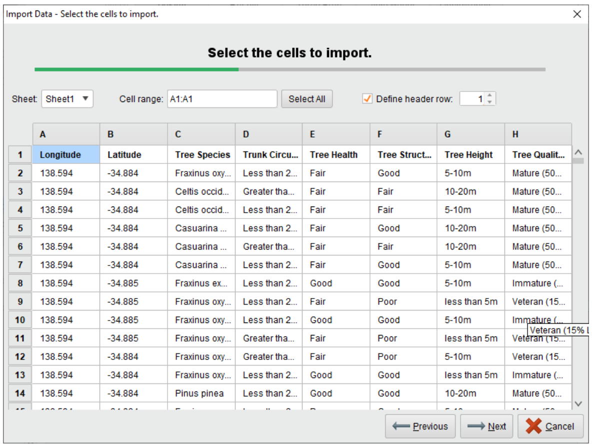


Figure 3. Input dataset

Figure 3 enter the stage to insert a dataset into RapidMiner with the attributes Tree Species, Trunk Circumference, Tree Health, Tree Structure, Tree Height, and Tree Quality life.

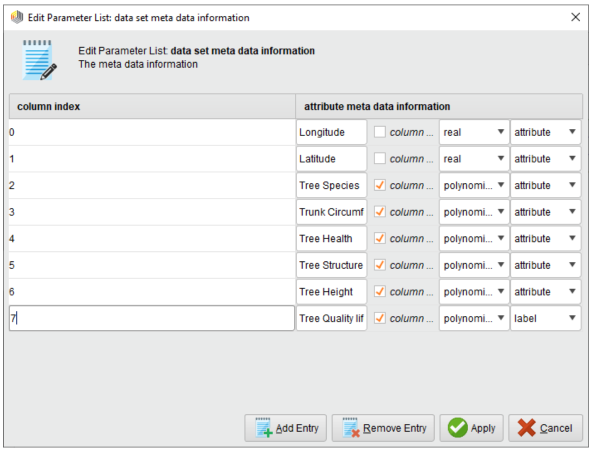


Figure 4. label determination process

Figure 4 to determine which records will be labeled for reference to the final output for data processing. In this study, the authors labeled Tree Quality life data.

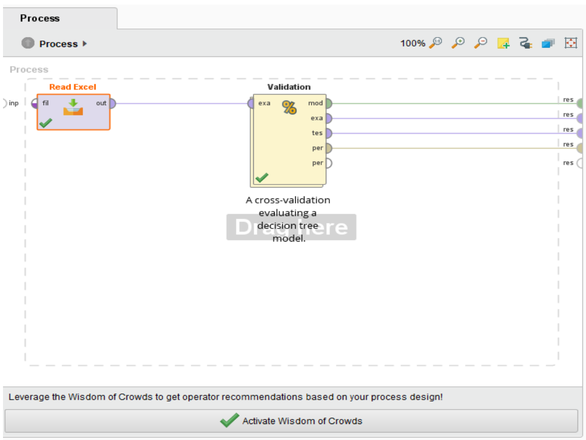


Figure 5. Process input validation

Figure 5 explains the classification of the Naive Bayes algorithm to determine the value of accuracy using the Validation model with a 10 fold validation scenario.

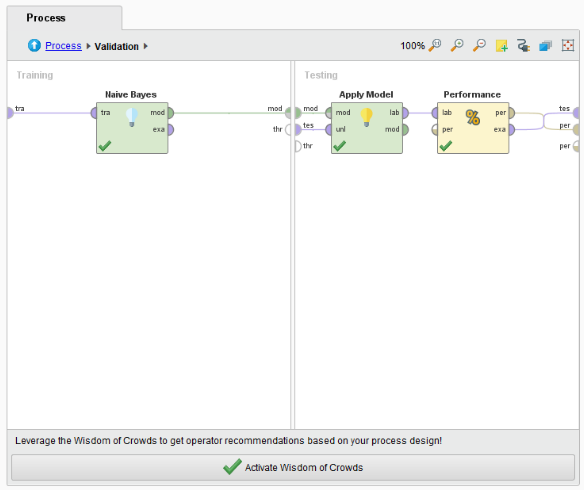
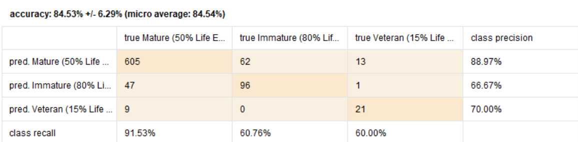


Figure 6. Application of the model.

Figure 6 explains the application of the model, enter the algorithm that will be used. In this study, the authors use Naive Bayes to process the data to be processed.

The accuracy obtained from is displayed in the Confusion Matrix table as follows.



accuracy: 84.53% +/- 6.29% (micro average: 84.54%)

	true Mature (50% Life E...	true Immature (80% Lif...	true Veteran (15% Life ...	class precision
pred. Mature (50% Life ...	605	62	13	88.97%
pred. Immature (80% Li...	47	96	1	66.67%
pred. Veteran (15% Life ...	9	0	21	70.00%
class recall	91.53%	60.76%	60.00%	

Figure 7. Accuracy value from Naive Bayes

Based on the figure 7 above, it can be seen and known the resulting accuracy is 84.53%. This shows the results of the accuracy produced in the excellent category.

Table 3. Precision, recall, and AUC results

Precision	*Recall*	*AUC*
0,840	0,848	0,873

From the Table 3, data processing by Naive Bayes produces a precision value of 0.840%, a recall value of 0.848%, and an AUC of 0.873%. These results indicate that the results are included in the excellent category.

VOL. 6. NO. 1 AUGUST 2020
P-ISSN: 2685-8223 | E-ISSN: 2527-4864
DOI: 10.33480/jitk.v6i1.1405

SimpleDistribution

```
Distribution model for label attribute Tree Quality life

Class Mature (50% Life Expectancy) (0.774)
5 distributions

Class Immature (80% Life Expectancy) (0.185)
5 distributions

Class Veteran (15% Life Expectancy) (0.041)
5 distributions
```

Figure 8. Naive Bayes distribution

From the figure 8 above, the Mature class with a life expectancy of 50% is 0.774%. Class Immature with an 80% life expectancy of 0.185% and a Veterans class with a life expectancy of 15% of 0.041%. Each class of distribution of the three as many as 5 classes. Mature Class with the largest life expectancy of 50% being the largest.

CONCLUSION

Based on research that has been done with Tools RapidMiner 9.0 on a dataset of life expectancy in urban trees obtained from the South Australian government website using the Naive Bayes method produces an accuracy of 84.53%. With a Precision value of 0.840%, a recall value of 0.848%, and an AUC of 0.873%. While Class Mature with the expected level of tree life of 50% to the largest of 0.774% This pattern can be used as a benchmark for determining tree life expectations for future classification purposes.

REFERENCE

[1] L. M. P. Gumaja, M. Mardhiansyah, and E. Sribudiani, "EVALUASI KESEHATAN POHON PADA JALUR HIJAU JALAN ARIFIN ACHMAD KOTA PEKANBARU," *J. Ilmu-Ilmu Kehutan.*, 2019.

[2] R. Hijauku, "9 Manfaat Pohon Di Perkotaan," 2017. https://hijauku.com/2017/02/22/9-manfaat-pohon-di-perkotaan/.

[3] A. Junaedi, "GROWTH PERFORMANCE OF THREE NATIVE TREE SPECIES FOR PULPWOOD PLANTATION IN DRAINED PEATLAND OF PELALAWAN DISTRICT, RIAU," *Indones. J. For. Res.*, 2018, DOI: 10.20886/ijfr.2018.5.2.119-132.

[4] "IKLIM MIKRO DAN KEBUTUHAN RUANG TERBUKA HIJAU DI KOTA SEMARANG (The Micro Climate and The Need of Green Open Space for The City of Semarang)," *J. Mns. dan Lingkung.*, 2008, doi: 10.22146/jml.18685.

[5] S. Hasanbahri, D. Marsono, S. Hardiwinoto, and R. Sadono, "Serangan benalu pada beberapa kelas umur tanaman jati di wilayah hutan BKPH Begal, KPH Ngawi, Jawa Timur," *Mns. dan Lingkung.*, 2014.

[6] A. Subiakto, "Teknologi Perbanyakan Vegetatif Bibit Pohon Hutan Secara Masal," in *Prosidding Ekspose Hasil-Hasil Penelitian*, 2007.

[7] M. Mawazin and H. Suhaendi, "Pengaruh jarak tanam terhadap pertumbuhan diameter," *J. Penelit. Hutan dan Konserv. Alam*, vol. 5, no. 4, pp. 381–388, 2008, doi: 10.20886/jphka.2008.5.4.381-388.

[8] Y. Suranto, A. P. Tibertius, D. Marsono, and G. S. Johanes Pramana, "Pengaruh Umur Pohon, Bonita dan Posisi Aksial Batang Terhadap Struktur Makroskopis dan Kualitas Kayu Jati Sebagai Bahan Furnitur," *J. Mns. dan Lingkung.*, vol. 22, no. 1, pp. 84–93, 2015.

[9] Karmilasanti and Abdurachman, "Struktur Tegakan Pada Umur 1,2,3 dan 4 Tahun Pada Hutan Bekas Tebangan Di Areal TPTJ PT. Intracawood Manufacturing, Kalimantan Utara," in *Prosiding Seminar Peningkatan Ekonomi Masyarakat Melalui Integrasi Hasil Riset Aplikatif*, 2018.

[10] S. Rahmadhani, S. B. Yuwono, A. Setiawan, and I. S. Banuwa, "PEMILIHAN JENIS POHON MENJERAP DEBU DI MEDIAN JALAN KOTA BANDAR LAMPUNG," *J. Belantara*, 2019, doi: 10.29303/jbl.v2i2.181.

[11] D. Pertiwi, R. Safe'i, H. Kaskoyo, and I. Indriyanto, "IDENTIFIKASI TIPE KERUSAKAN POHON MENGGUNAKAN METODE FOREST HEALTH MONITORING (FHM)," *PERENNIAL*, 2019, doi: 10.24259/perennial.v15i1.6033.

[12] P. A. Widyarini and Y. B. S. Heddy, "Penilaian Estetika dan Fungsional Pohon Tepi Jalan Berdasarkan Persepsi Pengguna Jalan (Studi Kasus : Jl Ijen dan Jl Veteran Kota Malang) Aesthetic and Functional Assessment of Roadside Trees Based on Street User Perceptions," *Produksi Tanam.*, vol. 6, no. 9, pp. 2320–2327, 2018.

[13] A. Saleh, "Implementasi Metode Klasifikasi Naïve Bayes Dalam Memprediksi Besarnya Penggunaan Listrik Rumah Tangga," *Creat. Inf. Technol. J.*, 2015.

[14] N. Nurajijah and D. Riana, "Algoritma Naïve Bayes, Decision Tree, dan SVM untuk Klasifikasi Persetujuan Pembiayaan Nasabah Koperasi Syariah," *J. Teknol. dan Sist. Komput.*, 2019, doi: 10.14710/jtsiskom.7.2.2019.77-82.

[15] E. S. Palupi, S. M. Pahlevi, U. Bina, S. Informatika, P. Magister, and I. Komputer,

"Inti nusa mandiri," *Inti Nusa Mandiri*, vol. 14, no. 2, pp. 133–138, 2020, doi: https://doi.org/10.33480/inti.v14i2.1178 VOL.

[16] I. A. Medha and D. P. Sari, "IMPLEMENTASI FUZZY DECISION TREE UNTUK PREDIKSI PENYAKIT LIVER PADA DATASET ILPD (INDIAN LIVER PATIENT DATASET)," *POSITIF J. Sist. dan Teknol. Inf.*, 2019, doi: 10.31961/positif.v5i2.773.

[17] S. Angra and S. Ahuja, "Implementation of data mining algorithms on student's data using rapid miner," in *Proceedings of the 2017 International Conference On Big Data Analytics and Computational Intelligence, ICBDACI 2017*, 2017, doi: 10.1109/ICBDACI.2017.8070869.

[18] Y. Angraini, S. Fauziah, and J. L. Putra, "ANALISIS KINERJA ALGORITMA C4.5 DAN NAÏVE BAYES DALAM MEMPREDIKSI KEBERHASILAN SEKOLAH MENGHADAPI UN," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, 2020, doi: 10.33480/jitk.v5i2.1233.

[19] T. P. Wicaksono, N. Hidayat, and B. Rahayudi, "Implementasi Metode Naive Bayes Pada Diagnosis Penyakit Lambung," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, 2019.

[20] S. Mujilahwati, F. Widyahastuti, and N. Nafi'iyah, "Performansi Klasifikasi Kelas Siswa Menggunakan Naïve Bayes Dan Decision Tree," *Semin. Nas. Unisla*, pp. 155–158, 2018.

[21] G. Australia, "City of Prospect Tree Species in Reserves," *2016*, 2016. https://data.sa.gov.au/data/dataset/city-of-prospect-tree-species-in-reserves.