

Klasifikasi Tingkat Kepuasan Penggunaan Layanan Teknologi Informasi Menggunakan Decision Tree

Devy Fatmawati^{1*}, Wulan Trisnawati¹, Yuwan Jumaryadi², Gandung Triyono¹

¹Program Studi Magister Ilmu Komputer, Fakultas Teknologi Informasi, Universitas Budi Luhur, Jakarta, Indonesia

²Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Mercu Buana, Jakarta, Indonesia

Email: ^{1*}2111601353@student.budiluhur.ac.id, ²2111601361@student.budiluhur.ac.id, ³yuwan.jumaryadi@mercubuana.ac.id,

⁴gandung.triyono@budiluhur.ac.id

Email Penulis Korespondensi: ¹2111601353@student.budiluhur.ac.id

Abstrak—Dalam rangka meningkatkan pengelolaan layanan IT perlu dilakukan analisis terhadap kepuasan pengguna layanan. Penentuan kinerja layanan yang berkualitas salah satunya ditentukan dari kepuasan atas layanan yang dirasakan oleh penggunanya. Hasil analisis kepuasan pengguna dilakukan dengan menganalisis hasil survey kepuasan pengguna layanan IT Pusintek. Survey kepuasan pengguna layanan IT Pusintek dilakukan terhadap seluruh pengguna layanan pada tahun 2022 dengan menggunakan 5 (lima) indikator kualitas pelayanan yaitu tangible, reliability, responsiveness, assurance, dan empathy. Hasil survey yang telah didapatkan kemudian akan dianalisis dengan menggunakan Decision Tree. Berdasarkan hasil pengujian yang dilakukan, didapatkan akurasi decision tree sebesar 98,10% dan Naïve Bayes sebesar 97,95%, sedangkan untuk recall decision tree sebesar 99,51% dan Naive Bayes sebesar 98,78%. Akan tetapi untuk tingkat precision naive bayes lebih baik dari decision tree, dimana precision naive bayes sebesar 99,02% dan Decision Tree 98,47%.

Kata Kunci: Decision Tree; Kepuasan; Layanan; Naive Bayes; Survey

Abstract—In order to improve the management of IT services, it is necessary to analyze the satisfaction of service users. One of the determinations of quality service performance is determined from the satisfaction with the service felt by its users. The results of the user satisfaction analysis were carried out by analyzing the results of the Pusintek IT service user satisfaction survey. The Pusintek IT service user satisfaction survey will be conducted for all service users in 2022 using 5 (five) service quality indicators, namely tangible, reliability, responsiveness, assurance, and empathy. The survey results that have been obtained will then be analyzed using a Decision Tree. Based on the results of the tests performed, the accuracy of the decision tree was 98.10% and Naïve Bayes was 97.95%, while for the recall decision tree was 99.51% and Naive Bayes was 98.78%. However, for the level of precision naive bayes, it is better than the decision tree, where the precision naive bayes is 99.02% and the decision tree is 98.47%.

Keywords: Decision Tree; Satisfaction; Service; Naive Bayes; Survey

1. PENDAHULUAN

Pusat Sistem Informasi dan Teknologi Keuangan (Pusintek) merupakan unit TIK di lingkungan salah satu kementerian di Indonesia yang bertugas untuk memberikan layanan IT kepada penggunanya melalui beberapa jenis layanan di bidang TIK yang disusun dalam Katalog Layanan IT (*IT Service Catalog Pusintek*). Kebutuhan layanan tersebut tidak hanya terbatas pada pengguna layanan yang berada di kantor pusat, tetapi juga melingkupi pengguna layanan di daerah.

Dalam rangka meningkatkan pengelolaan layanan IT perlu dilakukan analisis kepuasan pengguna layanan. Kepuasan merupakan respon pelanggan terhadap kebutuhan yang terpenuhi [1]. Adanya keluhan dari pengguna atas pelayanan yang diterima menunjukkan tingkat kualitas pelayanan dari penyedia jasa yang rendah [2][3][4]. Adapun indikator kualitas layanan yang menjadi perhatian dalam penelitian ini adalah *client satisfaction* dan *perceptions*.

Pengukuran kualitas layanan dapat membantu untuk meningkatkan kualitas layanan dari suatu instansi [5]. Adapun manfaat dari kepuasan penggunaan layanan bagi suatu instansi atau perusahaan yaitu dapat menciptakan komitmen yang kuat dalam membangun dan mempertahankan hubungan dengan perusahaan dalam jangka waktu lama, dapat meningkatkan kepercayaan pengguna terhadap produk/jasa yang diberikan, dapat meningkatkan kualitas hubungan yang baik antara pengguna dengan perusahaan, dan sebagai salah satu sarana promosi produk/jasa yang dilakukan oleh pengguna kepada pengguna lain, dimana pengguna yang puas akan kualitas yang diterima dapat bercerita dan memberikan rekomendasi kepada orang lain [6].

Agar dapat meningkatkan kepuasan terhadap kualitas layanan maka perlu dilakukan analisis terhadap kepuasan pengguna layanan yang diberikan oleh Pusintek. Kualitas layanan merupakan suatu ukuran untuk menilai sejauh mana perbedaan atas layanan yang didapatkan pengguna dengan layanan yang diharapkan pengguna. Kualitas layanan merupakan segala bentuk aktualisasi dari kinerja pelayanan yang dirasakan oleh pengguna yang terdiri dari *tangible*, *reliability*, *responsiveness*, *assurance*, dan *empathy* [7].

Penelitian ini bertujuan untuk menemukan faktor yang berpengaruh terhadap kepuasan pengguna terhadap layanan Teknologi Informasi dengan menggunakan teknik data mining. Berdasarkan data yang didapatkan dari Pusintek, ada 1316 layanan IT yang berhasil diselesaikan. Dari layanan yang berhasil diselesaikan tersebut, ada 1251 user yang puas dan 65 user yang tidak puas terhadap pelayanan yang diberikan.

Decision Tree termasuk teknologi untuk menemukan pola yang bermanfaat. Algoritma *decision tree* termasuk klasifikasi yang sederhana namun efektif [8]. Hal ini seperti penelitian yang dilakukan oleh Budiyantara, dkk (2020) dalam memprediksi mahasiswa lulus tepat waktu. Hasil akurasi dari Metode Decision Tree (C4.5) sebesar 98,04%, sedangkan akurasi Naïve Bayes sebesar 96,00%. Dan akurasi Metode K-Nearest Neighbor (K-NN) sebesar 90,00% [9]. *Decision tree* sangat berguna untuk analitik data dan *machine learning* karena memecah data kompleks menjadi bagian

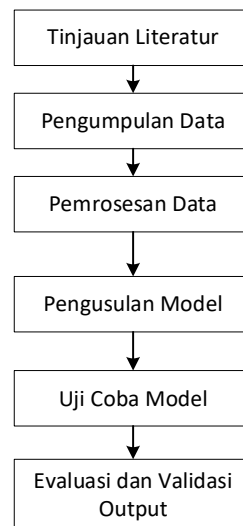
yang lebih mudah dikelola. Mereka sering digunakan untuk analisis prediksi, klasifikasi data, dan regresi [10]. *Decision tree* sering digunakan karena kejelasan dan kesederhanaannya dalam menemukan dan memprediksi data. Selain itu juga mudah dipahami karena dibuat berdasarkan aturan IF-THEN [11].

Teknik yang dapat digunakan untuk memvalidasi model untuk memvalidasi model untuk menilai bagaimana hasil statistik analisis akan menggeneralisasi data independen adalah *Cross validation* [12]. *k-fold cross validation* merupakan salah satu teknik dari *cross validation* dengan memecah data menjadi 'k' *dataset* data dengan ukuran yang sama. Pada penelitian ini data akan dipecah menjadi 5 jumlah 'k'. Adapun tujuan dari penelitian ini adalah untuk mengetahui kinerja algoritma *Decision Tree* dan Naive Bayes dengan menggunakan teknik *Cross validation* dengan jumlah 'k' sebanyak 5 untuk memprediksi kepuasan pengguna terhadap layanan IT pada Pusintek di salah satu kementerian di Indonesia.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Pada gambar 1 merupakan langkah-langkah penelitian yang dilakukan.



Gambar 1. Tahapan Penelitian

Dalam melakukan penelitian untuk memprediksi kepuasan pengguna layanan TIK Pusintek, peneliti melakukan pengumpulan informasi dari berbagai sumber seperti studi literatur serta melakukan kegiatan survey dengan menyebarkan kuesioner kepada pengguna layanan TIK. Kuesioner adalah metode pengumpulan informasi dengan cara menyediakan responden dengan daftar pertanyaan atau pernyataan tertulis berdasarkan teori [13]. Informasi survey disebar ke pengguna layanan TIK Pusintek mulai tahun 2022 dalam bentuk kuesioner. Indikator Evaluasi diukur dengan 5 karakteristik yaitu kekonkretan, keandalan, ketanggapan, keamanan dan empati dengan skala Likert. Selanjutnya dilakukan analisis atas sejumlah kuesioner yang berhasil didapatkan dengan menggunakan metode Decision Tree C4.5 dan Naive Bayes. Adapun data yang berhasil dikumpulkan untuk dianalisis berjumlah 1316 data.

2.2 Decision Tree C4.5

Salah satu algoritma untuk melakukan klasifikasi adalah Decision Tree C4.5 [14]. Algoritma C4.5 bekerja dengan cara mengunjungi setiap *decision node*, dan memilih percabangan yang optimal [15]. Pemrosesan data dilakukan dengan melakukan pengolahan data dari hasil kuesioner yang kemudian di export ke bentuk excel untuk dilakukan analisis, cleansing data dari noisy dan duplikasi data. Setelah proses tersebut selesai, kemudian dilanjutkan dengan melakukan pemrosesan data. Adapun tahapan dalam decision tree C4.5 adalah mempersiapkan data training, dan menghitung root dari decision tree.

2.3 Naïve Bayes

Algoritma Naïve Bayes merupakan model yang diusulkan untuk memprediksi kepuasan pengguna terhadap layanan TIK yang diberikan oleh Pusintek. Bayes merupakan teknik untuk memprediksi yang berbasis probabilitas sederhana dengan berdasarkan penerapan teorema Bayes yang berasumsikan independensi yang kuat. Adapun rumus umum untuk teorema Bayes adalah [16] :

$$P(A|B) = \frac{P(B | A) \cdot P(A)}{P(B)} \quad (1)$$

Keterangan :

B : Data klasifikasi yang belum diketahui.

- A : Kelas spesifik pada data hipotesis
- P(A) : Probabilitas hipotesis A.
- P(B) : Probabilitas B.
- P(B|A) : Probabilitas B atas dasar kondisi hipotesis A.
- P(A|B) : Probabilitas hipotesis A berdasarkan kondisi B.

2.4 Confusion Matrix

Evaluasi dalam penelitian ini dijalankan dengan menganalisis hasil klasifikasi. Confusion Matrix adalah metode untuk menganalisis seberapa baik contoh pengelompokan yang dibangun untuk mengenali data dari level yang berbeda [17]. Kinerja model klasifikasi dapat diukur dari tingkat akurasinya [18][19]. Keakuratan diperoleh dari dataset yang diklasifikasikan dengan benar seperti pada Tabel 1.

Tabel 1. Confusion Matrix

Klasifikasi Benar	Diklasifikasikan Sebagai	
	+	-
+	TP	FN
-	FN	TP

Keterangan:

- TP (True Positive): proporsi positif yang diklasifikasikan positif dalam dataset.
- FN (False Negative): proporsi positif yang diklasifikasikan negatif dalam dataset.
- FP (False Positive): proporsi negatif yang diklasifikasikan positif dalam dataset.
- TN (True Negative): proporsi negatif yang diklasifikasikan negatif dalam dataset.

Adapun untuk persamaan dalam confusion matrix sebagai berikut

1. Nilai akurasi merupakan proporsi untuk jumlah prediksi yang benar. Berikut ini merupakan rumus yang digunakan untuk menghitung akurasi.

$$Akurasi = \frac{TP+TN}{TP+FP+TN+FN} \tag{2}$$

2. Precision probabilitas bahwa item yang dipilih relevan. Berikut ini merupakan rumus untuk menghitung akurasi.

$$Precision = \frac{T}{TP+FN} \tag{3}$$

3. Recall adalah probabilitas bahwa item yang relevan akan dipilih. Berikut ini merupakan rumus untuk menghitung tingkat pengembalian:

$$Recall = \frac{T}{TP+FP} \tag{4}$$

Jika nilai precision dan nilai recall tinggi, maka rekomendasi yang dihasilkan dianggap baik [20]. Validasi dari penelitian ini diterapkan terhadap hasil suvey berupa pengisian kuesioner oleh 1316 responden kepuasan pengguna layanan Pusintek yang kemudian data diuji dengan metode Decision Tree.

3. HASIL DAN PEMBAHASAN

3.1 Analisis Data

Berikut ini merupakan analisis terhadap kepuasan layanan Teknologi Informasi yang diukur berdasarkan *tangible*, *reliability*, *responsiveness*, *assurance*, dan *empathy* yang dilakukan.

Tabel 2. Perhitungan Manual Probabilitas Tangible

	Puas	Prob Puas	Tidak Puas	Prob Tidak Puas
Tinggi	1238	0,9896	20	0,3077
Rendah	13	0,0104	45	0,6923

Pada Tabel 2 merupakan perhitungan terhadap dimensi *tangible*. Ada 1238 user yang merasakan tingkat kepuasan tinggi, dan 13 user yang merasakan tingkat kepuasan rendah terhadap dimensi *tangible*. Adapun user yang merasakan kepuasan tinggi terhadap dimensi *tangible* dan tidak puas terhadap layanan yang diberikan berjumlah 20. Dan user yang merasakan kepuasan rendah terhadap dimensi *tangible* dan tidak puas terhadap layanan yang diberikan berjumlah 40.

Tabel 3. Perhitungan Manual Probabilitas Reliability

	Puas	Prob Puas	Tidak Puas	Prob Tidak Puas
Tinggi	1228	0,9816	1	0,0154
Rendah	23	0,0184	64	0,9846

Pada Tabel 3 merupakan perhitungan terhadap dimensi *reliability*. Ada 1228 user yang merasakan tingkat kepuasan tinggi, dan 23 user yang merasakan tingkat kepuasan rendah terhadap dimensi *reliability*. Adapun user yang merasakan kepuasan tinggi terhadap dimensi *reliability* dan tidak puas terhadap layanan yang diberikan berjumlah 1. Dan user yang merasakan kepuasan rendah terhadap dimensi *reliability* dan tidak puas terhadap layanan yang diberikan berjumlah 64.

Tabel 4. Perhitungan Manual Probabilitas Responsiveness

	Puas	Prob Puas	Tidak Puas	Prob Tidak Puas
Tinggi	1241	0,9920	7	0,1077
Rendah	10	0,0080	58	0,8923

Pada Tabel 4 merupakan perhitungan terhadap dimensi *responsiveness*. Ada 1241 user yang merasakan tingkat kepuasan tinggi, dan 10 user yang memiliki tingkat kepuasan rendah terhadap dimensi *responsiveness*. Adapun user yang merasakan kepuasan tinggi terhadap dimensi *responsiveness* dan tidak puas terhadap layanan yang diberikan berjumlah 7. Dan user yang merasakan kepuasan rendah terhadap dimensi *responsiveness* dan tidak puas terhadap layanan yang diberikan berjumlah 58.

Tabel 5. Perhitungan Manual Probabilitas Assurance

	Puas	Prob Puas	Tidak Puas	Prob Tidak Puas
Tinggi	1216	0,9720	2	0,0308
Rendah	35	0,0280	63	0,9692

Pada Tabel 5 merupakan perhitungan terhadap dimensi *assurance*. Ada 1216 user yang merasakan tingkat kepuasan tinggi, dan 35 user yang merasakan tingkat kepuasan rendah terhadap dimensi *assurance*. Adapun user yang merasakan kepuasan tinggi terhadap dimensi *assurance* dan tidak puas terhadap layanan yang diberikan berjumlah 2. Dan user yang merasakan kepuasan rendah terhadap dimensi *assurance* dan tidak puas terhadap layanan yang diberikan berjumlah 63.

Tabel 6. Perhitungan Manual Probabilitas Empathy

	Puas	Prob Puas	Tidak Puas	Prob Tidak Puas
Tinggi	1227	0,9808	4	0,0615
Rendah	24	0,0192	61	0,9385

Pada Tabel 6 merupakan perhitungan terhadap dimensi *assurance*. Ada 1227 user yang merasakan tingkat kepuasan tinggi, dan 24 user yang memiliki tingkat kepuasan rendah terhadap dimensi *assurance*. Adapun user yang merasakan kepuasan tinggi terhadap dimensi *assurance* dan tidak puas terhadap layanan yang diberikan berjumlah 4. Dan user yang merasakan kepuasan rendah terhadap dimensi *assurance* dan tidak puas terhadap layanan yang diberikan berjumlah 61.

3.2 Pemrosesan Data

Pada Tabel 7 merupakan dataset yang digunakan dalam penelitian. Dataset tersebut didapatkan dari kuesioner yang dibagikan kepada pengguna layanan Pusintek tahun 2022.

Tabel 7. Dataset Awal

Data ke-	T	R1	R2	R3	R4	Re1	Re2	Re3	A1	A2	E1	E2
1	4	4	4	4	4	4	4	4	4	4	4	4
2	5	5	5	5	5	5	5	5	5	5	5	5
3	5	5	5	5	5	5	5	5	5	5	5	5
4	5	5	5	5	5	5	5	5	5	5	5	5
5	4	3	3	3	3	3	4	4	3	3	3	3
6	5	5	5	5	5	5	5	5	5	5	5	5
7	4	5	5	5	4	5	4	4	5	5	4	5
8	5	4	4	4	4	4	5	5	4	4	4	4
9	5	5	5	5	5	5	5	5	5	5	5	5
10	2	2	2	2	2	2	2	3	2	2	2	2
..
..
..
1315	5	5	5	5	5	5	5	5	5	5	5	5
1316	4	4	4	4	4	4	4	4	4	4	4	4

Data yang telah didapatkan akan dikelompokkan menjadi dua kelompok berdasarkan data sampel per indikator, yaitu rendah dan tinggi. Hal ini dapat dilihat pada Tabel 8.

Tabel 8. Pengelompokan Nilai Kuesioner

No	Nilai	Kelas
1	1 - 3	Rendah
2	4 - 5	Tinggi

Selain itu, penulis juga mengelompokan hasil kepuasan menjadi dua kelompok yaitu puas dan tidak puas sebagaimana terlihat pada Tabel 9.

Tabel 9. Klasifikasi Hasil Nilai Kuesioner

No	Nilai	Kelas
1	1 - 3	Tidak Puas
2	4 - 5	Puas

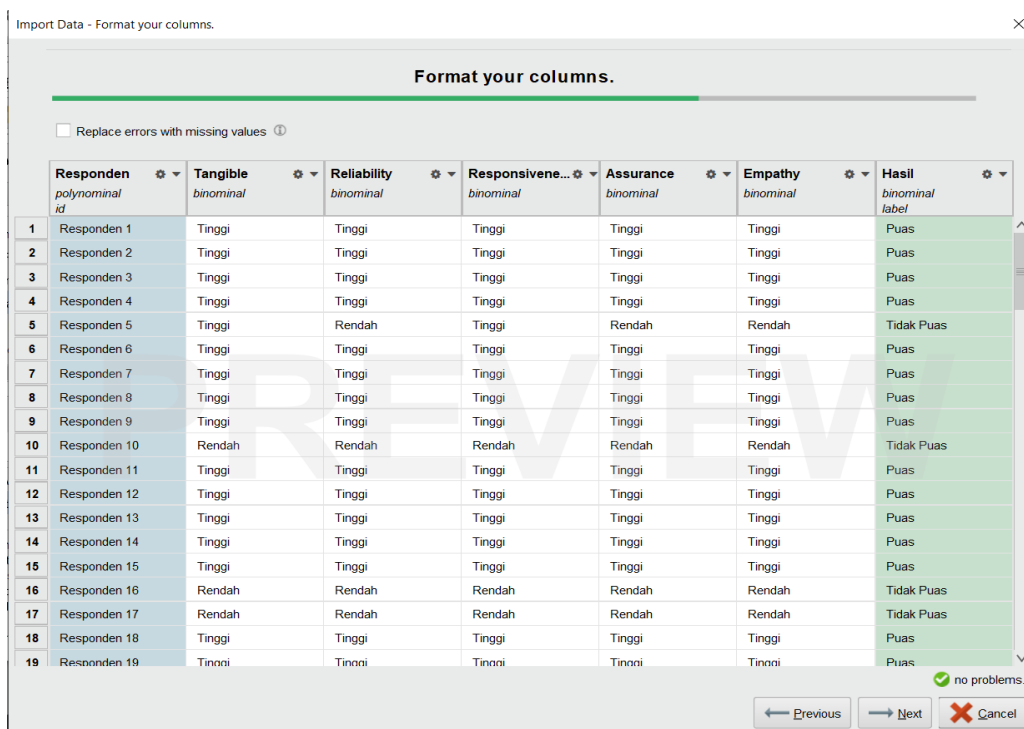
Pada Tabel 10 merupakan dataset yang telah diolah dengan menggunakan Microsoft Excel.

Tabel 10. Hasil Pengolahan Dataset

Data ke-	T	R	Re	A	E	Hasil
1	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
2	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
3	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
4	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
5	Tinggi	Rendah	Tinggi	Rendah	Rendah	Tidak Puas
6	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
7	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
8	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
9	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
10	Rendah	Rendah	Rendah	Rendah	Rendah	Tidak Puas
..
..
..
1315	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas
1316	Tinggi	Tinggi	Tinggi	Tinggi	Tinggi	Puas

3.3 Proses Penambahan Data Menggunakan RapidMiner

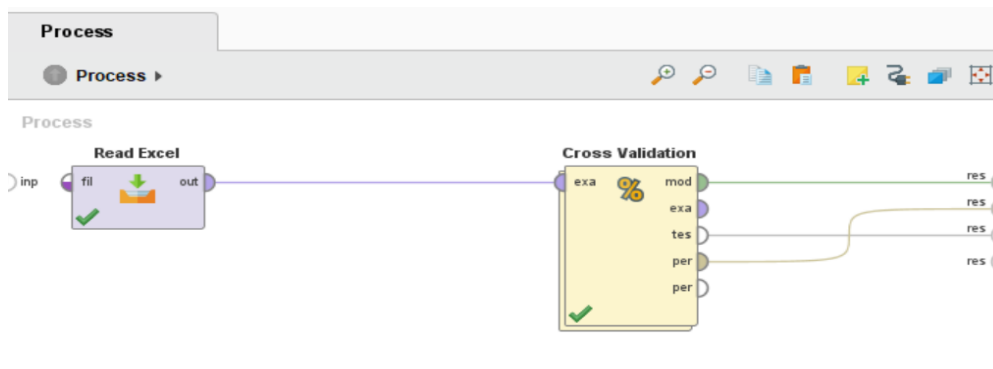
Untuk mendapatkan nilai prediksi kepuasan pengguna layanan IT Pusintek, data yang diperoleh akan di proses oleh Rapidminer dimana pengujiannya menggunakan Cross Validation. Gambar 2 adalah proses import data ke RapidMiner.



Gambar 2. Proses import data ke RapidMiner

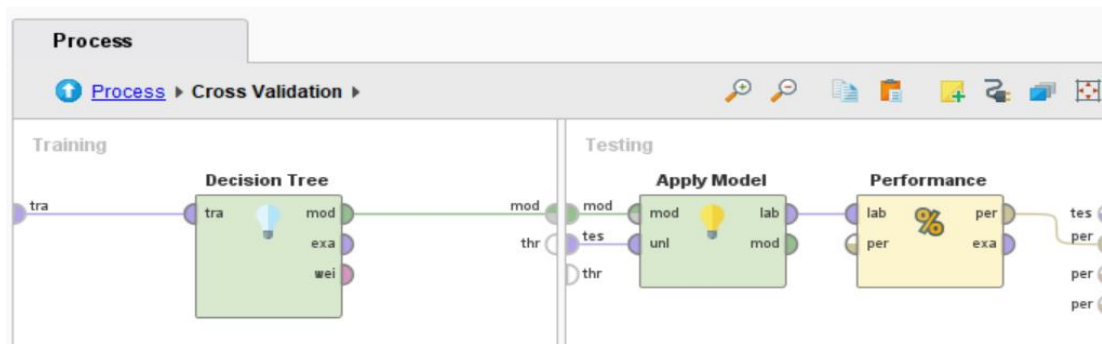
3.4 Uji Coba Model

Algoritma Decision Tree digunakan sebagai pemodelan terhadap proses klasifikasi yang dilakukan, adapun pemodelan Naive Bayes dilakukan untuk membandingkan tingkat akurasi yang ada.. Data disiapkan dalam format Microsoft excel untuk diimpor ke dalam Rapidminer. Pada Gambar 3 merupakan proses import data untuk melakukan pemodelan dengan menggunakan teknik Cross Validation.



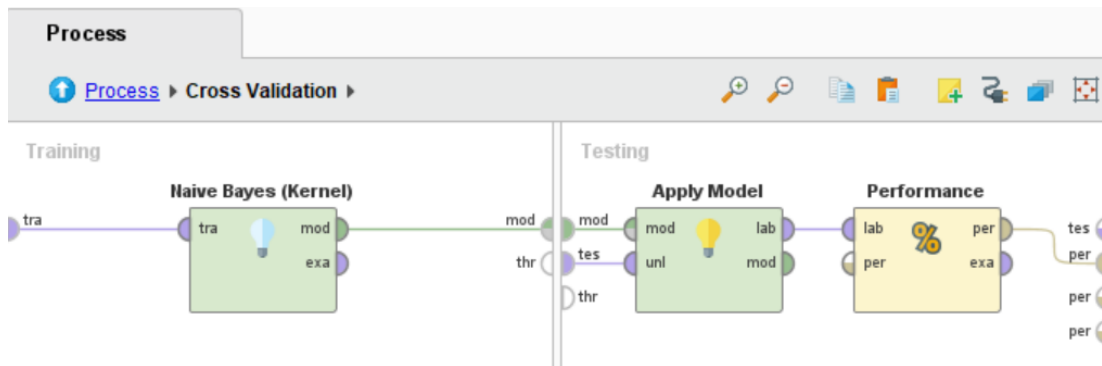
Gambar 3. Import Data untuk Pemodelan

Setelah data di impor ke dalam Rapidminer, maka tahap selanjutnya adalah menyambungkan data tersebut dengan operator Cross validation. Adapun jumlah ‘k’ pada k-fold yang digunakan adalah 5. Hal ini dapat dilihat pada Gambar 3.



Gambar 4. Pemodelan Decision Tree

Pada Gambar 4 merupakan teknik pemodelan dengan menggunakan Decision Tree. Adapun pemodelan yang dilakukan untuk mengetahui tingkat akurasi menggunakan teknik Cross Validation dengan jumlah ‘k’ pada k-fold yang digunakan adalah 5.



Gambar 5. Pemodelan Naive Bayes

Pada Gambar 5 merupakan pemodelan dengan menggunakan Naive Bayes. Adapun pemodelan yang dilakukan untuk mengetahui tingkat akurasi menggunakan teknik Cross Validation.

Tabel 11. Hasil Pengujian Akurasi

No	Algoritma	Kinerja		
		Accuracy	Recall	Precision
1.	Decision Tree	98.10	99.51%	98.47%
2.	Naive Bayes	97.95	98.78%	99.02%

Pada Tabel 11 merupakan hasil pengujian algoritma naive bayes dan decision tree berdasarkan accuracy, recall, dan precision.

4. KESIMPULAN

Berdasarkan hasil pengujian data untuk kepuasan pelanggan di Pusintek menggunakan *Decision Tree* dengan teknik Cross Validation, dan jumlah 'k' pada k-fold cross validation adalah 5 didapatkan tingkat akurasi dan recall Decision Tree lebih tinggi dari pada Naive Bayes, dimana akurasi decision tree sebesar 98,10% dan *Naive Bayes* sebesar 97,95%, sedangkan untuk recall decision tree sebesar 99.51% dan Naive Bayes sebesar 98.78%. Akan tetapi untuk tingkat precision naive bayes lebih baik dari decision tree, dimana precision naive bayes sebesar 99.02% dan Decision Tree 98.47%. Untuk mendapatkan hasil yang lebih baik, diharapkan pada penelitian selanjutnya dapat menggunakan parameter yang berbeda, sehingga dapat diketahui tingkat akurasi dari parameter yang lain.

REFERENCES

- [1] D. T. S. Kaihatu, *Manajemen Komplain*. CV. Andi Offset, 2015.
- [2] Fattya Ariani and Andi Taufik, "Perbandingan Metode Klasifikasi Data Mining untuk Prediksi Tingkat Kepuasan Pelanggan Telkomsel Prabayar," *SATIN - Sains dan Teknol. Inf.*, vol. 6, no. 2, pp. 46–55, 2020, doi: 10.33372/stn.v6i2.666.
- [3] B. Tarhini and D. Hayek, "Business Revolution in a Digital Era," in *14th International Conference on Business Excellence*, 2020, p. 405.
- [4] J. Wang, J. Y. Lai, and Y. H. Lin, "Social media analytics for mining customer complaints to explore product opportunities," *Comput. Ind. Eng.*, vol. 178, no. January, p. 109104, 2023, doi: 10.1016/j.cie.2023.109104.
- [5] R. Mardiyanto and M. Ismowati, "Analisis Indeks Kepuasan Masyarakat Dalam Upaya Peningkatan Kepuasan Kualitas Pelayanan Masyarakat Di Kantor Kecamatan Kotabaru Kabupaten Karawang," *Transparansi J. Ilm. Ilmu Adm.*, vol. 9, no. 2, pp. 184–197, 2018, doi: 10.31334/trans.v9i2.23.
- [6] Suherman, D. Maulana, and V. Mustikaningtyas, "Penerapan Data Mining untuk Klasifikasi Kepuasan Pelanggan Transportasi Online (Ojek Online) Menggunakan Algoritma C4.5," *Pros. SAINTEK Sains dan Teknol.*, vol. 1, no. 1, pp. 165–175, 2022.
- [7] S. Kim and H. S. Kim, "A Study on the Effect of Medical Service Quality on Customer Satisfaction during COVID-19 for Foreigners in Korea," *Sustain.*, vol. 15, no. 7, pp. 1–12, 2023, doi: 10.3390/su15075953.
- [8] J. J. Purnama, H. M. Nawawi, S. Rosyida, Ridwansyah, and Risandar, "Klasifikasi Mahasiswa Her Berbasis Algoritma SVM Dan Decision Tree," *J. Teknol. Inf. dan Ilmu Komput.*, vol. 7, no. 6, pp. 1253–1260, 2020, doi: 10.25126/jtiik.202073080.
- [9] A. Budiyantara, I. Irwansyah, E. Prengki, P. A. Pratama, and N. Wiliani, "Komparasi Algoritma Decision Tree, Naive Bayes Dan K-Nearest Neighbor Untuk Memprediksi Mahasiswa Lulus Tepat Waktu," *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 2, pp. 265–270, 2020, doi: 10.33480/jitk.v5i2.1214.
- [10] S. Zaki, N. I. Ghali, A. Abo-Elfetoh, and A. M. Idrees, "Predictive Analysis of Big data in Egypt Census 2017 Comparison of Four ML Predictive Models," *SSRN Electron. J.*, pp. 1–10, 2022, doi: 10.2139/ssrn.4225813.
- [11] Y. A. Alsariera, Y. Baashar, G. Alkawsi, A. Mustafa, A. A. Alkahtani, and N. Ali, "Assessment and Evaluation of Different Machine Learning Algorithms for Predicting Student Performance," *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/4151487.
- [12] H. Azis, P. Purnawansyah, F. Fattah, and I. P. Putri, "Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung," *Ilk. J. Ilm.*, vol. 12, no. 2, pp. 81–86, 2020, doi: 10.33096/ilkom.v12i2.507.81-86.
- [13] A. H. D. Purwanto, M. Nashar, Y. Jumaryadi, W. Wibowo, and A. Mekaniwati, "Improving medium small micro enterprise (MSME) performance," *Int. J. Adv. Appl. Sci.*, vol. 9, no. 5, pp. 37–46, May 2022, doi: 10.21833/ijaas.2022.05.005.
- [14] Q. Guo, C. Minghua, D. Yueli, An Du, and Y. Linlei, "Prediction of students' course failure based on campus card data," in *Proceedings - 2019 International Conference on Robots and Intelligent System, ICRIS 2019*, 2019, pp. 361–364. doi: 10.1109/ICRIS.2019.00097.
- [15] D. Y. Putri, R. Andreswari, and M. A. Hasibuan, "Analysis of Students Graduation Target Based on Academic Data Record Using C4.5 Algorithm Case Study: Information Systems Students of Telkom University," in *The 6th International Conference on Cyber and IT Service Management (CITSM 2018)*, 2018, pp. 1–6. doi: 10.1109/CITSM.2018.8674366.
- [16] B. Deta and T. Mauritsius, "Implementation of Data Mining to Determine Student Majors Using The Machine Learning," *J. Theor. Appl. Inf. Technol.*, vol. 100, no. 11, pp. 3798–3809, 2022.
- [17] E. N. Fitri, S. Winarno, F. Budiman, A. Rohmani, J. Zeniarja, and E. Sugiarto, "Decision Tree Simplification Through Feature Selection Approach In Selecting Fish Feed Sellers," *J. Tek. Inform.*, vol. 4, no. 2, pp. 301–309, 2023.
- [18] D. Chicco and G. Jurman, "The Matthews correlation coefficient (MCC) should replace the ROC AUC as the standard metric for assessing binary classification," *BioData Min.*, vol. 16, no. 1, pp. 1–23, 2023, doi: 10.1186/s13040-023-00322-4.
- [19] F. Sutomo *et al.*, "Optimization Of The K-Nearest Neighbors Algorithm Using The Elbow Method On Stroke Prediction," *J. Tek. Inform.*, vol. 4, no. 1, pp. 125–130, 2023.
- [20] N. W. Wardani and N. K. Ariasih, "Analisa Komparasi Algoritma Decision Tree C4.5 dan Naive Bayes untuk Prediksi Churn Berdasarkan Kelas Pelanggan Retail," *Int. J. Nat. Sci. Eng.*, vol. 3, no. 3, p. 103, 2019, doi: 10.23887/ijnse.v3i3.23113.