

Penerapan K-Nearest Neighbor Untuk Klasifikasi Tingkat Kelulusan Pada Siswa

Esty Purwaningsih^{1*}, Ela Nurelasari²

^{1,2)} Program Studi Sistem Informasi, Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98, Senen, Jakarta Pusat
Email: * esty.epw@bsi.ac.id

Abstrak. Tingkat kelulusan pada siswa salah satunya terbentuk dari hasil pencapaian kinerja siswa dalam suatu pembelajaran. Prediksi dari tingkat kelulusan berdasarkan kinerja siswa bermanfaat untuk menganalisa siswa yang kurang berprestasi dalam kegiatan akademik dan juga dapat memberikan dukungan pada siswa yang mengalami kesulitan saat pembelajaran. Terdapat faktor yang menjadi pertimbangan dalam memprediksi tingkat kelulusan siswa diantaranya nilai akademik, sikap dan juga faktor sosial. Namun itu saja tidak cukup untuk dijadikan faktor dalam memprediksi kinerja siswa. Para pengajar pun belum efektif untuk memilah faktor mana yang mempengaruhi kinerja siswa. Predikat prestasi terhadap kinerja siswa diperoleh dari hasil sebuah prediksi. Prediksi dilakukan dengan menggunakan metode K-Nearest Neighbor (KNN). K-Nearest Neighbor sering digunakan pada klasifikasi kinerja siswa karena kesederhanaannya juga dapat memberikan hasil yang signifikan dan kompetitif. Hasil dari prediksi tingkat kelulusan siswa dengan metode KNN didapat rata-rata akurasi dengan nilai sebesar 96,49%. Pengolahan data dilakukan dengan menggunakan tools rapid miner. Output dari implementasi pada prediksi tingkat kelulusan dapat dijadikan sebagai acuan bagi siswa untuk meningkatkan prestasi dan predikat studi lanjut dimasa yang akan datang.

Kata kunci: *kinerja siswa, kelulusan, k-nearest neighbor (knn), data mining, klasifikasi.*

1 Pendahuluan

Saat ini, Pendidikan adalah salah satu bagian terpenting dan esensial bagi kehidupan masyarakat. Ini digunakan untuk meningkatkan dan meningkatkan pertumbuhan individu secara akademis dan finansial. Seorang individu yang berpendidikan harus berkontribusi tidak hanya untuk keluarganya tetapi juga untuk m

asyarakat dan komunitas. Ini semua dicapai melalui pembelajaran yang tepat [1]

Untuk dapat menghasilkan SDM (sumber daya manusia) yang cakap, berwawasan, kompetitif dan kreatif, lembaga pendidikan diminta untuk

menyelenggarakan pendidikan yang bermutu dan berkualitas bagi peserta didiknya. Sehingga untuk mengimplementasikannya dilakukan ujian sekolah ataupun ujian nasional guna menentukan kelulusan pada tingkatan pendidikan serta menjadi tolak ukur dalam satuan pendidikan.

Berdasarkan Peraturan Menteri 'Pendidikan Dan 'Kebudayaan Republik Indonesia Nomor 43 Tahun 2019 Tentang Penyelenggaraan Ujian Yang Diselenggarakan Satuan Pendidikan Dan Ujian Nasional bahwa peserta yang diajarkan atau dididik dapat dinyatakan lulus apabila telah menyelesaikan masa studi dan memperoleh nilai sikap atau perilaku dengan nilai kategori minimal baik serta dapat ikut serta dalam tes yang diadakan oleh instansi ataupun satuan pendidikan yang ditetapkan. Untuk dapat mengukur hasil ujian para siswa beserta kelulusannya bisa ditunjukkan dari kualitas dan kelemahan siswa tersebut dalam hal pembelajaran [2] sehingga diupayakan untuk selalu mendukung siswa yang kesulitan dalam pembelajaran dalam satuan/program pendidikan [3].

Prestasi siswa, kemajuan siswa dan potensi siswa sangat penting untuk mengukur hasil belajar, pemilihan materi pembelajaran dan kegiatan pembelajaran. Namun, pekerjaan yang ada tidak menyediakan alat analisis yang cukup untuk menganalisis bagaimana kinerja siswa, faktor mana yang akan memengaruhi kinerja mereka, dengan cara apa siswa dapat membuat kemajuan, dan apakah siswa memiliki potensi untuk tampil lebih baik [4].

Tingkat keberhasilan kelulusan dari para siswa terdapat berbagai macam faktor yang mempengaruhi [3]. Satu dari faktor utamanya adalah nilai harian yang berdampak ada nilai ujian, namun tidak itu saja dalam memprediksi kinerja siswa dalam peningkatan kelulusan faktor lainnya juga turut andil diantaranya perilaku dan disiplin yang juga mempengaruhinya [5].

Meningkatnya jumlah kelulusan pada siswa dalam suatu satuan/program pendidikan dapat mengangkat dan meningkatkan popularitas satuan/program pendidikan tersebut. Jika satuan/program pendidikan dikatakan menurun dalam ratingnya, maka dapat memungkinkan berpengaruh pada daya tarik calon siswa baru yang akan masuk ke dalam satuan/program pendidikan tersebut. Sehingga tentu saja dapat menimbulkan kekhawatiran pada satuan/program pendidikan [6].

Nilai yang dapat dijadikan faktor penunjang dalam keberhasilan siswa untuk dapat lulus diantaranya terdapat nilai ujian baik teori maupun praktik, nilai tengah semester (uts), nilai ujian akhir serta nilai sikap. Penelitian terkait kinerja siswa, kebanyakan memakai metode klasifikasi [3].

Dalam menganalisis tingkat keberhasilan/kelulusan dari siswa/peserta didik, penelitian ini menerapkan metode K-Nearest Neighbor. Konsep KNN

tergolong sebagai ‘konsep klasifikasi yang ‘sederhana dan menjadi populer dalam penelitian terkait klasifikasi [7]. Menurut Imandoust dan Bolandraftar [8] terdapat keunggulan yang dimiliki diantaranya yaitu adanya kemudahan, serta efektivitas, dan intuitif, juga performa dari klasifikasi yang sangat kompetitif dalam kebanyakan domain. Tujuan dari penelitian pada studi kasus ini adalah untuk data mengetahui prediksi dari tingkat keberhasilan siswa dalam kelulusannya berdasarkan dari kinerja siswa serta mengetahui proses kerja dari metode K-Nearest Neighbor dalam mengklasifikasikan tingkat keberhasilan siswa untuk lulus.

2 Kajian Literature

K-Nearest Neighbor (KNN) menurut Gray et al, mempunyai kinerja yang baik dalam klasifikasi untuk menangani dataset dengan jumlah yang kecil. [9]. Metode ini membandingkan kesamaan catatan training yang terdekat dengannya. Nilai k mewakili jumlah tetangga yang dibandingkan. Nilai k menunjukkan banyaknya k data terdekat yang dibandingkan [10].

Terdapat beberapa penelitian mengenai kinerja dan kelulusan siswa salah satu diantaranya menggunakan model K-Nearest Neighbor [2]. Adapun penelitian sebelumnya terkait kinerja siswa, bahwa prediksi dari hasil pembelajaran siswa terdapat beberapa pengklasifikasi terintegrasi. Pengklasifikasi terintegrasi terdiri dari tiga algoritma pelengkap, yaitu *Decision Tree*, *K-Nearest Neighbor*, dan *Aggregating One-Dependence Estimators* (AODE) yang memberikan solusi umum untuk memprediksi hasil kerja siswa [2].

Penelitian [11] menyatakan Informasi kemajuan pembelajaran siswa tidak cukup untuk membentuk indikator dalam mengevaluasi kinerja siswa. Klasifikasi dalam hasil pembelajaran siswa yang efektif adalah dengan metode KNN. Dengan menggunakan set data yaitu student performance menghasilkan nilai rata-rata akurasi yaitu 74,068 dan nilai rata-rata nilai akurasi terbesar yaitu 76,553 untuk metode KNN dan Information Gain.

Penelitian [12], jenis klasifikasi yang cocok untuk meramalkan kinerja dari siswa-siswa adalah dengan algoritma KNN. Setelah data diolah, disimpulkan bahwa terdapat prediksi paling baik yaitu pada nilai K sejumlah 3, 6, dan 9. Peneliti mencoba memberikan nilai K dari 3 hingga 60 kemudian prediksi dibandingkan, kemudian didapatkan nilai terbaik.

Pada penelitian ini merumuskan model siswa dengan atribut terkait kinerja dan atribut terkait non-kinerja dengan Student Attribute Matrix (SAM), dan menyediakan alat estimasi kinerja siswa dengan menggunakan Back Propagation

Neural Network (BP-NN), yang dapat memperkirakan prestasi / atribut siswa menurut pengetahuan awal siswa serta kinerja / atribut siswa lain yang memiliki kesamaan karakteristik. Hasil evaluasi menunjukkan bahwa alat yang diajukan dapat memberikan hasil yang benar dan akurat, dan juga pemahaman yang lebih baik tentang kemajuan siswa [4].

Kemampuan untuk memprediksi kinerja siswa dapat berguna dalam banyak cara berbeda yang terkait dengan pembelajaran jarak jauh tingkat universitas. Nilai siswa dalam beberapa tugas tertulis dapat menjadi set pelatihan untuk algoritme machine learning yang diawasi. Saat ini, penggabungan pengklasifikasi diusulkan sebagai arah baru untuk meningkatkan akurasi klasifikasi. Namun, sebagian besar algoritma ansambel beroperasi dalam mode batch. Penelitian kali ini menggunakan ansambel pengklasifikasi online yang menggabungkan versi Naive Bayes, algoritma 1-NN dan WINNOWER [13].

Penelitian ini membandingkan kinerja C4.5, Naive Bayes, AODE, dan Multi-labeled K-Nearest Neighbor menganalisis kinerja lingkungan akademik. Hasil eksperimen menyimpulkan bahwa Multi-labeled K- Nearest Neighbor memiliki akurasi waktu yang paling baik dalam klasifikasi jika dibandingkan dengan teknik lain dengan signifikansi hasil pemeriksaan dan aktivitas lain yang dipengaruhi oleh aturan yang ditetapkan. Penelitian ini sangat berguna untuk mengidentifikasi rasio anak lamban belajar untuk memperbaiki kegagalan sejak dini dan mengambil tindakan untuk memperbaiki siswa yang lemah secara sempurna. Kedepannya, teknik yang sesuai dari algoritma K-Nearest Neighbor multi-label akan digunakan untuk analisis dasar keterampilan siswa dan juga untuk mengetahui tingkat kognitif siswa melalui metode observasi, metode perseptual, pengujian metode manual, metode pengujian online, umpan balik dll [14].

Kelulusan siswa merupakan hal yang diutamakan untuk mencapai prestasi di dunia pendidikan, kelulusan juga berpengaruh terhadap nilai akreditasi suatu pendidikan itu sendiri. Peneliti menggunakan KNN untuk memprediksi kinerja dari siswa SMK Swasta Anak Bangsa berdasarkan nilai kehadiran, nilai sikap dan nilai pengetahuan. Hasil yang diperoleh yaitu nilai pada $K=5$, tingkat akurasinya 93,55% dan dijadikan sebagai nilai K-Optimal [15].

3 Metodologi

Dalam penelitian ini akan dilakukan beberapa tahapan penelitian, sebagai berikut:

- a. Dataset

Data yang digunakan dalam pengolahan adalah data pada tahun 2013 dan 2014 didapat dari salah satu instansi sekolah swasta di area Bekasi dengan jumlah siswa kelas VI sebanyak 104 siswa dengan atribut nilai PAS, Nilai US Teori, Nilai US Praktek, Nilai UTS, Nilai Prilaku Siswa.

Tabel 1. Sampel data kinerja siswa kelas VI tahun 2013-2014

no	nama	jnskel	pas	usteori	uspraktek	uts	sikap	output
1	Adam Syahputra Umam	L	85	74	65	90	80	L
2	Addelyo Firza Rafandi Betha	L	77	74	64	98	80	L
3	Audra Triananda	L	74	6	66	48	70	T
4	Bunga Kireina Humaira	P	85	74	66	88	70	L
5	Cessa Kinanty Puti Heza	P	80	75	65	90	80	L
6	Davina Naja Suryantoro	P	78	74	68	98	85	L
7	Faza Alisha Ramadina	P	88	64	68	87	70	L
8	Ghinan Alhanief Mahardika Saefullah	L	81	58	65	63	70	T
9	Meidyana Salsabila Putri Wasesa	P	89	74	67	78	80	L
10	Naila Faraya Thaib	P	71	60	65	78	80	L

b. *Pre-processing*

Fitur pemilihan atribut dilakukan melalui *select attribute* dan *replace missing* untuk menghindari adanya kekosongan data dan noise. Pada langkah ini setdata dilakukan pembersihan (*cleansing*) terhadap *missing value* dan *noise*, agar ketika diolah tidak ada masalah dalam hasil. Data yang sudah dilakukan *preprocessing* diperoleh sebanyak 104 record.

Tabel 2. Atribut dan Nilai Kategori

No	Atribut	Nilai
1	Nilai PAS	<75
		=75
		>75
2	Nilai US Teori	<75
		=75
		>75
3	Nilai US Praktek	<75
		=75
		>75
4	Nilai UTS	<75
		=75
		>75

		Sangat Baik
5	Nilai Perilaku Siswa	Baik
		Kurang Baik

Tabel 3. Jumlah Siswa Dengan Kategori Nilai

Kategori	Nilai PAS	Nilai US	Nilai UTS
A	69	38	76
B	22	13	14
C	13	53	13
D	0	0	1
Total	104	104	104

Tabel 4. Data Dengan Kategori Gender

Jenis Kelamin	2014	2013	Jumlah Data
Laki-laki	25	19	44
Perempuan	38	22	60
Jumlah Data	63	41	104

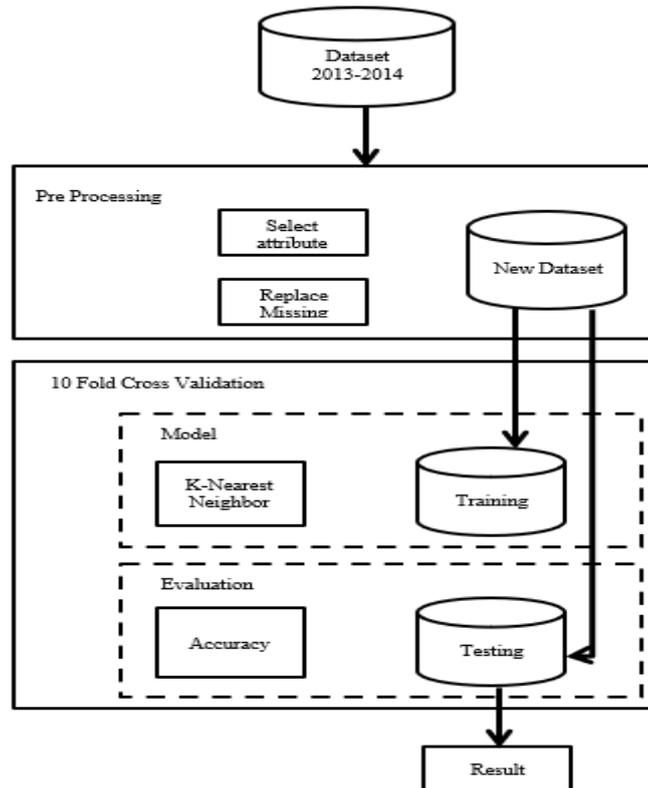
Tabel 5. Jumlah Siswa Dengan Kategori Prilaku Siswa

Kategori	Jumlah
Baik	88
Sangat Baik	0
Kurang Baik	16
Total	104

- c. Pengujian Metode dengan *10-fold Cross Validation*
 Data diuji menggunakan metode k-Nearest Neighbor. Dataset baru dibagi menjadi 2 (dua) yaitu (1) data *training*, (2) data *testing* menggunakan *10 fold cross validation*. Kemudian data *training* diklasifikasikan menggunakan *K-Nearest Neighbor* yang akan menghasilkan akurasi dalam klasifikasi tingkat kelulusan pada siswa. Pengujian menggunakan teknik *validation* dengan tujuan untuk mendapatkan nilai *accuracy*. Dataset diuji dengan metode yang diusulkan melalui aplikasi 'RapidMiner' 9.0.
- d. *Result*
 Hasil uji dari metode k-Nearest Neighbor dapat terlihat tingkat rata-rata nilai akurasi, tingkat akurasi dapat dipengaruhi oleh jumlah klastering data sehingga tujuan penerapan metode k-Nearest Neighbor untuk

memprediksi tingkat kelulusan pada siswa berdasarkan kinerja siswa dapat tercapai.

Uraian diatas merupakan penjelasan dari kerangka pemikiran dalam tahapan penelitian, sehingga dapat digambarkan dengan kerangka pemikiran sebagai berikut:



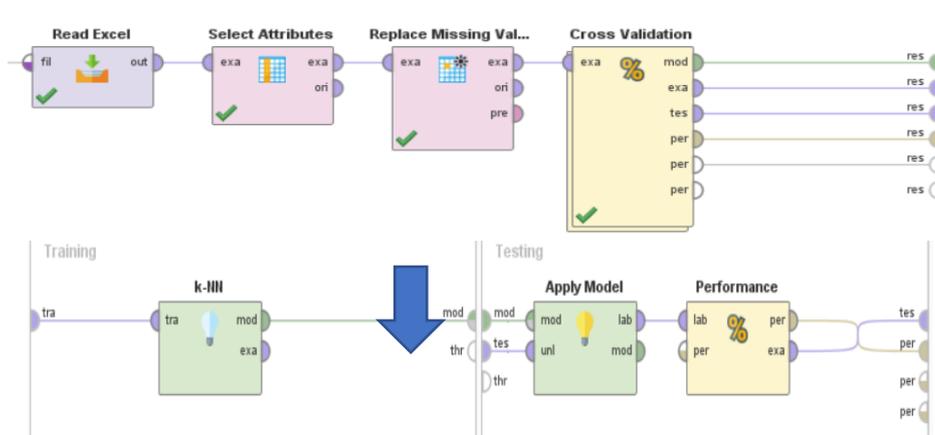
Gambar 1. Kerangka pemikiran

4 Hasil dan Pembahasan

Metode K-Nearest Neighbor (KNN) diterapkan dalam penelitian ini dengan tujuan untuk mengetahui nilai akurasi dalam kinerja siswa melalui tingkat kelulusan. Dalam mengolah data dan pengujian pada metode KNN menggunakan *RapidMiner* sebagai *software* untuk implementasinya.

Dalam penelitian ini dilakukan pengujian dengan metode K-Nearest

Neighbor (KNN), penelitian juga dilakukan percobaan sebanyak 10 kali dengan memberikan nilai k 1 sampai 10 menggunakan dataset kelulusan siswa pada kelas VI yang telah terkumpul sebanyak 104 record. Berikut adalah pengolahan datanya dengan menggunakan *tools rapidminer 9.0*. Menurut Bramer, menyuguhkan keputusan yang didapat pada data training dan testing merupakan cara kerja dari *Confusion matrix* memberikan keputusan yang diperoleh dalam *training* dan *testing* [16], dan selain itu *confusion matrix* menurut gorunescu dapat memberikan *performance* yang didasarkan pada objek yang benar atau salah [16].



Gambar 2. Pengolahan Data dengan Rapid Miner

Tabel 6. Perbandingan K-Nearest Neighbor

Nilai k	K-Terdekat
1	97.00
2	97.00
3	99.00
4	99.00
5	96.09
6	96.09
7	95.18
8	95.18
9	95.18
10	95.18
Rata-rata	96,49

Sumber : Data Olahan 2021

Hasil percobaan ini diperoleh akurasi rata-rata 96,49 dengan K-Nearest Neighbor. Hasil percobaan menunjukkan bahwa K-Nearest Neighbor dapat meningkatkan akurasi klasifikasi kinerja siswa dengan sangat baik. Dari perbandingan diatas, tingkat akurasi juga dapat dipengaruhi oleh jumlah klastering data.

5 Kesimpulan

Penelitian ini dilakukan pengujian model dengan menggunakan algoritma K-Nearest Neighbor (K-NN) dengan menggunakan data tingkat kelulusan siswa yang lulus dan tidak lulus. Model yang dihasilkan diuji untuk mendapatkan nilai akurasi dari algoritma klasifikasi data mining sehingga didapat pengujian data mahasiswa dengan rata-rata nilai akurasi pada klastering data $k=1$ sampai $k=10$ dengan menggunakan K-Nearest.

Metode *K-Nearest Neighbor* (KNN) dapat memprediksi tingkat kelulusan siswa. Metode *k-nearest neighbor* (KNN) yang diproses dengan *tools rapidminer 9.0* didapatkan rata-rata akurasi sebesar 96,49%. Tingkat akurasi pengujian model kelulusan siswa dengan menggunakan algoritma K-Nearest Neighbor (K-NN) juga dipengaruhi oleh jumlah klastering data. Sistem yang digunakan dengan *rapid miner* dapat membantu dalam menentukan *kriteria* data siswa lulus mencapai KKM dan tidak lulus mencapai KKM sehingga membantu dalam pengambilan keputusan.

6 Referensi

- [1] K. Maheswari, A. Priya, A. Balamurugan, and S. Ramkumar, "Analyzing student performance factors using KNN algorithm," *Mater. Today Proc.*, Feb. 2021.
- [2] M. Pandey and S. Taruna, "Towards the integration of multiple classifier pertaining to the Student's performance prediction," *Perspect. Sci.*, vol. 8, no. September 2016, pp. 364–366, 2016.
- [3] Y. Altujjar, W. Altamimi, I. Al-Turaiki, and M. Al-Razgan, "Predicting Critical Courses Affecting Students Performance: A Case Study," *Procedia Comput. Sci.*, vol. 82, pp. 65–71, 2016.
- [4] F. Yang and F. W. B. Li, "Study on student performance estimation, student progress analysis, and student potential prediction based on data mining," *Comput. Educ.*, vol. 123, pp. 97–108, Aug. 2018.
- [5] F. A. Sianturi, "Analisa Decision Tree Dalam Pengolahan Data Siswa,"

- MEANS (Media Inf. Anal. dan Sist.*, vol. 3, no. 2, pp. 166–172, 2018.
- [6] A. Sijabat, “Penerapan Data Mining Untuk Pengolahan Data Siswa Dengan Menggunakan Metode Decision Tree (Studi Kasus : Yayasan Perguruan Kristen Andreas),” *Maj. Ilm. Inf. dan Teknol. Ilm.*, vol. V, no. 3, pp. 7–12, 2015.
 - [7] A. R. Widiarti, “K-nearest neighbor performance for Nusantara scripts image transliteration,” *J. Teknol. dan Sist. Komput.*, vol. 8, no. 2, pp. 150–156, 2020.
 - [8] C. S. D. Prasetya, “Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor,” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 4, no. 3, p. 194, 2017.
 - [9] A. Baita, Y. Pristyanto, and I. Pratama, “Perbandingan Algoritme Klasifikasi Untuk Prediksi Kinerja Siswa Di Kelas,” *INFOS J.*, vol. 1, no. 4, pp. 1–4, 2019.
 - [10] Z. E. Rasjid and R. Setiawan, “Performance Comparison and Optimization of Text Document Classification using k-NN and Naïve Bayes Classification Techniques,” in *Procedia Computer Science*, 2017, vol. 116, pp. 107–112.
 - [11] T. Setiyorini and R. T. Asmono, “Penerapan Metode K-Nearest Neighbor Dan Information Gain Pada Klasifikasi Kinerja Siswa,” *JITK (Jurnal Ilmu Pengetah. dan Teknol. Komputer)*, vol. 5, no. 1, pp. 7–14, 2019.
 - [12] S. Wiyono and T. Abidin, “Implementation of K-Nearest Neighbour (Knn) Algorithm To Predict Student’S Performance,” *Simetris J. Tek. Mesin, Elektro dan Ilmu Komput.*, vol. 9, no. 2, pp. 873–878, 2018.
 - [13] S. Kotsiantis, K. Patriarcheas, and M. Xenos, “A combinational incremental ensemble of classifiers as a technique for predicting students’ performance in distance education,” *Knowledge-Based Syst.*, vol. 23, no. 6, pp. 529–535, Aug. 2010.
 - [14] M. Mayilvaganan and D. Kalpanadevi, “Comparison of classification techniques for predicting the performance of students academic environment,” 2014.
 - [15] S. R. Rani, S. R. Andani, and D. Suhendro, “Penerapan Algoritma K-Nearest Neighbor untuk Prediksi Kelulusan Siswa pada SMK Anak Bangsa,” *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 670, 2019.
 - [16] E. Purwaningsih, “Seleksi Mobil Berdasarkan Fitur dengan Komparasi Metode Klasifikasi Neural Network, Support Vector Machine, dan Algoritma C4.5,” *J. Pilar Nusa Mandiri*, vol. XII, no. 2, pp. 153–160, 2016.