

Sistem Rekomendasi Pemilihan Kerja Siswa SMK Dengan Menggunakan Algoritma Decision Tree

Muhamad Faisal Halim¹, Much. Rifqi Maulana², Hermanus Wim Hapsoro³,
Arochman⁴, Agus Ilyas⁵

^{1,2,3,4,5}Teknik Informatika STMIK Widya Pratama

Email : ¹ffaisalhaalim@gmail.com, ²rifqi@stmik-wp.ac.id, ³wimhapsoro06@gmail.com,
⁴arochman.aryanta@gmail.com, ⁵ilyasagus@gmail.com

Abstrak

Banyaknya lulusan SMK setiap tahunnya menjadi salah satu pondasi untuk memajukan bangsa ini, namun hal tersebut tidak selalu diikuti dengan ketersediaan lapangan pekerjaan yang memadai, oleh karena itu Banyak lulusan SMK yang menganggur dan sulit memasuki dunia kerja. Peran bursa kerja khusus (BKK) pada sekolah sangat penting, karena berhubungan langsung dengan pemetaan kompetensi siswa dan perusahaan dengan lowongan pekerjaan baru. Namun sering kali kinerja pada sistem BKK tidak berjalan dengan baik dan optimal, dan susahnya melakukan pemetaan kompetensi siswa yang masih dilakukan dengan manual, dengan demikian obyektivitas hasilnya rendah. Sistem rekomendasi perusahaan untuk lulusan SMK menjadi solusi yang dapat menunjang hal-hal tersebut. Dengan memanfaatkan data lulusan sebelumnya dengan menggunakan variable nilai akademis matematika raport, rata-rata raport, rata-rata us, ujian praktik kejuruan sebagai fitur prediktor, pemetaan kompetensi siswa untuk disalurkan pada perusahaan dapat dilakukan dengan mudah dan obyektivitas menjadi lebih tinggi. Dengan memanfaatkan algoritma decision tree untuk menentukan hasil rekomendasi dan CRISP-DM sebagai metode pengembangan sistem diperoleh evaluasi metrik akurasi sebesar 61%. Maka dengan adanya sistem rekomendasi ini proses dalam menyeleksi lulusan untuk disalurkan ke perusahaan tertentu dapat dilakukan dengan mudah dan lebih efisien dan hasil dari rekomendasi dapat digunakan sebagai pengambilan keputusan kedepannya nanti.

Kata Kunci : Sistem Rekomendasi, Decision Tree, CRISP-DM

1. PENDAHULUAN

Sekolah Menengah Kejuruan atau SMK adalah salah satu jenjang pendidikan menengah dengan kekhususan mempersiapkan lulusannya untuk siap bekerja [1], SMK sebagai subsistem pendidikan nasional seharusnya mengutamakan mempersiapkan peserta didiknya Untuk mampu memilih karir, memasuki lapangan kerja, berkompetisi, dan mengembangkan diri dengan sukses di lapangan kerja yang cepat berkembang, lulusan SMK dituntut untuk memiliki keterampilan teknis (hard skills) dan keterampilan sosial (soft skills) yang sesuai dengan tuntutan pekerjaan. Perubahan dunia kerja yang terus berkembang ini menuntut lulusan SMK untuk memiliki kemampuan yang kompetitif dan sesuai dengan kebutuhan pasar kerja saat ini [1]. Namun, setiap tahunnya jumlah lulusan SMK yang meningkat ternyata tidak selalu diikuti dengan ketersediaan lapangan kerja yang memadai, Kondisi ini tidak sesuai dengan PP No. 17 tahun 2010 yang menyatakan bahwa tujuan pendidikan kejuruan adalah untuk menciptakan lulusan yang siap bekerja [2]. Akibatnya, banyak lulusan yang menganggur dan sulit untuk memasuki dunia kerja. Menurut Data Kemendikbud, jumlah lulusan SMK di Indonesia meningkat dari 2016 hingga 2020. Pada tahun 2016, terdapat 4.222 siswa lulus; pada tahun 2020, terdapat 6.298 siswa. Hal ini dapat meningkatkan jumlah pengangguran lulusan SMK jika lapangan pekerjaan yang ada belum sebanyak dan belum dapat meningkatkan jumlahnya [2].

SMK Gondang wonopringgo merupakan sala satu subsistem pendidikan menengah yang setiap tahunnya meluluskan banyak siswa, dari data yang diperoleh sebagian hanya sebagian kecil dari lulusannya dapat disalurkan untuk langsung bekerja oleh BKK SMK Gondang. BKK adalah sebuah unit pelaksana yang dibentuk di SMK Negeri dan Swasta untuk memberikan pelayanan dan informasi terkait lowongan kerja [3]. Performa dari BKK menjadi salah satu faktor penentu keberhasilan lulusan SMK dalam terserap di dunia industri sesuai dengan bidang keahliannya [4]. Namun kenyataannya, pelaksanaan BKK tidak optimal dalam menjalankan peran yang menjadi tanggung jawabnya. Hasil wawancara dengan BKK SMK Gondang diketahui bahwa BKK seringkali menghadapi beberapa permasalahan, antara lain: (1) sistem BKK yang berjalan belum maksimal, (2) pemetaan keahlian siswa masih dilakukan secara manual.

Dari permasalahan yang dihadapi bahwa dalam melakukan pemetaan keahlian siswa, yang masih dilakukan manual, dimana hal tersebut berkemungkinan dapat menurunkan tingkat objektivitas, maka diperlukannya sistem rekomendasi perusahaan yang dibuat dengan memanfaatkan algoritma *decision tree* dan diharapkan bisa membantu dalam proses penyaringan dan pemetaan keahlian lulusan SMK Gondang agar dapat dilakukan dengan lebih mudah dan objektivitas menjadi lebih tinggi. Pada penelitian ini algoritma *decision tree* dengan spesifikasi algoritma CART sebagai algoritma klasifikasi akan digunakan untuk membangun sistem yang dimaksud. Tugas-tugas data mining seperti klasifikasi sering digunakan untuk meramalkan nama kelas kategorikal atau diskrit untuk kasus-kasus baru berdasarkan data historis [5].

2. METODE PENELITIAN

Dalam melakukan penelitian ini, digunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai sebuah model proses yang sistematis. Model proses CRISP-DM mencakup enam tahapan utama yaitu [6] :



Gambar 1 CRISP-DM Diagram

1. Tahap Pemahaman Bisnis (*Business Understanding*)
Tujuan dari tahapan ini adalah mengetahui tujuan apa yang akan dicapai dengan dibuatnya sistem rekomendasi perusahaan yang memanfaatkan *decision tree* untuk mengklasifikasikan data baru berdasarkan label atau nama kelas tujuannya, yang didasari pada data lama sebagai acuannya.
2. Tahap Pemahaman Data (*Data Understanding*)
Tahapan pemahaman data adalah proses mengumpulkan data lulusan terdahulu di SMK Gondang yang dilakukan pada bulan Oktober 2022.
3. Tahap Pengolahan Data (*Data Preparation*)
Proses pengolahan data adalah proses membentuk data final yang siap digunakan pada tahapan pemodelan [7]. Salah satu proses yang dilakukan adalah membersihkan data dari

kesalahan atau kecacatan yang terdapat pada data, yang sebelumnya berjumlah 125 data menjadi 110 data.

4. Tahap Pemodelan (*Modeling*)

Pemodelan yang dilakukan menggunakan algoritma *CART (Classification and Regression Tree)* pada *decision tree* untuk melakukan klasifikasi, Proses penghitungan *CART Algorithm* melibatkan pembentukan pohon keputusan dari data latih dengan menghitung tingkat ketidakmurnian (*impurity*) dari setiap variabel prediktor. Tingkat ketidakmurnian dihitung dengan menggunakan beberapa metode, seperti Indeks Gini atau Entropi [8]. Pada setiap simpul keputusan (*decision node*), algoritma *CART* akan memilih variabel prediktor dengan tingkat ketidakmurnian terendah untuk memecah data menjadi dua kelompok yang lebih kecil [9]. Untuk mendapatkan nilai *impurity* dari setiap variabel yang digunakan dapat menggunakan rumus [10]:

$$\text{Gini :} \quad H(Q_m) = \sum_k P_{mk} (1 - P_{mk}) \quad 1$$

$$\text{Entropy :} \quad H(Q_m) = \sum_k P_{mk} \log(P_{mk}) \quad 2$$

5. Tahap Evaluasi (*Evaluation*)

Dalam rangka meningkatkan kualitas dan efektivitas pemodelan sebelum dilakukan *deployment*, digunakanlah metode evaluasi untuk mengukur kinerja sistem yang digunakan. Untuk mengukur kinerja sistem klasifikasi, dapat digunakan matriks akurasi (*confusion matrix*). Formula yang digunakan untuk menghitung akurasi adalah sebagai berikut [11]:

$$\text{Akurasi :} \quad = \frac{\text{akurasi}}{\text{jumlah data prediksi benar}} = \frac{\text{jumlah data prediksi benar}}{\text{jumlah prediksi yang dilakukan}} \quad 3$$

6. Tahap Penyebaran (*Deployment*)

Tahapan ini adalah tahapan dimana dilakukannya penyebaran dari model data mining yang sudah dibuat, tahap penyebaran bisa dilakukan dengan mengintegrasikan model dengan program yang sudah ada, atau membuat program tersendiri berdasarkan model yang sudah dibuat.

3. HASIL DAN PEMBAHASAN

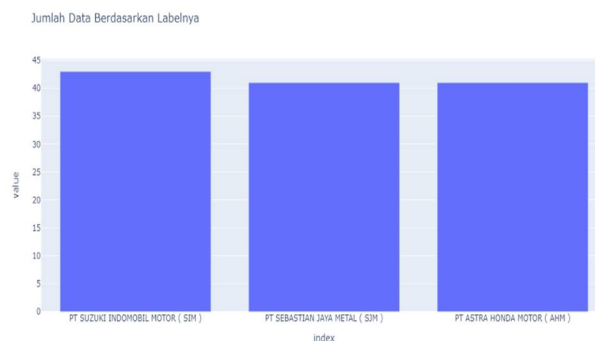
SMK Gondang wonopringgo merupakan sala satu subsistem pendidikan menengah yang setiap tahunnya meluluskan banyak siswa, dari data yang diperoleh sebagian hanya sebagian kecil dari lulusannya dapat disalurkan untuk langsung bekerja oleh BKK SMK Gondang, Hasil wawancara dengan BKK SMK Gondang diketahui bahwa BKK seringkali menghadapi beberapa permasalahan, antara lain: (1) sistem BKK yang berjalan belum maksimal, (2) pemetaan keahlian siswa masih dilakukan secara manual. Oleh karena itu dengan dibuatnya sistem rekomendasi ini diharapkan diharapkan bisa membantu dalam proses penyaringan dan pemetaan keahlian lulusan SMK Gondang agar dapat dilakukan dengan lebih mudah dan objektivitas menjadi lebih tinggi.

Dalam penelitian ini data yang digunakan merupakan data lulusan SMK Gondang tahun ajaran 2019/2020 dan 2020/2021, dimana data tersebut memiliki 7 fitur dengan 4 fitur prediktor dan 1 fitur target. Data yang dipakai memiliki 3 label yang berupa nama perusahaan tempat lulusan sudah bekerja, yaitu : PT. Astra Honda Motor (AHM), PT. Sebastian Jaya Metal (SJM), dan PT. Suzuki Indomobil Motor (SIM), sesuai dengan table 1 berikut :

Tabel 1 data lulusan

No.	Nama	matematika raport	rata-rata raport	rata- rata us	ujian praktik kejuruan	perusahaan
1	Abdurrahman asyari	77.1	79.2	70.2	84.0	SJM
2	Adam mubarak	77.2	78.8	77.3	78.5	SIM
3	Adhim setiawan	76.2	82.5	75.9	94.8	AHM
4	Aditya ardiansyah	77.5	79.2	74.6	85.0	SJM
5	Aditya suryana	78.5	81.4	74.8	86.3	SIM
...	...	...	...	...	...	...
125	Zaki umar ibrahim	76.7	80.2	72.3	84.3	SIM

Dari data yang digunakan dapat diambil beberapa informasi sebagai berikut, informasi ini diperoleh akan ditangani untuk bisa merubah data mentah sebelumnya menjadi dataset yang siap untuk dilakukan pemodelan. Proses mendapatkan informasi dari data dilakukan dengan bantuan Bahasa pemrograman python dan beberapa library pendukung lainnya.



Gambar 2 Jumlah data berdasarkan label

Gambar 1 merupakan visualisasi dari jumlah data berdasarkan labelnya, hal ini perlu diperhatikan, karena jika terjadi ketidak seimbangan jumlah data, model yang dibuat memiliki kemungkinan tinggi tidak dapat memahami pola data secara maksimal. Beberapa informasi data lainnya juga perlu diketahui untuk menambah obyektivitas pada model adalah jumlah missing value pada data, deteksi outliers dan lainnya.

```
df.isnull().sum()

# Output
# no                0
# nama              0
# matematika raport 0
# rata-rata raport  0
# rata-rata us      0
# ujian praktik kejuruan 0
# perusahaan        0
# dtype: int64
```

Gambar 3 Deteksi Missing Value

Dari Gambar 2 dapat diambil kesimpulan bahwa data yang digunakan tidak memiliki missing value, data dengan missing value akan mengganggu proses pelatihan model data mining yang akan dibuat [8]. Untuk mengatasi missing value dapat dilakukan dalam beberapa cara, namun cara yang umum dilakukan adalah dengan menghapus data tersebut [5], dengan demikian pastinya akan terdapat pengurangan jumlah dataset.



Gambar 4 Deteksi Outliers

Deteksi outliers dapat dilakukan dengan menggunakan visualisasi seperti Gambar 3, titik-titik yang berada diluar box merupakan data outliers yang terdapat dari dataset yang digunakan. outliers atau pencilan adalah data yang jauh dari nilai tengah (mean) dari kumpulan data atau data yang berbeda jauh dengan mayoritas data lainnya [5]. Outlier dapat memengaruhi analisis dan peramalan statistik karena menyebabkan perkiraan yang salah dan pemodelan yang buruk.

```
print("Jumlah Dataset : {}".format(df.shape[0]))
print("Data Outliers Fix : {}".format(df_new.shape[0]))

# OutPut
# Jumlah Dataset : 125
# Data Outliers Fix : 110
```

Gambar 5 Perbandingan jumlah data

Mengatasi outliers menggunakan pendekatan IQR (Interquartile Range) akan mempengaruhi jumlah dataset, metode ini akan memilah dan menghapus data yang menjadi outliers dari total keseluruhan data yang dimiliki. Pada Gambar 4 terlihat perbedaan terhadap jumlah data yang sebelumnya berjumlah 125 data, dan setelah dilakukan penghapusan outliers menjadi 110 data. Sebelum melakukan pemodelan langkah-langkah diatas diperlukan untuk

membuat dataset yang berbasis hambatan Ketika dilakukan pemodelan, termasuk dengan melakukan normalisasi data. Proses normalisasi data dikatakan dapat mempermudah model dalam memahami pola dari data [8].

Proses normalisasi data dilakukan terhadap data dengan tipe numerik, yang akan dirubah dengan menggunakan metode MinMax Scaller dengan hasil akhir data menjadi skala antara 0 sampai 1.

Tabel 2 Data Sebelum Normalisasi

Matematika Raport	Rata-rata Raport	Rata-rata US	Praktik Kejuruan
77.1	79.2	70.2	84.0
77.2	78.8	77.3	78.5
76.2	82.5	75.9	94.8
77.5	79.2	74.6	85.0
78.5	81.4	74.8	86.3

Tabel 3 Data Setelah Normalisasi

Matematika Raport	Rata-rata Raport	Rata-rata US	Praktik Kejuruan
0.119205	0.152941	0.180451	0.358382
0.125828	0.105882	0.714286	0.040462
0.059603	0.541176	0.609023	0.982659
0.145695	0.152941	0.511278	0.416185
0.211921	0.411765	0.526316	0.491329

Pembuatan model dan proses melatih model akan dilakukan setelah data dirasa sudah optimal, proses pemodelan dan evaluasi sering kali dilakukan bersamaan. Proses pembuatan model pada penelitian ini dibantu dengan menggunakan pustaka *sklearn* pada *python*, Pustaka ini memanfaatkan algoritma *Decision Tree CART* untuk membuat model klasifikasi.

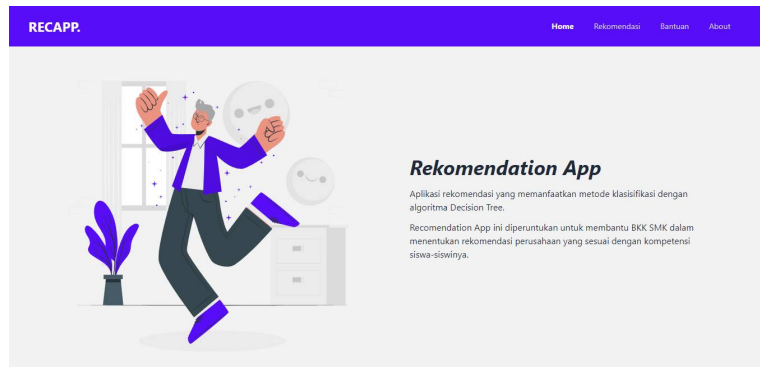
Metrik evaluasi menjadi salah satu penanda dalam mengetahui apakah model dan data yang digunakan sudah optimal atau belum, dalam evaluasi model dapat dilakukan dengan beberapa cara, salah satunya adalah mengetahui tingkat akurasi suatu model. Gambar 5 berikut ini menunjukkan berbagai metrik evaluasi yang diperoleh dari model klasifikasi yang dibuat.

	precision	recall	f1-score	support
PT ASTRA HONDA MOTOR (AHM)	0.78	0.70	0.74	10
PT SEBASTIAN JAYA METAL (SJM)	0.71	0.45	0.56	11
PT SUZUKI INDOMOBIL MOTOR (SIM)	0.71	1.00	0.83	12
accuracy			0.73	33
macro avg	0.73	0.72	0.71	33
weighted avg	0.73	0.73	0.71	33

Gambar 6 Metrik Evaluasi

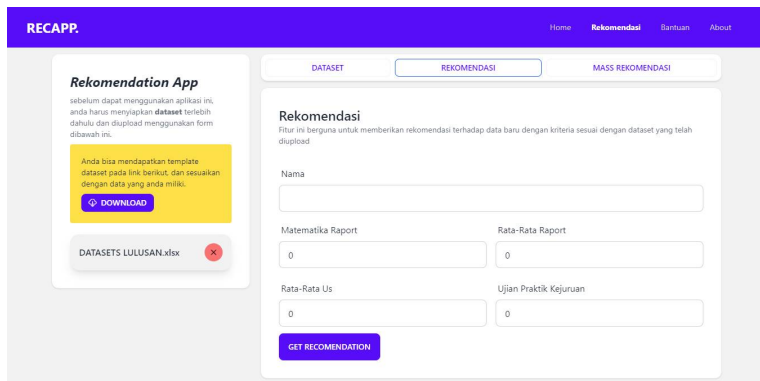
Pada Gambar 5 diketahui bahwa metrik akurasi yang diperoleh dari model yang dibuat adalah sebesar 0.73 atau 73%, tinggi rendah suatu nilai metrik dapat dipengaruhi oleh banyak faktor. Setelah model sudah selesai pengintegrasian model terhadap program lain diperlukan agar model dapat digunakan sebagai fitur untuk pengguna. Proses penyebaran atau deployment model data dapat dilakukan dalam beberapa cara, cara yang digunakan dalam penelitian ini adalah dengan membuat program berdasarkan model yang dibuat yaitu program sistem

rekomendasi. Dengan memanfaatkan framework untuk mempermudah pembuatan, aplikasi sistem rekomendasi ini dapat dibangun.



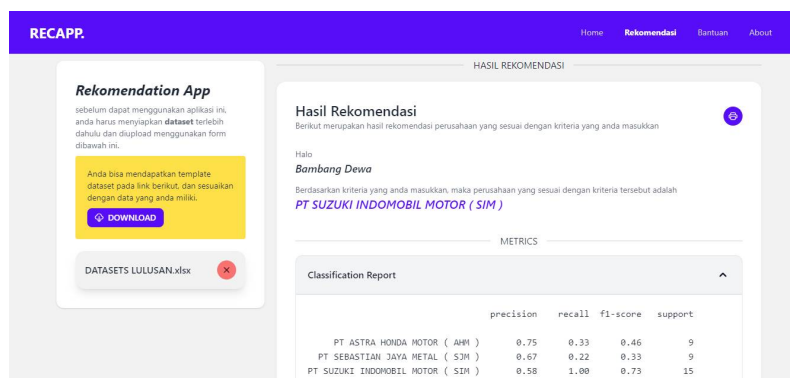
Gambar 7 Halaman Utama Aplikasi

Pengguna dapat melakukan perhitungan rekomendasi dengan masuk pada halaman rekomendasi dengan cara klik pada menu rekomendasi. Dengan mengisi inputan yang tersedia sebagai syarat perhitungan seperti pada gambar 7 Berikut ini.



Gambar 8 Halaman Rekomendasi

Dengan mengisi inputan yang ada yaitu nama dan beberapa nilai, inputan tersebut akan dikalkulasi untuk bisa mendapatkan hasil dari rekomendasi. Selanjutnya aplikasi akan menghasilkan hasil sebagai berikut.



Gambar 9 Hasil Rekomendasi

Metrik model yang dibuat menghasilkan akurasi 61%, dan dapat disimpulkan bahwa model yang dibuat masih belum terlalu baik dalam memahami pola data yang digunakan, hal ini dapat disebabkan oleh banyak faktor, misalnya : data yang tidak cocok, data yang kurang, atau bahkan kesalahan dalam tahapan data preparation.

Jika diperhatikan dengan seksama pada kelas PT SUZUKI INDO MOOR (SIM) memiliki nilai precision yang paling kecil dari yang lain, artinya dalam memahami data dengan kelas tersebut model tidak terlalu optimal jika dibandingkan dengan model saat memahami pola dari kelas yang lain, dengan begitu semoga kedepannya dengan perubahan pada beberapa faktor hal tersebut dapat teratasi.

4. KESIMPULAN

Pemodelan pada aplikasi system rekomendasi yang dibuat menggunakan algoritma Decision Tree CART dapat menghasilkan prediksi dengan melihat pola dari dataset lulusan tahun sebelumnya dengan menggunakan nilai akademis sebagai kriteria. Hasil dari pemodelan menunjukkan bahwa sistem rekomendasi yang dibuat dapat memberikan prediksi dengan akurasi sebesar 61% untuk kalkulasi data baru. Sehingga dapat disimpulkan bahwa sistem rekomendasi yang dibuat mampu memberikan rekomendasi perusahaan yang sesuai dengan kriteria yang diperoleh siswa selama sekolah.

Namun, perlu diingat bahwa sistem yang dibuat hanya dapat memberikan rekomendasi, Meskipun demikian, sistem rekomendasi ini dapat menjadi salah satu solusi untuk mengatasi kesulitan dalam memilih karir bagi lulusan SMK, sehingga dapat membantu mereka memasuki lapangan kerja, berkompetisi, dan mengembangkan diri dengan sukses di lapangan kerja yang cepat berkembang.

DAFTAR PUSTAKA

- [1] A. T. Turistiati, H. Faikar, dan A. Ramadhan, "Pelatihan Soft Skills Dan Pendampingan Siswa-Siswi SMK Di Kota Bogor Untuk Persiapan Memasuki Dunia Kerja," *Jurnal Komunitas : Jurnal Pengabdian kepada Masyarakat*, vol. 1, no. 2, 2019, [Daring]. Tersedia pada: <http://ojs.stiami.ac.id>
- [2] R. B. Miko, O. Wijaya, dan E. Diah Utami, "Determinan Pengangguran Lulusan SMK di Indonesia Tahun 2020," *Seminar Nasional Official Statistics*, vol. 2021, no. 1, hlm. 801–810, Nov 2021, doi: 10.34123/SEMNASOFFSTAT.V2021I1.1048.
- [3] A. Amalina dan Y. D. Putri, "Pengembangan Sistem Informasi Berbasis Web untuk Peningkatan Kinerja Unit Bursa Kerja Khusus SMK Negeri 1 Tanjung Raya," *Khazanah Informatika : Jurnal Ilmu Komputer dan Informatika*, vol. 3, no. 2, hlm. 73–79, Des 2017, doi: 10.23917/KHIF.V3I2.4910.
- [4] N. Adhi, Y. Pambayun, P. Teknologi, K. Universitas, dan N. Yogyakarta, "KINERJA BURSA KERJA KHUSUS (BKK) SMK NEGERI BIDANG KEAHLIAN TEKNOLOGI DAN REKAYASA DI KABUPATEN SLEMAN," *Jurnal Pendidikan Vokasi*, vol. 4, no. 2, 2014.
- [5] J. Han, M. Kamber, dan J. Pei, *Data Mining: Concepts and Techniques*. 2012. doi: 10.1016/C2009-0-61819-5.
- [6] D. T. Larose, *Discovering Knowledge in Data: An Introduction to Data Mining*. 2005. doi: 10.1002/0471687545.
- [7] A. Azevedo dan M. F. Santos, "KDD, SEMMA and CRISP-DM: a parallel overview," *IADS - DM*, 2008, Diakses: 23 Februari 2023. [Daring]. Tersedia pada: <https://recipp.ipp.pt/handle/10400.22/136>
- [8] I. Saputra dan D. A. Kristiyanti, *Machine Learning Untuk Pemula*. Bandung: Informatika Bandung, 2022.
- [9] L. Breiman, J. H. Friedman, R. A. Olshen, dan C. J. Stone, *Classification and regression trees*. CRC Press, 2017. doi: 10.1201/9781315139470/CLASSIFICATION-REGRESSION-TREES-LEO-BREIMAN.
- [10] "1.10. Decision Trees — scikit-learn 1.2.1 documentation." <https://scikit-learn.org/stable/modules/tree.html> (diakses 24 Februari 2023).

- [11] A. Soma Darmawan, D. Sugianti, M. Faisal Halim, T. Informatika, dan S. Widya Pratama, "Case Based Reasoning Untuk Mendeteksi Penyakit Kucing Dengan Metode KNN," *Smart Comp :Jurnalnya Orang Pintar Komputer*, vol. 11, no. 2, hlm. 109–115, Apr 2022, doi: 10.30591/SMARTCOMP.V11I2.3452.