



Comparative Analysis of C.45 Algorithm and Support Vector Machine (SVM) for Student Department Classification

Analisis Perbandingan Algoritma C.45 dan Support Vector Machine (SVM) untuk Klasifikasi Jurusan Siswa

**Qurotul A'yuniyah¹, Septi Kenia Pita Loka²,
Ena Tasia³, Nanda Nazira⁴**

^{1,2,3,4}Program Studi Sistem Informasi, Fakultas Sains dan Teknologi,
Universitas Islam Negeri Sultan Syarif Kasim Riau

E-Mail: ¹qurotulyuyun510@gmail.com, ²12050323230@students.uin-suska.ac.id,
³enatasia6696@gmail.com, ⁴nandanazira0211@gmail.com

Corresponding Author: Qurotul A'yuniyah

Abstract

Specialization in majors is a decision made by new students to choose the "IPA" and "IPS" majors that exist in students. One of the high schools that uses this major system is SMA Negeri 15 Pekanbaru with the available majors, namely Science and Social Sciences. A system like this is very effective if the student has previously had preparation and knowledge about the major to be chosen. But on the contrary, if the student does not know the direction of ability regarding the major to be chosen, the student will be confused in choosing a major which can result in taking the wrong major. This study aims to classify the majors of SMAN 15 Pekanbaru students by comparing the C4.5 and SVM algorithms. So that the best accuracy value for this data is produced, namely the SVM algorithm of 65% and followed by the C4.5 algorithm of 56%.

Keywords: C4.5, Classification, Data Mining, Student Majors, SVM.

Abstrak

Peminatan jurusan merupakan suatu keputusan yang dilakukan oleh peserta didik baru untuk memilih jurusan "IPA" dan "IPS" yang ada pada siswa. Salah satu sekolah menengah atas yang menggunakan sistem penjurusan ini yaitu SMA Negeri 15 Pekanbaru dengan jurusan yang tersedia yaitu jurusan IPA dan IPS. Sistem seperti ini sangat efektif dilakukan, jika sebelumnya siswa tersebut telah memiliki persiapan dan pengetahuan mengenai jurusan yang akan dipilih. Namun sebaliknya, jika siswa tersebut belum mengetahui arah kemampuan mengenai jurusan yang akan dipilih maka siswa tersebut akan kebingungan dalam memilih jurusan dan bisa berakibat pada salah ambil jurusan. Penelitian ini bertujuan untuk mengklasifikasi jurusan siswa SMAN 15 Pekanbaru dengan membandingkan algoritma C4.5 dan SVM. Sehingga dihasilkan nilai akurasi terbaik untuk data ini yaitu algoritma SVM sebesar 65% dan disusul algoritma C4.5 sebesar 56%.

Kata Kunci: C4.5, Data Mining, Klasifikasi, Penjurusan Siswa, SVM.

1. PENDAHULUAN

Berdasarkan Undang-Undang Republik Indonesia Nomor 20 Tahun 2003 tentang Sistem Pendidikan Nasional Pasal 1 Ayat 1 menyatakan bahwa Pendidikan adalah usaha dan terencana untuk mewujudkan suasana belajar dan proses pembelajaran agar peserta didik secara aktif mengembangkan potensi dirinya untuk memiliki kekuatan spiritual keagamaan, pengendalian diri, kepribadian kecerdasan, akhlak mulia, serta keterampilan yang diperlukan dirinya, masyarakat, bangsa, dan negara. Pentingnya pendidikan di Indonesia merupakan salah satu upaya untuk membantu meningkatkan mutu jiwa anak-anak didik baik secara lahir ataupun batin, salah satunya pada pendidikan tingkat Sekolah Menengah Atas (SMA) [1].

Sekolah Menengah Atas (disingkat SMA; bahasa Inggris: *Senior High School* atau *High School*), adalah jenjang pendidikan menengah pada pendidikan formal di Indonesia setelah lulus Sekolah Menengah Pertama (atau sederajat). Sekolah menengah atas ditempuh dalam waktu 3 tahun, mulai dari kelas 10 sampai kelas 12 [2]. Sejak tahun ajaran 2016/2017, penjurusan jenjang SMA dan setingkatnya dimulai sejak siswa masuk yaitu mulai dari kelas X. jurusan yang tidak tepat bisa saja merugikan siswa tersebut dan juga karirnya di masa yang akan datang [3].

Peminatan jurusan merupakan suatu keputusan yang dilakukan oleh peserta didik baru untuk memilih jurusan "IPA" dan "IPS" yang ada pada siswa. Sebaliknya, kesempatan yang sangat baik bagi siswa akan hilang karena kurang tepatnya dalam penentuan peminatan. Kurikulum yang digunakan pada sekolah SMA Negeri 15 Pekanbaru saat ini adalah kurikulum yang digunakan yaitu 2013. Ditahun sebelumnya pemilihan jurusan baru dilakukan pada saat kenaikan atau saat kenaikan kelas XI. Tentu saja hal ini bisa mengakibatkan sering terjadinya kesalahan dalam memilih jurusan oleh calon siswa SMA [4]. Pendidikan kejuruan adalah bagian dari sistem pendidikan yang mempersiapkan seseorang agar lebih mampu bekerja pada suatu kelompok pekerjaan atau satu bidang pekerjaan dari pada bidang-bidang pekerjaan lainnya. Dengan pengertian bahwa setiap bidang studi adalah pendidikan kejuruan sepanjang bidang studi tersebut dipelajari lebih mendalam dan kedalaman tersebut dimaksudkan sebagai bekal memasuki dunia kerja. Maka setiap tahun ajaran baru, siswa yang ingin mendaftar di SMA mengajukan pilihan jurusan yang dipilih selama belajar di SMA tersebut [5].

Salah satu sekolah menengah atas yang menggunakan sistem penjurusan ini yaitu SMA Negeri 15 Pekanbaru dengan jurusan yang tersedia yaitu jurusan IPA dan IPS. Saat ini, pemilihan jurusan di SMA Negeri 15 Pekanbaru dilakukan pada saat calon siswa dinyatakan lulus masuk di sekolah ini. Sistem seperti ini sangat efektif dilakukan, jika sebelumnya siswa tersebut telah memiliki persiapan dan pengetahuan mengenai jurusan yang akan dipilih. Namun sebaliknya, jika siswa tersebut belum mengetahui arah kemampuan mengenai jurusan yang akan dipilih maka siswa tersebut akan kebingungan dalam memilih jurusan dan bisa berakibat pada salah ambil jurusan.

Menurut *Educational Psychologist dari Integrity Development Flexibility (IDF)*, Irene Guntur, sebanyak 87 persen mahasiswa di Indonesia salah jurusan. Salah jurusan memicu pengangguran. Rata-rata pertumbuhan angkatan kerja sekitar 2 juta orang pertahun. Dari jumlah tersebut, 63% angkatan kerja bekerja tidak sesuai dengan kualifikasi. Ketidaksihinggaan pekerjaan dengan latar belakang pendidikan dan kemampuan tenaga kerja di bawah kualifikasi menimbulkan penurunan produktivitas, kepuasan kerja, dan rendahnya jenjang karier [6]. Dengan banyaknya jumlah pelajar yang mengalami salah jurusan, maka dibutuhkan klasifikasi untuk memprediksi jurusan siswa menggunakan pengolahan data yaitu data mining.

Penelitian yang dilakukan oleh Rina Kurniasari dan Azizah Fatmawati (2019) mengenai Penjurusan Siswa Sekolah Menengah Atas menggunakan algoritma C.45, hasil dari penelitian ini sendiri berupa aplikasi data mining dengan Algoritma C4.5 untuk memprediksi penjurusan IPA, IPS, atau Bahasa. Tingkat akurasi yang didapatkan sebesar 97,42 % [10]. Penelitian lain mengenai algoritma Super Vector Machine (SVM) dilakukan oleh Putri Kurnia Handayani (2021) mengenai Analisis klasifikasi Parkinson' Dataset Menggunakan Algoritma (SVM). Dari penelitian tersebut didapat hasil Algoritma SVM berhasil mengenali pola klasifikasi pada Parkinson's Disease Dataset dengan metode pendekatan PPE. Pendekatan PPE ini mampu meningkatkan kinerja algoritma SVM dalam analisis pola klasifikasi, dibandingkan dengan pendekatan tradisional dan nonstandard dengan akurasi sebesar 95% [11].

Penelitian mengenai komparasi algoritma C.45 dan *Super Vector Machine (SVM)* yang dilakukan oleh Krno dan Said (2019) mengenai Perbandingan Algoritma *Decision Tree* dengan *Support Vector Machine* pada konsentrasi mahasiswa informatika, Hasil percobaan menunjukkan metode pohon keputusan data yaitu sebesar 92,50% dengan perbandingan 60 data latih yang dibandingkan dengan data vektor mesin yaitu sebesar akurasi 92,56% [12]. Penelitian lain mengenai perbandingan kedua algoritma tersebut pernah dilakukan oleh Rizky Ade Putranot,dkk (2015), mengenai Penentuan Jurusan pada Siswa SMA Menggunakan perbandingan algoritma C.45 dan SVM. Hasil penelitian tersebut didapat Tingkat akurasi dan laju error lebih baik pada decision tree dimana nilai total akurasi sebesar 88,57% dan error 11,43% untuk klasifikasi decision tree dan total akurasi sebesar 87,14% dan error 12,86% untuk klasifikasi SVM [13].

Berdasarkan permasalahan yang sudah dijelaskan pada pendahuluan maka dibuatlah suatu pengelompokan terhadap data jurusan di SMA Negeri 15 Pekanbaru. Oleh karena itu pada penelitian ini diangkatlah sebuah judul mengenai analisis perbandingan Algoritma C.45 dan *Support Vector Machine (SVM)* untuk Klasifikasi Jurusan Siswa di SMA Negeri 15 Pekanbaru.

2. METODOLOGI PENELITIAN

Pada penelitian ini akan melakukan beberapa tahapan seperti yang terlihat pada Gambar 1 yaitu langkah-langkah dalam menyelesaikan penelitian dengan menerapkan algoritma C4.5 dan *Support Vector Machine (SVM)*.

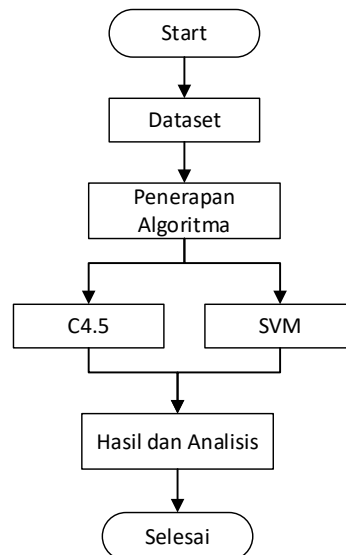


Figure 1. Metodologi Penelitian

2.1 Data Mining

Data mining adalah proses untuk memperoleh informasi yang berguna dari *database* yang besar [14]. Berdasarkan tugasnya, *Data Mining* dibagi menjadi 6 kelompok yaitu *Description*, *Estimation*, *Prediksi*, *Klasifikasi*, dan *Clustering* [15].

2.2 Klasifikasi

Klasifikasi merupakan sebuah proses untuk menemukan model yang menggambarkan dan membedakan kelas dari konsep data. Model ini pakai untuk memprediksi label kelas objek yang label kelasnya belum diketahui.

2.3 Algoritma Support Vector Machine (SVM)

SVM merupakan model baru pada teori pelatihan statistik yang diterapkan untuk memprediksi berbeda parameter [16]. Tujuan utama dari algoritma ini untuk mendapatkan fungsi yang membangun bidang hiper atau batas. Bidang *hiper* ini digunakan untuk memisahkan berbagai kategori titik data masukan [17].

2.4 Algoritma C4.5

Algoritma C4.5 memiliki input berupa sampel pelatihan dan sampel, sampel pelatihan berupa data sampel yang akan digunakan untuk membangun pohon yang sudah diuji kebenarannya, sedangkan sample adalah data *field* yang nantinya akan kita buat digunakan sebagai parameter dalam klasifikasi data (Titin Handayani Sinaga, dkk, 2020). Algoritma C4.5 menggunakan kriteria gain untuk menentukan fitur yang merupakan pemecah simpul di pohon yang diinduksi [18].

2.5 Decission Tree

Decision tree merupakan teknik berbasis pohon di mana setiap jalan mulai dari root dijelaskan oleh pemisahan data urutan sampel sampai hasil Boolean pada simpul tercapai [19]. Decision tree memiliki alur sederhana yang memilih label kelas variabel *output* menggunakan nilai dari satu atau lebih *input* variabel [20].

3. HASIL DAN ANALISIS

3.1 Pengumpulan Data

Pada penelitian ini data yang digunakan merupakan data jurusan siswa yang diambil langsung dari SMA Negeri 15 Pekanbaru. Terdapat 9 atribut yang digunakan pada data jurusan siswa yaitu Nama, Jenis Kelamin, B.Indonesia, B.inggris, IPS, MMT, IPA, Tes Psikotes dan Kelas. Terkonfirmasi pada data jurusan siswa ada sebanyak 192 data siswa. Pada penelitian ini data yang digunakan adalah hasil *cleaning* data yang dapat dilihat pada Tabel 1 dibawah ini.

Table 1. Data Penelitian

Nama	Jenis Kelamin	B. Indonesia	B. Inggris	IPS	...	Kelas
S-1	1	84,0	81,0	86,6	...	IPS
S-2	2	79,6	84,5	84,6	...	IPS
S-3	1	83,0	81,3	83,1	...	IPA
S-4	2	85,0	84,8	88,6	...	IPA
S-5	1	90,6	90,1	90,8	...	IPA
S-6	2	84,3	84,8	83,3	...	IPS
S-7	1	81,4	82,0	84,6	...	IPA
S-8	2	82,0	89,0	80,8	...	IPA
S-9	2	85,0	85,0	85,0	...	IPA
S-10	1	86,3	81,0	87,1	...	IPA
...
S-192	1	81,8	76,6	80,8	...	IPS

3.2 Implementasi Algoritma C4.5 dan SVM

Implementasi yang dilakukan menggunakan algoritma C4.5 dan SVM terhadap *dataset* yang berjumlah 192 data, dan pengolahan data menggunakan *tool Google Colab* sehingga memberikan hasil berupa perbandingan nilai *Accuracy*, *Precision*, *Recall*, *F1-Score* dan *Support*, dapat dilihat pada Tabel 2.

Table 2. Hasil Perbandingan

Parameter	Algoritma C4.5	Algoritma SVM
<i>Accuracy</i>	0,56	0,65
<i>Precision</i>	0,54	0,66
<i>Recall</i>	0,74	0,73
<i>F1-Score</i>	0,62	0,69
<i>Support</i>	19	26

4. KESIMPULAN

Data Jurusan yang dikumpulkan dari SMA NEGERI 15 PEKANBARU sebanyak 192 data dengan 25 data testing. Klasifikasi yang dilakukan pada penelitian ini menggunakan tools Goole Colab dengan algoritma C4.5 dan SVM, validasi dalam menguji nilai akurasi maka diperoleh hasil bahwa algoritma SVM memiliki akurasi yang tinggi dengan presentase 65% dan algoritma C4.5 memiliki akurasi rendah dengan presentase 56%.

REFERENCES (HEADING 1, 10 pt)

- [1] I. W. C. Sujana, "Fungsi Dan Tujuan Pendidikan Indonesia," *Adi Widya J. Pendidik. Dasar*, vol. 4, no. 1, p. 29, 2019, doi: 10.25078/aw.v4i1.927.
- [2] D. Apriadi and R. Kuswandhie, "Sistem Pendukung Keputusan Penentuan Jurusan Pada Sma Bina Satria," *JUSIM (Jurnal Sist. Inf. Musirawas)*, vol. 5, no. 02, pp. 87–95, 2020, doi: 10.32767/jusim.v5i02.970.
- [3] S. M. Monalisa and F. Hadi, "Algoritma C4.5 dalam Penentuan Jurusan Siswa Baru," *Ultim. J. Tek. Inform.*, vol. 12, no. 2, pp. 108–113, 2020, doi: 10.31937/ti.v12i2.1838.
- [4] V. Anestiviya, A. Ferico, and O. Pasaribu, "Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus : Sman 1 Natar)," *J. Teknol. dan Sist. Inf.*, vol. 2, no. 1, pp. 80–85, 2021, [Online]. Available: <http://jim.teknokrat.ac.id/index.php/JTSI>
- [5] E. B. Sambani and F. Nuraeni, "Penerapan Algoritma C4.5 Untuk Klasifikasi Pola Penjurusan di Sekolah Menengah Kejuruan (SMK) Kota Tasikmalaya," *CSRID (Computer Sci. Res. Its Dev. Journal)*, vol. 9, no. 3, p. 144, 2018, doi: 10.22303/csrid.9.3.2017.144-152.
- [6] A. M. Fadlillah and D. Ruhjati, "Edukasi Perencanaan Karier bagi Pelajar SMA di Kecamatan Limo , Depok Career Planning Education for High School Students in Kecamatan Limo , Depok," *J. Mitra*, vol. 3, pp. 164–178, 2019.
- [7] R. Adha, N. Nurhaliza, U. Sholeha, and M. Mustakim, "Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia," *SITEKIN J. Sains, Teknol. dan Ind.*, vol. 18, no. 2, pp. 206–211, 2021.
- [8] A. Qurotul, E. Tasia, N. Nazira, P. F. Pratama, M. R. Anugrah, and J. Adhiva, "Implementasi Algoritma Naïve Bayes Classifier (NBC) untuk Klasifikasi Penyakit Ginjal Kronik," vol. 4, no. September, pp. 72–76, 2022, doi: 10.30865/json.v4i1.4781.
- [9] H. Annur, "Klasifikasi Masyarakat Miskin Menggunakan Metode Naive Bayes," *Ilk. J. Ilm.*, vol. 10, no. 2, pp. 160–165, 2018, doi: 10.33096/ilkom.v10i2.303.160-165.
- [10] R. Kurniasari and A. Fatmawati, "Penerapan Algoritma C4.5 Untuk Penjurusan Siswa Sekolah

- Menengah Atas,” *Komputa J. Ilm. Komput. dan Inform.*, vol. 8, no. 1, pp. 19–27, 2019, doi: 10.34010/komputa.v8i1.3045.
- [11] P. K. Handayani, “Penerapan Algoritma Support Vector Machine (Svm) Untuk Analisis Pola Klasifikasi Pada Parkinson’S Dataset,” *Indones. J. Technol. Informatics Sci.*, vol. 3, no. 1, pp. 31–35, 2021, doi: 10.24176/ijtis.v3i1.7530.
- [12] K. Ganjar Prasetyo and S. Mirza Pahlevi, “021-31904598 Fax. 021-31904599 2 Program Pascasarjana Magister Ilmu Komputer Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri (STMIK Nusa Mandiri),” *Jl. Salemba Raya*, vol. 5, no. 5, pp. 11–26, 2019.
- [13] S. Rizky Ade Putranto, Triastuti Wuryandari, “Perbandingan Analisis Klasifikasi Antara Decision Tree Dan Support Vector Machine Multiclass,” *J. Gaussian*, vol. 4, no. 4, pp. 1007–1016, 2015, [Online]. Available: Data Mining
- [14] N. Yulia, R. Saragih, and I. Ambarita, “Data Mining Pengelompokan Anak Stunting Berdasarkan Usia , Penyebab dan Pekerjaan Orang Tua Dengan Menggunakan Metode Clustering (Studi Kasus : Dinas Kesehatan Kabupaten Langkat),” *Semin. Nas. Inform. (SENATIKA)Prosiding Senat. 2021*, p. 12, 2021, [Online]. Available: <http://www.ejournal.pelitaindonesia.ac.id/ojs32/index.php/SENATIKA/article/view/1174/661>
- [15] A. Ridwan and T. N. Sari, “The comparison of accuracy between naïve bayes clasifier and c4.5 algorithm in classifying toddler nutrition status based on anthropometry index,” *J. Phys. Conf. Ser.*, vol. 1764, no. 1, 2021, doi: 10.1088/1742-6596/1764/1/012047.
- [16] M. Majedi-Asl, R. Daneshfaraz, and ..., “Simulation of bridge pier scour depth base on geometric characteristics and field data using support vector machine algorithm,” *J. Appl. ...*, no. February, 2020, doi: 10.22126/ARWW.2021.5747.1189.
- [17] M. Jain, S. Narayan, P. Balaji, K. P. Bharath, A. Bhowmick, and R. Karthik, “2002.07590”.
- [18] E. T. Lestari, Mustakim, and J. Adhiva, “Implementation Naive Bayes Classifier Algorithm and K-Nearest Neighbor For Obesity Nutritional Status of Children with Disabilities Implementasi Algoritma Naive Bayes Classifier dan K-Nearest Neighbor Untuk Klasifikasi Status Gizi Obesitas Anak Disabilitas,” pp. 1–11, 2022.
- [19] B. Charbuty and A. Abdulazeez, “Classification Based on Decision Tree Algorithm for Machine Learning,” *J. Appl. Sci. Technol. Trends*, vol. 2, no. 01, pp. 20–28, 2021, doi: 10.38094/jastt20165.
- [20] S. Tangirala, “Evaluating the impact of GINI index and information gain on classification using decision tree classifier algorithm,” *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 2, pp. 612–619, 2020, doi: 10.14569/ijacsa.2020.0110277.