



## EVALUASI METODE DATA MINING DECISION TREE C4.5 DAN NAÏVE BAYES DALAM MEMPREDIKSI PEMILIHAN JURUSAN SISWA (STUDI KASUS SMA MASTER INDONESIA)

Khairudin

Universitas Pamulang

dosen02591@unpam.ac.id

Kata kunci:	Abstrak
Data mining, <i>Decision tree</i> , <i>naive bayes</i> , pemilihan jurusan	Penentuan Pemilihan jurusan akan membantu Siswa lebih fokus terhadap apa yang diminati dan disesuaikan dengan nilai akademis yang dimilikinya. Banyak siswa yang masih belum mengenal minat dan kemampuan yang dimilikinya. Hal tersebut membuat siswa cenderung memilih dan menjalani konsentrasi yang tidak sesuai dengan minat dan kemampuannya. Perbandingan kinerja <i>Decision tree</i> dan <i>naive bayes</i> bertujuan untuk mengukur tingkat akurasi terbaik masing-masing. Dan Hasil dari komparasi Evaluasi pada metode antara <i>decision tree</i> C4.5, dan <i>naive bayes</i> yang digabungkan dengan metode seleksi fitur <i>forward selection</i> untuk kasus ketepatan pemilihan jurusan siswa didapatkan tingkat akurasi tertinggi dengan algoritma terpilih C4.5 dengan nilai akurasi sebesar 90,35%.

### Pendahuluan

Pemilihan jurusan pendidikan merupakan keputusan penting bagi siswa di tingkat sekolah menengah atas (Rufaidah, 2015). Keputusan ini akan mempengaruhi jalur pendidikan dan karir mereka di masa depan (Krisbiyanto & Nadhifah, 2022). Oleh karena itu, penting untuk mengembangkan metode yang efektif dalam memprediksi pemilihan jurusan siswa berdasarkan faktor-faktor yang relevan. Salah satu pendekatan yang menjanjikan adalah menggunakan metode data mining untuk menganalisis data siswa dan membangun model prediksi yang akurat (Narulita dkk., 2021).

Dalam penelitian ini, kami akan melakukan evaluasi terhadap dua metode data mining yang umum digunakan, yaitu Decision Tree C4.5 dan Naïve Bayes, dalam konteks pemilihan jurusan siswa di SMA Master Indonesia. SMA Master Indonesia merupakan lembaga pendidikan yang terkenal dengan beragam program studi yang ditawarkan kepada siswanya. Tujuan utama dari penelitian ini adalah untuk membandingkan kinerja kedua metode dalam memprediksi pemilihan jurusan siswa dan menentukan metode yang paling akurat dalam konteks studi kasus ini.

Data yang digunakan dalam penelitian ini meliputi informasi mengenai nilai raport siswa, nilai ujian nasional (UN), minat siswa, dan atribut-atribut lain yang relevan. Dengan menggunakan metode data mining, kami akan membangun model prediksi berdasarkan data tersebut dan melakukan pengujian serta evaluasi terhadap kinerja model. Hasil penelitian ini

diharapkan dapat memberikan wawasan yang berguna bagi lembaga pendidikan dalam memperbaiki proses pemilihan jurusan siswa yang lebih efektif dan akurat.

Melalui evaluasi metode data mining Decision Tree C4.5 dan Naïve Bayes dalam memprediksi pemilihan jurusan siswa di SMA Master Indonesia, diharapkan dapat memberikan kontribusi yang signifikan dalam pengembangan pendekatan prediksi yang dapat digunakan oleh lembaga pendidikan dalam membantu siswa dan orang tua dalam mempertimbangkan pilihan jurusan yang sesuai dengan minat dan kemampuan siswa. Penelitian ini juga dapat membuka peluang untuk penelitian lebih lanjut dalam mengoptimalkan metode data mining lainnya untuk prediksi pemilihan jurusan siswa. Pendidikan merupakan salah satu tolak ukur kemajuan suatu bangsa. Bangsa yang besar memiliki perhatian terhadap pendidikan yang berkualitas untuk rakyatnya. Pendidikan yang berkualitas akan menghasilkan generasi yang berkualitas untuk membangun bangsa dan negaranya. Sekolah Menengah Atas (SMA) adalah jenjang pendidikan menengah pada pendidikan formal di Indonesia. Pada umumnya, SMA mempunyai dua jurusan yaitu Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). IPA merupakan jurusan yang paling diminati oleh siswa karena lulusan IPA bebas memilih fakultas ketika akan melanjutkan ke perguruan tinggi. Sedangkan jurusan IPS mempunyai pilihan yang terbatas.

SMA Master Indonesia merupakan salah satu Sekolah Menengah Atas di kota Depok yang beralamat di Jalan Raya Margonda. Seperti pada umumnya sekolah ini mempunyai pilihan jurusan IPA dan IPS. Penjurusan siswa ini bertujuan untuk mengarahkan peserta didik agar lebih fokus mengembangkan kemampuan dan minat yang dimiliki. Pemilihan jurusan yang tidak tepat dapat sangat merugikan siswa dan kehidupannya dimasa yang akan datang. Dengan sistem penjurusan yang ada pada SMA, diharapkan dapat memaksimalkan potensi, bakat atau talenta individu yang dimiliki oleh siswa. Penentuan jurusan ini akan berdampak terhadap kegiatan akademik selanjutnya dan mempengaruhi pemilihan bidang ilmu atau studi bagi siswa-siswi yang ingin melanjutkan ke perguruan tinggi. Penentuan jurusan yang dilakukan secara manual mempunyai banyak kelemahan karena penentuan jurusan pada umumnya hanya dilakukan berdasarkan nilai akademiknya saja tanpa mempertimbangkan tingkat kecerdasan, minat dan bakat yang dimiliki siswa, kemampuan pembiayaan jika ingin melanjutkan ke perguruan tinggi, serta factor-faktor lain yang mungkin akan berpengaruh terhadap kehidupan siswa tersebut.

Penelitian ini memiliki signifikansi yang penting dalam beberapa aspek. Pertama, penelitian ini bertujuan untuk mengevaluasi metode data mining Decision Tree C4.5 dan Naïve Bayes dalam memprediksi pemilihan jurusan siswa di SMA Master Indonesia. Dalam konteks dunia pendidikan yang semakin kompleks, penggunaan metode data mining dapat memberikan kontribusi yang besar dalam pengambilan keputusan yang lebih baik dan akurat dalam pemilihan jurusan siswa. Hasil penelitian ini dapat memberikan panduan praktis bagi lembaga pendidikan dalam mengembangkan strategi dan kebijakan yang tepat dalam hal pemilihan jurusan siswa.

Kedua, penelitian ini juga penting karena dilakukan dalam konteks studi kasus SMA Master Indonesia. Dalam konteks tersebut, penelitian ini memberikan wawasan yang spesifik tentang kinerja metode data mining dalam memprediksi pemilihan jurusan siswa di sekolah tersebut. Dengan memahami kinerja metode ini dalam konteks yang nyata, penelitian ini dapat memberikan masukan yang lebih relevan dan bermanfaat bagi SMA Master Indonesia dalam meningkatkan proses pemilihan jurusan siswa yang lebih efektif.

Meskipun telah ada penelitian terkait penggunaan metode data mining dalam memprediksi pemilihan jurusan siswa, masih terdapat beberapa celah penelitian yang dapat diidentifikasi. Salah satunya adalah kurangnya penelitian yang secara spesifik menggabungkan evaluasi kinerja metode Decision Tree C4.5 dan Naïve Bayes dalam konteks pemilihan jurusan siswa di Indonesia, terutama dalam lingkungan sekolah menengah atas.

Dalam konteks ini, penelitian sebelumnya lebih fokus pada satu metode atau menggunakan dataset yang berbeda.

Selain itu, ada juga kekurangan dalam penelitian terkait dalam hal penggunaan atribut-atribut yang relevan dan spesifik untuk pemilihan jurusan siswa. Beberapa penelitian sebelumnya cenderung hanya mempertimbangkan atribut akademik seperti nilai raport atau hasil ujian nasional, tanpa mempertimbangkan faktor-faktor lain seperti minat, bakat, atau preferensi siswa. Oleh karena itu, terdapat kebutuhan untuk melakukan penelitian yang lebih komprehensif dengan memasukkan atribut-atribut yang lebih luas dan beragam dalam memprediksi pemilihan jurusan siswa.

Dengan mengisi kesenjangan penelitian ini, penelitian kami diharapkan dapat memberikan kontribusi yang berharga dalam pemahaman dan pengembangan metode data mining dalam memprediksi pemilihan jurusan siswa.

## **Metode**

Penelitian ini menggunakan metode penelitian kuantitatif dengan pendekatan studi kasus. Tujuan penelitian ini adalah untuk melakukan evaluasi terhadap metode data mining Decision Tree C4.5 dan Naïve Bayes dalam memprediksi pemilihan jurusan siswa di SMA Master Indonesia (Suhada dkk., 2021). Sampel penelitian terdiri dari siswa-siswa yang telah memilih jurusan di SMA Master Indonesia. Jumlah sampel yang diambil dalam penelitian ini dapat ditentukan berdasarkan kebutuhan analisis statistik yang akurat dan reliabel. Pengumpulan data dilakukan dengan menggunakan metode dokumentasi dan angket. Data yang diperoleh meliputi nilai raport siswa, nilai ujian nasional (UN) siswa, informasi pemilihan jurusan siswa, dan atribut-atribut lain yang relevan.

### **Metode Analisis Data:**

Data yang dikumpulkan akan dianalisis menggunakan metode data mining Decision Tree C4.5 dan Naïve Bayes. Proses analisis akan melibatkan pembangunan model prediksi berdasarkan atribut-atribut yang ada, pelatihan model menggunakan data latih, serta pengujian dan evaluasi kinerja model menggunakan data uji terpisah.

### **Validitas dan Reliabilitas:**

Untuk memastikan validitas dan reliabilitas hasil penelitian, langkah-langkah berikut akan diimplementasikan: (1) Validitas konstruk akan diperiksa dengan memastikan bahwa atribut yang digunakan dalam analisis relevan dengan pemilihan jurusan siswa; (2) Validitas internal akan dijaga dengan memastikan bahwa langkah-langkah analisis dilakukan secara konsisten dan tepat; (3) Validitas eksternal akan diperhatikan dengan memastikan bahwa sampel penelitian mencerminkan populasi siswa SMA Master Indonesia secara umum; dan (4) Reliabilitas akan diperiksa dengan menggunakan teknik pengujian ulang (test-retest) atau alat ukur yang telah teruji keandalannya (Budiastuti & Bandur, 2018).

### **Etika Penelitian:**

Penelitian ini akan mematuhi prinsip-prinsip etika penelitian, termasuk penggunaan data yang bersifat rahasia atau sensitif dengan izin yang sesuai, anonimitas peserta penelitian, dan perlindungan hak-hak peserta penelitian.

Dalam penelitian ini, evaluasi metode data mining Decision Tree C4.5 dan Naïve Bayes dalam memprediksi pemilihan jurusan siswa di SMA Master Indonesia akan dilakukan dengan mengumpulkan data yang relevan, menerapkan analisis yang sesuai, dan memastikan validitas serta reliabilitas hasil penelitian. Dengan menggunakan metode penelitian ini, diharapkan dapat memberikan pemahaman yang lebih baik tentang kinerja kedua metode dalam memprediksi pemilihan jurusan siswa dan memberikan wawasan yang berguna bagi lembaga pendidikan dalam meningkatkan proses pemilihan jurusan siswa yang lebih efektif.

## **Hasil dan Pembahasan**

Penggunaan seleksi fitur forward selection pada algoritma C4.5 dan Naive Bayes dalam kasus ketepatan pemilihan konsentrasi siswa dengan melibatkan atribut-atribut seperti nilai raport IPS, nilai raport IPA, nilai UN Matematika, nilai UN IPS, dan nilai UN IPA meningkatkan hasil akurasi. Hal ini menunjukkan bahwa penggunaan seleksi fitur dapat membantu meningkatkan kinerja dan akurasi prediksi dari kedua algoritma tersebut.

Hasil uji kinerja algoritma klasifikasi menunjukkan bahwa pada kasus ketepatan pemilihan jurusan siswa, algoritma C4.5 tanpa penambahan seleksi fitur forward selection memiliki akurasi sebesar 83,38%, sedangkan algoritma Naive Bayes tanpa penambahan seleksi fitur forward selection memiliki akurasi sebesar 80,33%. Namun, ketika dilakukan penambahan seleksi fitur forward selection pada algoritma Decision Tree, akurasi meningkat menjadi 82,49%. Selain itu, algoritma Decision Tree tanpa seleksi fitur forward selection menghasilkan akurasi yang sangat tinggi sebesar 90,35%.

Hasil temuan menunjukkan bahwa penggunaan seleksi fitur forward selection secara positif mempengaruhi kinerja dan akurasi prediksi dari algoritma C4.5 dan Naive Bayes dalam kasus pemilihan jurusan siswa. Dengan melibatkan atribut-atribut yang relevan, seleksi fitur membantu mengidentifikasi atribut-atribut yang memiliki pengaruh signifikan terhadap prediksi pemilihan jurusan siswa, sehingga menghasilkan model yang lebih akurat.

Selain itu, hasil uji kinerja menunjukkan bahwa algoritma C4.5 tanpa penambahan seleksi fitur forward selection memiliki akurasi yang lebih tinggi dibandingkan dengan algoritma Naive Bayes tanpa seleksi fitur. Namun, ketika dilakukan penambahan seleksi fitur forward selection pada algoritma Decision Tree, akurasi meningkat secara signifikan. Hal ini menunjukkan bahwa penggunaan seleksi fitur pada algoritma Decision Tree memberikan kontribusi yang signifikan dalam meningkatkan akurasi prediksi pemilihan jurusan siswa.

Dalam keseluruhan penelitian ini, dapat disimpulkan bahwa penggunaan algoritma C4.5 dengan penambahan seleksi fitur forward selection memiliki kinerja yang lebih baik dibandingkan dengan algoritma Naive Bayes dalam memprediksi pemilihan jurusan siswa. Namun, algoritma Decision Tree tanpa seleksi fitur forward selection menghasilkan akurasi yang sangat tinggi, namun ketika dilakukan seleksi fitur, akurasi sedikit menurun. Penelitian ini memberikan wawasan penting bagi lembaga pendidikan dalam mengoptimalkan proses pemilihan jurusan siswa dengan memanfaatkan algoritma data mining yang tepat. Metode data mining merupakan pendekatan yang efektif dalam menganalisis dan menggali pola-pola tersembunyi dalam data. Dalam penelitian ini, kami melakukan evaluasi terhadap dua metode data mining yang populer, yaitu Decision Tree C4.5 dan Naive Bayes, dalam memprediksi pemilihan jurusan siswa di SMA Master Indonesia. Tujuan utama dari evaluasi ini adalah untuk menentukan metode mana yang memberikan hasil prediksi yang lebih akurat dan andal.

Dalam proses evaluasi, kami menggunakan data siswa SMA Master Indonesia yang terdiri dari berbagai atribut, termasuk nilai akademik, minat siswa, dan latar belakang pendidikan. Kami melatih model Decision Tree C4.5 dan model Naive Bayes menggunakan data tersebut, kemudian menguji kedua model dengan menggunakan data tes yang terpisah. Hasil evaluasi menunjukkan bahwa kedua metode memberikan performa yang baik dalam memprediksi pemilihan jurusan siswa, namun Decision Tree C4.5 memberikan akurasi yang sedikit lebih tinggi dibandingkan dengan Naive Bayes dengan nilai akurasi sebesar 8% dan 82% masing-masingnya.

Dalam konteks pemilihan jurusan siswa, hasil evaluasi ini memberikan wawasan yang berharga bagi para pengambil keputusan di SMA Master Indonesia. Metode Decision Tree C4.5 dapat digunakan sebagai alat bantu yang efektif dalam membantu siswa dan orang tua

dalam mempertimbangkan pilihan jurusan yang sesuai dengan minat dan kemampuan siswa. Namun, perlu dicatat bahwa hasil evaluasi ini spesifik untuk studi kasus SMA Master Indonesia dan perlu dilakukan penelitian lebih lanjut untuk memvalidasi hasil ini pada populasi yang lebih luas.

### **Kesimpulan**

Berdasarkan hasil implementasi algoritma C4.5 dan Naive Bayes pada kasus ketepatan pemilihan jurusan siswa dapat diambil beberapa kesimpulan sebagai berikut :

a. Dengan dataset yang sama, penggunaan seleksi fitur forward selection pada algoritma C4.5 dan Naive Bayes pada kasus ketepatan pemilihan konsentrasi siswa dengan melibatkan atributatribut : Nilai raport ips, Nilai raport ipa, Nilai Un Matematika, Nilai Un ips, dan Nilai Un Ipa meningkatkan hasil akurasi.

b. Hasil uji kinerja algoritma klasifikasi untuk kasus ketepatan pemilihan Jurusan siswa untuk algoritma C4.5 tanpa penambahan seleksi fitur forward selection diperoleh nilai akurasi sebesar 83,38%, Sedangkan pada algoritma Naive Bayes tanpa penambahan seleksi fitur forward selection diperoleh nilai akurasi sebesar 80,33%, Kemudian pada algoritma Decision tree penambahan seleksi fitur forward selection diperoleh nilai akurasi hanya sebesar 82,49% dan sedangkan tanpa penambahan seleksi fitur forward selection hasilnya sangat besar 90.35% Kinerja antara algoritma C4.5 tanpa penambahan seleksi fitur forward selection dengan algoritma C4.5 ditambah seleksi fitur forward selection lebih unggul bila dibandingkan dengan algoritma Naive Bayes pada kasus ketepatan pemilihan Jurusan siswa.

### **Daftar Pustaka**

- Budiastuti, D., & Bandur, A. (2018). Validitas dan reliabilitas penelitian. *Jakarta: Mitra Wacana Media*.
- Krisbiyanto, A., & Nadhifah, I. (2022). Pengaruh Lokasi dan Citra Sekolah Terhadap Keputusan Siswa Memilih Sekolah di Sekolah Menengah Atas Negeri. *Academicus: Journal of Teaching and Learning*, 1(1), 20–31.
- Narulita, S., Oktaga, A. T., & Susanti, I. (2021). Pengujian Akurasi Model Prediksi Menggunakan Metode Data Mining Classification Decision Tree Algoritma C4. 5 untuk Penentuan Peminatan Peserta Didik. *Media Aplikom*, 13(2), 65–79.
- Rufaidah, A. (2015). Pengaruh intelegensi dan minat siswa terhadap putusan pemilihan jurusan. *Faktor: Jurnal Ilmiah Kependidikan*, 2(2).
- Suhada, K., Elanda, A., & Aziz, A. (2021). Klasifikasi Predikat Tingkat Kelulusan Mahasiswa Program Studi Teknik Informatika dengan Menggunakan Algoritma C4. 5 (Studi Kasus: STMIK Rosma Karawang). *Dirgamaya: Jurnal Manajemen Dan Sistem Informasi*, 1(2), 14–27.