

Klasifikasi Minat Karir Siswa Sekolah Menengah Atas Menggunakan Algoritma C4.5

Oleh:

Dina Nofitasary,

Yunianita Rahmawati

Progam Studi Informatika

Universitas Muhammadiyah Sidoarjo

Mei , 2024



Pendahuluan

Banyaknya siswa SMA saat ini yang mengalami kebingungan dalam menentukan minat karir menuju jenjang selanjutnya[1], mulai dari ketidakpahaman akan yang dipilih hingga adanya pengaruh lingkungan dalam menentukan karir[2]. Siswa-siswi yang awalnya merasa yakin pun menjadi ragu dengan minat karir yang diambil. Akibatnya banyak siswa yang memilih tanpa dasar pemahaman[3] dan tidak maksimalnya baik ilmu maupun kemampuan yang didapat karena tidak sesuai dengan minat dan bakat yang dimiliki. Maka dari itu peneliti melakukan penelitian di SMA Negeri 1 Wringinanom, Kab. Gresik, Prov. Jawa Timur. Memungkinkan pihak sekolah membantu dalam meyakinkan siswa untuk memilih minat karir siswa sehingga sesuai dengan minat dan bakat siswa.

Pendahuluan

Berdasarkan data statistik Kementerian Pendidikan dan Kebudayaan pada tahun 2021, sebanyak 1,3 juta dari 3,6 juta siswa lulus yang melanjutkan studi ke perguruan tinggi[4]. Tidak sedikit yang memilih hanya menganggur disebabkan tidak yakinnya dengan minat karir yang di fokuskan dengan bakat masing-masing siswa[5]. Ini Sebagian dari banyaknya potensi yang bisa terjadi sesuai kondisi dari setiap masing-masing siswa[6].

Pendahuluan

Dari berbagai algoritma, dalam kasus hampir sama, yang berhubungan dengan minat karir siswa SMA sangat sesuai jika diproses menggunakan metode C4.5 adalah teknik model prediksi yang dapat digunakan untuk klasifikasi dan prediksi tugas[7]. seperti penelitian sebelumnya Dimana perbandingan menggunakan nilai precision dan accuracy hasil uji dari lima algoritma yang mendapatkan hasil bahwa dari K-Nearest Neighbor, Naive Bayes, Decision Tree (C4.5), Rule Induction, dan Deep Learning yang menghasilkan metode Decision Tree (C4.5) mengungguli algoritma lainnya dengan akurasi sebesar 89,09% dan nilai akurasi sebesar 84,00% [8].

Pendahuluan

Dengan adanya penelitian ini bertujuan untuk membantu pihak sekolah untuk meyakinkan siswa dalam memilih minat karir siswa di SMA menggunakan algoritma C4.5 sehingga lebih terarah sesuai dengan minat dan kemampuan dari siswa. Minat karir yang terdapat di SMA Negeri 1 Wringinanom dikelompokkan menjadi kuliah, bekerja, kedinasaan, dan menikah. Menggunakan data-data yang diambil dengan data riwayat pendidikan selama bersekolah dengan menggunakan data terbaru sehingga lebih akurat dan dapat dijadikan rujukan bagi sekolah dalam sistem Pendidikan. Banyak faktor yang mempengaruhi seperti jurusan yang ternyata memiliki pengaruh penting dalam minat karir siswa[9]. Bukan hanya itu, hobi juga dapat mempengaruhi secara internal siswa itu sendiri[10]. Dengan menggunakan variabel yang lebih bervariasi yaitu, jenis kelamin, jurusan, prestasi akademik, prestasi non akademik, hobi, kehadiran, surat peringatan, berkarir. Memakai metode Algoritma C4.5 yang diharapkan dapat mendapat hasil yang sesuai dan menghasilkan akurasi yang lebih tepat dan sesuai dengan tingkat keyakinan siswa terhadap masalah minat karir yang diangkat.

Pertanyaan Penelitian (Rumusan Masalah)

Bagaimana mengklasifikasikan minat karir siswa sekolah menengah atas setelah lulus menggunakan algoritma C4.5?

Metode

Penelitian ini dilakukan oleh siswa kelas X dan XI SMA Negeri 1 Wringinanom pada tanggal 24-25 April 2024 dengan beberapa tahap. Langkah langkah penelitian ini disajikan dalam gambar gambar 1, yaitu: studi Pustaka, pengumpulan data, pre-processing, implementasi algoritma, dan pengetahuan hasil.



Gambar 1. Proses Kerja Penelitian

Metode

3.1. Studi Pustaka

Dalam tahapan pada penelitian ini studi Pustaka tentang mengkasifikasikan minat karir siswa Sekolah Menengah Atas menggunakan Algoritma C4.5 untuk mengetahui keyakinan dalam pilihan karir siswa

3.2. Pengumpulan Data

Penelitian ini dilakukan pada tanggal 24-25 April 2024 dan di SMA Negeri 1 Wringinanom, Kab. Gresik, Jawa Timur. Bahan yang digunakan dalam penelitian ini adalah data sekunder para siswa kelas X dan XI SMA Negeri 1 Wringinanom dan alat berupa perangkat keras berupa laptop maupun perangkat lunak berupa google form sebagai alat pengumpulan data. Kemudian data diinput kedalam microsoft excel dan dikelompokan sesuai minat karir. Perlu melakukan pengelompokan data siswa ke dalam beberapa kelompok [15], yaitu bekerja, kuliah, menikah, dan kedinasan. Banyaknya data yang didapat dari hasil kuisisioner yang dilakukan sebanyak 76 data yang bersal dari 2 kelas.

Metode

3.3. Pre-Processing

Pada tahapan ini perlu melibatkan 4 rangkaian proses yang diperlukan dalam mempersiapkan data mentah sebelum digunakan untuk melatih model. Berikut adalah langkah-langkah pre-processing yang diperlukan :

1) Load Data

Pada tahapan ini perlu melakukan impor dataset yang akan digunakan untuk melatih data model. Data yang digunakan dapat berupa CSV maupun format lainnya.

2) Handling Missing Value

Tahapan ini menganalisa adanya nilai yang hilang dalam dataset. Dengan menghapus baris atau kolom yang mengandung nilai yang hilang atau mengisi mereka dengan nilai tertentu.

3) Encode Categorical Data

Jika dalam dataset terdapat fitur-fitur kategorikal, perlu dilakukan konversi menjadi numerik.

Metode

3.4. Implementasi Algoritma

Dalam implementasi algoritma dilakukan perhitungan dengan Google Colab menggunakan algoritma C4.5 dengan data yang sebelumnya telah dikelompokkan menjadi 4, yaitu data kuliah, data bekerja, data menikah, dan data kedinasan kemudian diinputkan kedalam microsoft excel. Sehingga data tersebut dapat diolah dengan algoritma C4.5 menggunakan program Python

Terdapat beberapa tahapan yang dilakukan dalam pengoprasian menggunakan algoritma C4.5. berikut adalahh tahapan yang harus dilakukan :

1) Mencluster Minat Siswa

Dalam tahap ini, perlu melakukan pengelompokan data siswa ke dalam beberapa kelompok. Pengelompokan dilakukan sesuai dengan minat karir siswa baik itu kuliah, bekerja, kedinasan bahkan menikah. Hal ini dilakukan agar dapat mempermudah dalam menentukan pohon keputusan sesuai minat karir menggunakan algoritma C4.5.

Metode

2) Perhitungan Total Entropy

Langkah selanjutnya, mencari nilai dari Entropy Total dari data kasus yang telah dikumpulkan. Berikut adalah perhitungan secara manual :

$$\text{Entropy}(S) = -\sum_{i=1}^c \left[\left(\frac{p_i}{n} \right) \log_2 \left(\frac{p_i}{n} \right) \right] \dots \dots \dots 1$$

Keterangan :

S : Himpunan Kasus.

n : Jumlah partisi atribut.S

pi : Proporsi dari Si.terhadap S

Metode

3) Perhitungan Entropy dan Gain

Agar mendapatkan root node (node akar), perlu melakukan perhitungan pada gain dan entropy tiap atribut. Berikut adalah bentuk perhitungan yang digunakan :

Entropy.:

$$\text{Entropy}(S) = -\sum_{i=1}^n p_i \log_2(p_i) \dots\dots\dots 2$$

Keterangan :

S : Himpunan Kasus

n : Jumlah partisi atribut S

pi : Proporsi dari Si terhadap S

Gain :

$$\text{Gain}(S,A) = \text{Entropi}(S) - \sum_{v \in \text{Values}(A)} \frac{|S_v|}{|S|} \cdot \text{Entropi}(S_v) \dots\dots 3$$

Keterangan :

S = himpunan data.

A = atribut yang akan digunakan untuk membagi S.

Values (A) = nilai-nilai yang mungkin dari atribut A.

|Sv| = jumlah instance dalam S yang memiliki nilai v pada atribut A.

|S| = total jumlah instance dalam S.

Entropi (S) = entropi dari S, dihitung menggunakan rumus entropi total yang telah dijelaskan sebelumnya.

Entropi(Sv) = entropi dari subset Sv, yaitu himpunan instance dalam S yang memiliki nilai v pada atribut A

Langkah ini dilakukan berulang hingga semua record terpartisi.

Metode

4) Menentukan Root Node

Setelah mendapatkan hasil entropy dan gain dari data kasus yang telah didapatkan sebelumnya, kemudian mencari Gain tertinggi untuk dijadikan akar dari pohon keputusan. Jika entropy dari atribut tersebut bernilai 0 maka perhitungan di selesaikan pada akar pohon keputusan tersebut. Kemudian melakukan perhitungan dengan mengeksekusi data sesuai dengan variabel. Sehingga mendapatkan hasil akhir berupa persentase akurasi dan pohon Keputusan.

Metode

3.5 Pengetahuan Hasil

Pada tahapan ini, memperoleh hasil pengolahan data kasus dengan menggunakan Algoritma C4.5 diatas menghasilkan pohon Keputusan. Dimana pada data siswa yang memilih minat karir akan dilakukan pembagian data training dan data testing sebesar 60:40 yang bersala dari data kuisisioner dan diambil secara random. Dengan hasil berupa pohon keputusan dan persentase akurasi.

Akurasi yang didapat berasal dari perbandingan prediksi yang dihasilkan dengan yang sebenarnya dari data pengujian. Menghitung jumlah prediksi yang benar dan akurasi model. Akurasi dapat dihitung menggunakan rumus berikut :

$$\text{Akurasi} = \frac{\text{Total jumlah prediksi}}{\text{Jumlah prediksi benar}} \times 100\% \dots \dots \dots 4$$

Hasil

Dari hasil uji, Algoritma Decision Tree C4.5 sebagai algoritma dalam memprediksi minat karir siswa SMA Negeri 1 Wringinanom, Kab. Gresik, Jawa Timur sebanyak 2 kelas yaitu data sekunder para siswa kelas X dan XI dengan keseluruhan jumlah data sebanyak 76 data. Dimana data itu terbagi menjadi 4, yaitu mencakup 45 data yang memilih kuliah, 12 data siswa yang memilih untuk bekerja, 8 data memilih menikah dan 11 data siswa yang memilih kedinasan menggunakan Algoritma nantinya. Dimana dilakukan pembagian data training dan data testing sebesar 60:40. Hasil akhir yang didapat bahwa akurasi paling tinggi didapat oleh mereka yang bekerja sebesar 80%, kemudian yang memilih kuliah sebesar 72,2%, lalu yang memilih menikah sebesar 50% dan yang terakhir adalah kedinasan sebesar 80%. Maka siswa yang memilih bekerja dan kedinasan telah yakin dengan pilihannya, sedangkan yang memilih kuliah juga tidak terlalu yakin, dan yang paling tidak yakin dengan pilihan yang mereka ambil adalah menikah dengan persentase keyakinan terendah. Maka siswa dengan pilihan menikah dan kuliah perlu diberikan keyakinan dengan pilihannya agar dapat meminimalisir terjadinya ketidaksesuaian karir sesuai minat dan bakat siswa.

Pembahasan

Dari data diatas terbagi menjadi 4, yaitu mencakup 12 data siswa yang memilih untuk bekerja, 45 data yang memilih kuliah, 8 data memilih menikah dan 11 data siswa yang memilih kedinasan. Dimana masih terdapat siswa yang sudah yakin maupun masih ragu dengan pilihan yang diambil.

Maka dilakukan penelitian ini agar dapat mengetahui minat karir siswa yang sudah yakin dan tepat dengan apa yang mereka pilih dengan menggunakan algoritma C4.5 dan Google Colab dengan harapan dapat memberikan hasil yang maksimal. Berikut adalah tahapan yang harus dilakukan :

1) Input Data

Data diinputkan nantinya akan diolah menggunakan algoritma C4.5 menggunakan bahasa Python dengan perintah-perintah tertentu.

Pembahasan

2) Kolom Unique-Integer

Langkah ini mengubah salah satu kolom dalam data menjadi bilangan bulat untuk digunakan dalam pemodelan atau analisa. Fungsi ini akan menghasilkan array dengan bilangan bulat dan dapat merepresentasikan kelas-kelas yang berbeda dalam kolom tersebut

3) Testing dan Training

Langkah ini adalah pembagian antara data uji dan data latih. Mengevaluasi seberapa baik model yang telah dilatih dapat memprediksi target. Ini membantu memastikan bahwa model tidak hanya mempelajari pola yang spesifik untuk data latih tetapi juga mampu menggeneralisasikan pola tersebut dengan baik. Dimana perbandingan data uji sebesar 60% dan data latih sebesar 40%.

Pembahasan

4) Pohon Keputusan

Pada bagian ini akan memunculkan pohon Keputusan yang berasal dari data yang diinputkan sebelumnya.

5) Akurasi

Meningkatkan akurasi model dengan membuat pohon keputusan yang memaksimalkan kemampuan untuk memisahkan kelas target dengan benar. Maka evaluasi akurasi menjadi penting untuk memastikan model yang dihasilkan dapat membuat prediksi yang benar dengan menggunakan data yang belum pernah dilihat sebelumnya.

Temuan Penting Penelitian

Berdasarkan hasil uji, persentase tingkat akurasi keyakinan minat karir siswa SMA yang diperoleh sebesar 80% untuk pilihan bekerja dan kedinasan, 72,2% dengan pilihan kuliah, dan 50% dengan pilihan menikah. Dengan hasil akurasi persentase keyakinan terendah, maka siswa yang memilih menikah perlu diyakinkan kembali dengan pilihannya. Hasil persentase ini dapat dijadikan acuan bagi pihak sekolah untuk membantu meyakinkan siswa dalam memilih minat karir sesuai dengan minat dan bakat masing-masing.

Manfaat Penelitian

1. Untuk memprediksi minat karir dari siswa sekolah menengah atas setelah lulus
2. Untuk mengetahui karir siswa yang cocok sesuai minat agar terarah dan tercapai tujuan siswa sekolah menengah atas
3. Menggunakan metode Algoritma C4.5 agar mendapatkan persentase hasil yang lebih akurat

Referensi

- [1] P. P. Rahayu, "Perancangan Karir Di Masa Pandemi COVID-19 Bagi Siswa Kelas XII SMA Negeri 5 Semarang," *Communnity Development Journal*, vol. 3, no. 1, hlm. 106–110, 2022.
- [2] I. Saripah¹, D. R. Priliani², dan N. A. Nadhirah, "Indonesian Journal of Guidance and Counseling: Theory and Application," *IJGC*, vol. 12, no. 1, hlm. 95–118, 2023, doi: 10.15294/ijgc.v12i1.7043.
- [3] P. K. Karir, S. Dengan, T. Holland, R. Asri, A. M. Yusuf, dan A. Afdal, "Peningkatan Kematangan Karir Siswa Dengan Teori Holland," *Featured Research 121 SCHOULID: Indonesian Journal of School Counseling*, vol. 6, no. 2, hlm. 121–132, 2021, doi: 10.23916/08935011.
- [4] A. Doahir dan A. N. Qolbi, "Analisis Potensi Siswa Untuk Kuliah Dengan Classification Menggunakan Metode Decision Tree," *Jurnal Poros Teknik*, vol. 14, no. 1, hlm. 28–32, 2022.
- [5] M. Roghib, N. Rahaningsih, dan R. D. Dana, "Penerapan Algoritma C4.5 Untuk Seleksi Penjurusan Siswa Baru Pada Sekolah Menengah Kejuruan (Studi Kasus: Smk Plus Al-Hilal Arjawinangun)," *Jurnal Mahasiswa Teknik Informatika*, Vol. 8, No. 1, hlm. 861-866, Feb 2024.
- [6] M. Fikri Rais dan K. Kunci, "Analisis Dampak Kegiatan Ekstrakurikuler Terhadap Hasil Belajar Siswa SMA Adabiah Padang," *Jurnal Pendidikan dan Olahraga*, Vol 3, No 6, hal 7-15, 2020.
- [7] R. Haqmanullah Pambudi dan B. Darma Setiawan, "Penerapan Algoritma C4.5 Untuk Memprediksi Nilai Kelulusan Siswa Sekolah Menengah Berdasarkan Faktor Eksternal," *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, Vol. 2, No. 7, hlm. 2637-2643, 2018. [Daring]. Tersedia pada: <http://j-ptiik.ub.ac.id>
- [8] A. Purwanto dkk., "Perbandingan Minat Siswa Smu Pada Metode Klasifikasi Menggunakan 5 Algoritma," *Ikraith Informatika*, vol. 2, no. 1, hlm. 43-47, 2018.
- [9] Yusuf S.Nugroho dan Syarifah N.Haryati, "Klasifikasi dan Klastering Penjurusan Siswa SMA Negeri 3 Boyolali," *Khazanah Informatika*, vol 1, no 1, hlm. 1-6, 2015.
- [10] M. K. Rokhman, S. Sucipto, dan M. Masturi, "Mengatasi Prokrastinasi Akademik Melalui Konseling Behavioristik Dengan Teknik Behavior Contract," *Jurnal Prakarsa Paedagogia*, vol. 2, no. 1, hlm. 10–18, Jan 2020, doi: 10.24176/jpp.v2i1.4310.
- [11] T. Cahyono dan N. Made Diah Padmi, "Pemetaan Minat, Bakat dan Karakter Berbasis RIASEC sebagai Acuan Peminatan Lintas Mata Pelajaran pada Jenjang SMA Kurikulum Merdeka," *Islamic Counseling Journal*, vol. 6, no. 1, hlm. 90–97, 2023, doi: <http://dx.doi.org/10.24235/prophetic.v6i1.14764>.
- [12] T. Restiany, A. Sudrajat, B. dan Konseling, M. Karir Siswa, dan S. Menengah Atas, "Manajemen Bimbingan dan Konseling Sekolah untuk Menetapkan Minat Karir Siswa Sekolah Menengah Atas (Studi Kasus di SMA Negeri 1 Majalaya dan SMA Negeri 2 Majalaya Kabupaten Bandung)," *Jurnal Ilmiah Ilmu Pendidikan*, vol. 5, no. 6, hlm. 1849-1861, 2022. [Daring]. Tersedia pada: <http://Jiip.stkipyapisdompu.ac.id>
- [13] J. Amsir dan A. Tigor Arifyanto, "Faktor-Faktor Penghambat Perencanaan Karir Siswa," *Jurnal Bening*, vol. 5, hlm. 57–66, 2021, doi: <http://dx.doi.org/10.36709/bening.v5i1.12107>.
- [14] S. Syamsu, N. Suta Wijaya, T. Informatika, dan S. Akba, "Rules Generation Untuk Klasifikasi Data Bakat Dan Minat Berdasarkan Rumpun Ilmu Dengan Decision Tree," *Jurnal Teknologi Informasi dan Komunikasi*, vol. 9, no. 1, hlm. 40–51, 2019.
- [15] A. Mujafar, M. Hanafi, dan M. Maimunah, "Grouping community reading interests using the k-means clustering method (case study: Magelang district library and archive service)," *Barobudur Informatics Review*, vol. 2, no. 2, hlm. 91–102, Des 2022, doi: 10.31603/binr.6810.

