

Classification of Career Interests of High School Students Using the C4.5 Algorithm

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

Dina Nofitasary¹⁾, Yunianita Rahmawati ^{*,2)}

¹⁾Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

²⁾ Program Studi Informatika, Universitas Muhammadiyah Sidoarjo, Indonesia

*Email Penulis Korespondensi: yunianita@umsida.ac.id

Abstract. Many SMA students experience confusion in determining their next career interests, from not understanding what they have chosen to the influence of the environment in determining their career. So, the researcher aims to help the school convince students in choosing their career interests at SMA Negeri 1 Wringinanom, using SMA student career interest data which is processed using the C4.5 algorithm so that it is more targeted according to the students' interests and talents. A total of 2 classes of secondary data for students in grades X and XI with a total of 76 data with the variables gender, major, academic achievement, non-academic achievement, hobbies, attendance, warning letters, career. Where the data is divided into 4 with accuracy, namely choosing to work 80%, choosing to study 72.2%, choosing to marry 50% and choosing official service 80%. So students who choose work and officialdom are confident in their choice, those who choose college are not too sure, and those who are most unsure about the choice they make are civil servants with the lowest percentage of confidence.

Keywords - classification; career interests; SMA; Algorithm C4.5

Abstrak. Banyak siswa SMA yang mengalami kebingungan menentukan minat karir selanjutnya, mulai ketidak pahaman akan yang dipilih hingga pengaruh lingkungan dalam menentukan karir. Maka, peneliti bertujuan untuk membantu pihak sekolah meyakinkan siswa dalam memilih minat karir siswa di SMA Negeri 1 Wringinanom, menggunakan data minat karir siswa SMA yang diproses menggunakan algoritma C4.5 sehingga lebih terarah sesuai dengan minat dan bakat dari siswa. Sebanyak 2 kelas dari data sekunder siswa kelas X dan XI dengan keseluruhan jumlah data sebanyak 76 data dengan variabel jenis kelamin, jurusan, prestasi akademik, prestasi non akademik, hobi, kehadiran, surat peringatan, berkarir. Dimana data dibagi menjadi 4 dengan akurasi yakni memilih bekerja 80%, memilih kuliah 72,2%, memilih menikah 50% dan memilih kedinasan 80%. Maka siswa yang memilih bekerja dan kedinasan yakin dengan pilihannya, yang memilih kuliah tidak terlalu yakin, dan yang paling tidak yakin dengan pilihan yang mereka ambil adalah kedinasan dengan persentase keyakinan terendah.

Kata Kunci - klasifikasi; minat karir; SMA; Algoritma C4.5

I. PENDAHULUAN

Dokumen ini adalah petunjuk penulis dan template artikel yang baru untuk UMSIDA Preprints Server. Setiap artikel yang dikirimkan ke redaksi UMSIDA Preprints Server harus mengikuti petunjuk penulisan ini. Jika artikel tersebut tidak sesuai dengan panduan ini maka tulisan akan dikembalikan.

UMSIDA Preprints Server dikelola oleh Universitas Muhammadiyah Sidoarjo. UMSIDA Preprints Server menerima manuskrip atau artikel ilmiah TA/Skripsi/Tesis dari mahasiswa Universitas Muhammadiyah Sidoarjo. Artikel-artikel yang dimuat di UMSIDA Preprints Server adalah artikel yang telah melalui proses penelaahan oleh Dewan Penguji. Keputusan diterima atau tidaknya suatu artikel ilmiah di preprint server ini menjadi hak dari Moderator berdasarkan atas rekomendasi dari Dewan Penguji. Banyaknya siswa SMA saat ini yang mengalami kebingungan dalam menentukan minat karir menuju jenjang selanjutnya[1], mulai dari ketidak pahaman 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.

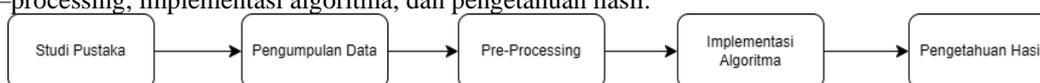
Berdasarkan data statistik Kementerian Pendidikan dan Kebudayaan pada tahun 2021, sebanyak 1,3 juta dari 3,6 juta siswa lulus yang melanjutkan studi ke perguruan tinggi[4]. Tidak sedikit yang memilih hanya menganggur disebabkan tidak yakinnnya dengan minat karir yang di fokuskan dengan bakat masing-masing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

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].

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, kedinasan, 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.

II. 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

A. Studi Pustaka

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

B. 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 dikelompokkan 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.

C. 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 :

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.

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 teretntu.

Encode Categorical Data

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

D. 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 adalah tahapan yang harus dilakukan :

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.

Perhitungan Total Entropy

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

$$\text{Entropy}(S) = - \sum_{i=1}^c p_i \cdot \log^2(p_i) \dots\dots\dots 1$$

Keterangan :

S : Himpunan Kasus.

n : Jumlah partisi atribut.S

pi : Proporsi dari Si terhadap S

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.:

$$\text{Entropy}(S) = - \sum_{i=1}^c p_i \cdot \log^2(p_i) \dots\dots\dots 2$$

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.

Entropy (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.

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.

E. 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 kuisioner 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 :

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

III. HASIL DAN PEMBAHASAN

A. Hasil pembahasan

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.

B. Pengujian program

Dari hasil kuisioner yang dilakukan mendapatkan sebesar 76 data. yang didapat dari 2 kelas, yaitu kelas X dan XI.

Tabel 1. Data Hasil Kuisioner

Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir	Pilihan
Laki-Laki	Ips	Tidak	Tidak	Seni	Tidak	Ya	Yakin Tidak	Bekerja
Laki-Laki	Ips	Tidak	Tidak	Fisik	Ya	Ya	Yakin	Bekerja
Perempuan	Ips	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Bekerja
Perempuan	Ips	Tidak	Ya	Fisik	Ya	Tidak	Yakin	Bekerja
Laki-Laki	Ipa	Tidak	Tidak	Fisik	Tidak	Tidak	Yakin	Bekerja
Perempuan	Ips	Tidak	Tidak	Fisik	Ya	Ya	Yakin Tidak	Bekerja
Laki-Laki	Ips	Ya	Ya	Fisik	Tidak	Tidak	Yakin Tidak	Bekerja
Perempuan	Ips	Tidak	Tidak	Seni	Ya	Tidak	Yakin	Bekerja
Perempuan	Ips	Tidak	Tidak	Fisik	Ya	Ya	Yakin	Bekerja
Laki-Laki	Ipa	Tidak	Ya	Fisik	Ya	Tidak	Yakin Tidak	Bekerja
Perempuan	Ips	Tidak	Tidak	Seni	Ya	Tidak	Yakin	Bekerja
Perempuan	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Bekerja
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Ya	Fisik	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Ya	Ya	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Ya	Ya	Fisik	Tidak	Tidak	Tidak	Kuliah
Perempuan	Ipa	Ya	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ips	Ya	Ya	Seni	Tidak	Tidak	Yakin	Kuliah
Laki-Laki	Ips	Ya	Tidak	Seni	Tidak	Ya	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ips	Ya	Ya	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ips	Ya	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Fisik	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ips	Ya	Ya	Seni	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Ya	Tidak	Seni	Ya	Tidak	Yakin	Kuliah
Perempuan	Ips	Ya	Tidak	Fisik	Ya	Ya	Yakin	Kuliah
Laki-Laki	Ipa	Ya	Tidak	Seni	Ya	Tidak	Yakin	Kuliah
Perempuan	Ips	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kuliah
Perempuan	Ips	Tidak	Tidak	Seni	Ya	Ya	Yakin	Kuliah
Perempuan	Ips	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Ya	Tidak	Seni	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Tidak	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah

Perempuan	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Ya	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Laki-Laki	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kuliah
Laki-Laki	Ipa	Tidak	Ya	Fisik	Ya	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Tidak	Kuliah
Laki-Laki	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Laki-Laki	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Tidak	Kuliah
Perempuan	Ipa	Tidak	Ya	Fisik	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Ya	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Ya	Tidak	Fisik	Tidak	Tidak	Yakin	Kuliah
Laki-Laki	Ipa	Tidak	Tidak	Fisik	Tidak	Tidak	Yakin	Kuliah
Laki-Laki	Ipa	Tidak	Tidak	Seni	Tidak	Ya	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Ya	Tidak	Tidak	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Laki-Laki	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kuliah
Laki-Laki	Ipa	Tidak	Ya	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Kuliah
Perempuan	Ips	Tidak	Ya	Seni	Tidak	Tidak	Yakin	Menikah
Laki-Laki	Ipa	Tidak	Ya	Fisik	Ya	Ya	Tidak	Menikah
Perempuan	Ipa	Tidak	Tidak	Seni	Tidak	Tidak	Yakin	Menikah
Perempuan	Ips	Tidak	Tidak	Fisik	Ya	Ya	Yakin	Menikah
Perempuan	Ipa	Tidak	Ya	Fisik	Tidak	Tidak	Tidak	Menikah
Laki-Laki	Ips	Tidak	Tidak	Seni	Ya	Ya	Yakin	Menikah
Perempuan	Ipa	Ya	Ya	Seni	Tidak	Tidak	Yakin	Menikah
Laki-Laki	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Menikah
Laki-Laki	Ipa	Ya	Ya	Fisik	Tidak	Tidak	Yakin	Kedinasan
Laki-Laki	Ipa	Ya	Tidak	Seni	Tidak	Tidak	Yakin	Kedinasan
Perempuan	Ips	Ya	Tidak	Seni	Tidak	Tidak	Yakin	Kedinasan
Laki-Laki	Ipa	Ya	Tidak	Fisik	Tidak	Tidak	Tidak	Kedinasan
Laki-Laki	Ipa	Ya	Tidak	Seni	Tidak	Tidak	Yakin	Kedinasan
Laki-Laki	Ipa	Ya	Tidak	Fisik	Tidak	Tidak	Tidak	Kedinasan
Perempuan	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kedinasan
Laki-Laki	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kedinasan
Laki-Laki	Ipa	Ya	Tidak	Fisik	Ya	Tidak	Yakin	Kedinasan
Perempuan	Ipa	Ya	Tidak	Seni	Tidak	Tidak	Tidak	Kedinasan
Laki-Laki	Ipa	Tidak	Tidak	Fisik	Ya	Tidak	Yakin	Kedinasan

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.

Table 2. Klasifikasi Minat Karir

Data Karir	Minat	Jumlah
Kuliah		45
Bekerja		12
Menikah		8
Kedinasan		11

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.

```
#Membaca Dataset dari File ke Pandas dataframe
Dataset = pd.read_csv('Masukan_Bekerja.csv', delimiter=';', header=0)

Dataset.head()
```

	Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir
0	Laki-Laki	IPS	TIDAK	TIDAK	Seni	TIDAK	YA	YAKIN
1	Laki-Laki	IPS	TIDAK	TIDAK	Fisik	YA	YA	TIDAK YAKIN
2	Perempuan	IPS	TIDAK	TIDAK	Seni	TIDAK	TIDAK	YAKIN
3	Perempuan	IPS	TIDAK	YA	Fisik	YA	TIDAK	YAKIN
4	Laki-Laki	IPA	TIDAK	TIDAK	Fisik	TIDAK	TIDAK	YAKIN

Gambar 2. Proses Input Data Bekerja

```
#Membaca Dataset dari File ke Pandas dataframe
Dataset = pd.read_csv('Masukan_Kuliah.csv', delimiter=';', header=0)

Dataset.head()
```

	Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir
0	Perempuan	IPA	TIDAK	TIDAK	Seni	TIDAK	TIDAK	YAKIN
1	Perempuan	IPA	TIDAK	YA	Fisik	YA	TIDAK	YAKIN
2	Perempuan	IPA	YA	YA	Seni	TIDAK	TIDAK	YAKIN
3	Perempuan	IPA	YA	YA	Fisik	TIDAK	TIDAK	TIDAK
4	Perempuan	IPA	YA	TIDAK	Seni	TIDAK	TIDAK	YAKIN

Gambar 3. Proses Input Data Kuliah

```
#Membaca Dataset dari File ke Pandas dataframe
Dataset = pd.read_csv('Masukan_Menikah.csv', delimiter=';', header=0)

Dataset.head()
```

	Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir
0	Perempuan	IPS	TIDAK	YA	Seni	TIDAK	TIDAK	YAKIN
1	Laki-Laki	IPA	TIDAK	YA	Fisik	YA	YA	TIDAK
2	Perempuan	IPA	TIDAK	TIDAK	Seni	TIDAK	TIDAK	YAKIN
3	Perempuan	IPS	TIDAK	TIDAK	Fisik	YA	YA	YAKIN
4	Perempuan	IPA	TIDAK	YA	Fisik	TIDAK	TIDAK	TIDAK

Gambar 4. Proses Input Data Menikah

```
#Membaca Dataset dari File ke Pandas dataframe
Dataset = pd.read_csv('Masukan_Kedinasan.csv', delimiter=';', header=0)

Dataset.head()
```

	Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir
0	Laki-Laki	IPA	YA	YA	Fisik	TIDAK	TIDAK	YAKIN
1	Laki-Laki	IPA	YA	TIDAK	Seni	TIDAK	TIDAK	YAKIN
2	Perempuan	IPS	YA	TIDAK	Seni	TIDAK	TIDAK	YAKIN
3	Laki-Laki	IPA	YA	TIDAK	Fisik	TIDAK	TIDAK	TIDAK
4	Laki-Laki	IPA	YA	TIDAK	Seni	TIDAK	TIDAK	YAKIN

Gambar 5. Proses Input Data Kedinasan

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

```
# Mengubah kelas (kolom "Berkarir") dari String ke Unique-Integer
Dataset["Berkarir"] = pd.factorize(Dataset["Berkarir"])[0]

Dataset.head()
```

	Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir
0	Laki-Laki	IPS	TIDAK	TIDAK	Seni	TIDAK	YA	0
1	Laki-Laki	IPS	TIDAK	TIDAK	Fisik	YA	YA	1
2	Perempuan	IPS	TIDAK	TIDAK	Seni	TIDAK	TIDAK	0
3	Perempuan	IPS	TIDAK	YA	Fisik	YA	TIDAK	0
4	Laki-Laki	IPA	TIDAK	TIDAK	Fisik	TIDAK	TIDAK	0

Gambar 6. Mengubah bilangan Data Bekerja

```
# Mengubah kelas (kolom "Berkarir") dari String ke Unique-Integer
Dataset["Berkarir"] = pd.factorize(Dataset["Berkarir"])[0]

Dataset.head()
```

	Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir
0	Perempuan	IPA	TIDAK	TIDAK	Seni	TIDAK	TIDAK	0
1	Perempuan	IPA	TIDAK	YA	Fisik	YA	TIDAK	0
2	Perempuan	IPA	YA	YA	Seni	TIDAK	TIDAK	0
3	Perempuan	IPA	YA	YA	Fisik	TIDAK	TIDAK	1
4	Perempuan	IPA	YA	TIDAK	Seni	TIDAK	TIDAK	0

Gambar 7. Mengubah bilangan Data Kuliah

```
# Mengubah kelas (kolom "Berkarir") dari String ke Unique-Integer
Dataset["Berkarir"] = pd.factorize(Dataset["Berkarir"])[0]

Dataset.head()
```

	Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir
0	Perempuan	IPS	TIDAK	YA	Seni	TIDAK	TIDAK	0
1	Laki-Laki	IPA	TIDAK	YA	Fisik	YA	YA	1
2	Perempuan	IPA	TIDAK	TIDAK	Seni	TIDAK	TIDAK	0
3	Perempuan	IPS	TIDAK	TIDAK	Fisik	YA	YA	0
4	Perempuan	IPA	TIDAK	YA	Fisik	TIDAK	TIDAK	1

Gambar 8. Mengubah bilangan Data Menikah

```
# Mengubah kelas (kolom "Berkarir") dari String ke Unique-Integer
Dataset["Berkarir"] = pd.factorize(Dataset["Berkarir"])[0]

Dataset.head()
```

	Jenis Kelamin	Jurusan	Prestasi Akademik	Prestasi Non Akademik	Hobi	Absensi	Surat Peringatan	Berkarir
0	Laki-Laki	IPA	YA	YA	Fisik	TIDAK	TIDAK	0
1	Laki-Laki	IPA	YA	TIDAK	Seni	TIDAK	TIDAK	0
2	Perempuan	IPS	YA	TIDAK	Seni	TIDAK	TIDAK	0
3	Laki-Laki	IPA	YA	TIDAK	Fisik	TIDAK	TIDAK	1
4	Laki-Laki	IPA	YA	TIDAK	Seni	TIDAK	TIDAK	0

Gambar 9. Mengubah bilangan Data Kedinasan

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%.

```
# Data yang akan dibagi
X = np.random.rand(12, 8) # contoh data, baris, fitur
y = np.random.randint(2, size=12) # contoh target, 0 atau 1, baris

# Membagi dataset menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=1)

# Membuat objek DecisionTreeClassifier
clf = DecisionTreeClassifier(random_state=1)

# Melihat ukuran data latih dan data uji
print("Ukuran data latih:", X_train.shape)
print("Ukuran data uji:", X_test.shape)

Ukuran data latih: (7, 8)
Ukuran data uji: (5, 8)

print("Data Latih : ", X_train)

Data Latih : [[0.70185991 0.09924104 0.121012 0.23696094 0.41147589 0.45046865
0.74173033 0.42945949]
[0.89910513 0.62003433 0.24000902 0.72942379 0.39675748 0.88745804
0.34177136 0.44810829]
[0.56889377 0.59360828 0.02394007 0.31604924 0.24843596 0.14870889
0.16006773 0.36562334]
[0.08174013 0.47904523 0.71720022 0.64170743 0.40816487 0.69600381
0.121012 0.23696094]
[0.0656306 0.36423791 0.45217097 0.10266314 0.43108446 0.89318141]]

print("Data Uji : ", X_test)

Data Uji : [[0.12909188 0.80441774 0.99886974 0.25414168 0.38770226 0.64679267
0.99396934 0.63137718]
[0.68311086 0.98711906 0.70824417 0.43945883 0.92808755 0.8306523
0.84558783 0.5457378 ]
[0.17857559 0.49365379 0.54098581 0.87396775 0.5100722 0.0330083
0.72924505 0.90038959]
[0.0656306 0.36423791 0.45217097 0.10266314 0.43108446 0.89318141]]
```

Gambar 10. Membagi Testing dan Training Data Bekerja

```
# Data yang akan dibagi
X = np.random.rand(45, 8) # contoh data, baris, fitur
y = np.random.randint(2, size=45) # contoh target, 0 atau 1, baris

# Membagi dataset menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=1)

# Membuat objek DecisionTreeClassifier
clf = DecisionTreeClassifier(random_state=1)

# Melihat ukuran data latih dan data uji
print("Ukuran data latih:", X_train.shape)
print("Ukuran data uji:", X_test.shape)

Ukuran data latih: (27, 8)
Ukuran data uji: (18, 8)
```

```

print("Data Latih : ",X_train)

Data Latih : [[0.88256619 0.31139595 0.79567971 0.9334648 0.92494354 0.31688106
0.79208088 0.66373386]
[0.86824472 0.53258571 0.22841572 0.00561383 0.45263872 0.45008713
0.29500655 0.80720716]
[0.95419335 0.41348539 0.71514908 0.27084065 0.85569051 0.13591219
0.09849466 0.5192087 ]
[0.82858466 0.61331118 0.65108376 0.20797435 0.91448663 0.8462969
0.33426021 0.40009618]
[0.95965524 0.34375865 0.5652416 0.92548136 0.43560514 0.39810738]

print("Data Uji : ",X_test)

Data Uji : [[7.84698256e-01 4.85516882e-01 5.43385731e-01 5.56181048e-01
6.32462825e-01 5.25633188e-01 4.46520085e-01 1.15464251e-01]
[9.18260289e-02 7.81667747e-01 1.60360403e-01 3.60511521e-01
1.54153034e-01 9.33656655e-01 3.31160026e-01 7.83707812e-01]
[8.67255233e-01 4.87368285e-01 5.09397392e-01 6.06721294e-01
9.55119196e-01 5.64555447e-01 1.09801801e-01 5.95609413e-01]
[3.82642902e-01 7.37205328e-01 9.82353917e-01 4.97275407e-01
7.83046317e-01 7.37425784e-01 7.41561200e-01 8.30823313e-01]

```

Gambar 11. Membagi Testing dan Training Data Kuliah

```

# Data yang akan dibagi
X = np.random.rand(8, 8) # contoh data, baris, fitur
y = np.random.randint(2, size=8) # contoh target, 0 atau 1, baris

# Membagi dataset menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=1)

# Membuat objek DecisionTreeClassifier
clf = DecisionTreeClassifier(random_state=1)

# Melihat ukuran data latih dan data uji
print("Ukuran data latih:", X_train.shape)
print("Ukuran data uji:", X_test.shape)

Ukuran data latih: (4, 8)
Ukuran data uji: (4, 8)

print("Data Latih : ",X_train)

Data Latih : [[0.00303295 0.38285904 0.41824838 0.91881907 0.60356628 0.71626689
0.09427345 0.50391187]
[0.30157232 0.07691754 0.87080963 0.81000378 0.32488494 0.69086441
0.31664073 0.01077328]
[0.99030233 0.71342728 0.12820374 0.40536679 0.53677386 0.15496292
0.39184603 0.41465755]
[0.35130495 0.79620683 0.35251369 0.68641553 0.97863119 0.09815998
0.88742682 0.09677143]]

print("Data Uji : ",X_test)

Data Uji : [[0.09149596 0.66944646 0.9381699 0.71146626 0.19247296 0.48118341
0.76030607 0.48245267]
[0.22368227 0.43553442 0.35028196 0.52597192 0.32884179 0.07617762
0.50304357 0.09998383]
[0.59074602 0.19968995 0.72994533 0.39078938 0.51482529 0.92047264
0.86328233 0.11892047]
[0.75559877 0.03972269 0.72381056 0.84003481 0.59535515 0.26317481
0.65467225 0.72313455]]

```

Gambar 11. Membagi Testing dan Training Data Menikah

```

# Load dataset
iris = load_iris()
X = iris.data
y = iris.target

# Data yang akan dibagi
X = np.random.rand(11, 8) # contoh data, baris, fitur
y = np.random.randint(2, size=11) # contoh target, 0 atau 1, baris

# Membagi dataset menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.4, random_state=1)

# Membuat objek DecisionTreeClassifier
clf = DecisionTreeClassifier(random_state=1)

# Melihat ukuran data latih dan data uji
print("Ukuran data latih:", X_train.shape)
print("Ukuran data uji:", X_test.shape)

Ukuran data latih: (6, 8)
Ukuran data uji: (5, 8)

```

```
print("Data Latih : ",X_train)

Data Latih : [[0.80514768 0.65953685 0.92693744 0.35027262 0.69355784 0.13296324
0.54262033 0.05807462]
[0.16217389 0.93583851 0.38795826 0.56344351 0.65999972 0.16591225
0.83664599 0.82735706]
[0.39931435 0.79512072 0.71177645 0.16197655 0.17721652 0.40709912
0.39492945 0.48844043]
[0.87487521 0.66657608 0.66809374 0.84459729 0.2124019 0.64661052
0.68313568 0.08834467]
[0.04448988 0.9962549 0.07338244 0.07554693 0.87004742 0.28026277
0.53740071 0.93282522]
[0.58382923 0.36501127 0.9021119 0.05104215 0.79255824 0.13033651
0.74516355 0.80896756]]

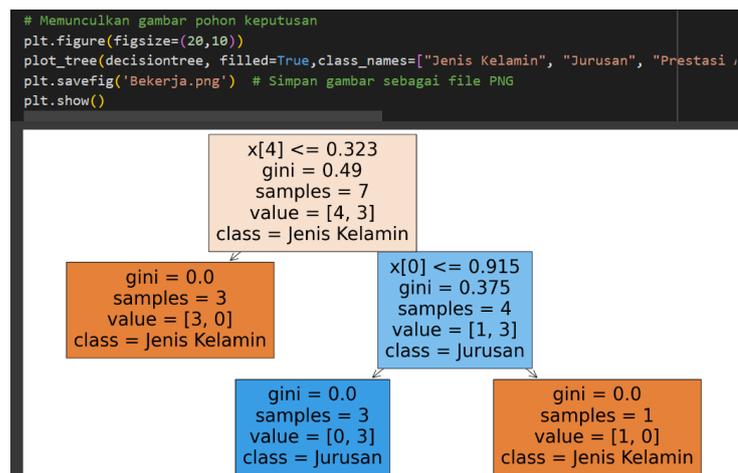
print("Data Uji : ",X_test)

Data Uji : [[0.64584086 0.37502734 0.43093612 0.18063327 0.84847395 0.31422849
0.48530221 0.97875316]
[0.12687532 0.62964807 0.53967311 0.98921209 0.78846403 0.76150437
0.60576436 0.15890381]
[0.25558003 0.5644956 0.11306038 0.72357179 0.46434675 0.35548579
0.06396673 0.71056282]
[0.73544827 0.32441218 0.1539399 0.67711372 0.13972671 0.42743417
0.76639163 0.59109087]
[0.29157458 0.67954479 0.34507974 0.78799108 0.65424651 0.9018983
0.69234196 0.52970184]]
```

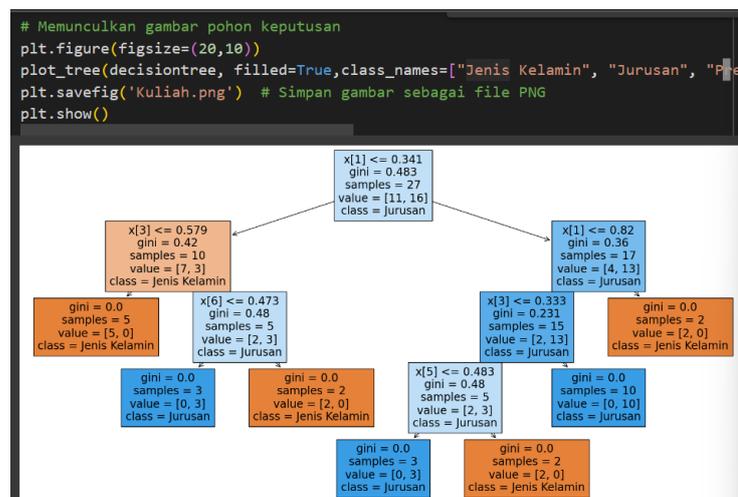
Gambar 12. Membagi Testing dan Training Data Kedinasan

4) Pohon Keputusan

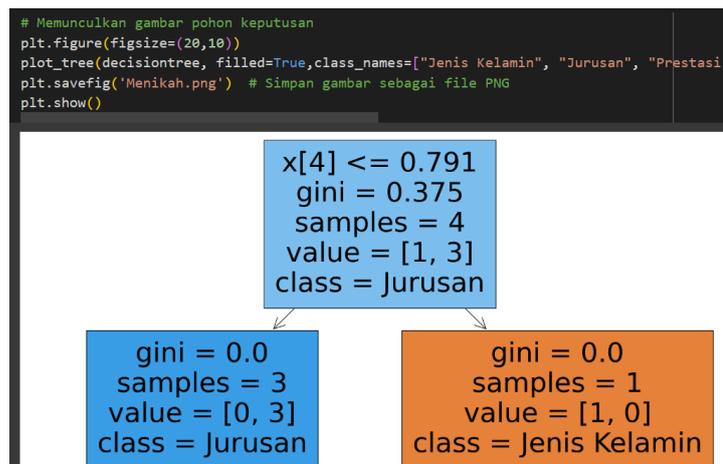
Pada bagian ini akan memunculkan pohon Keputusan yang berasal dari data yang diinputkan sebelumnya.



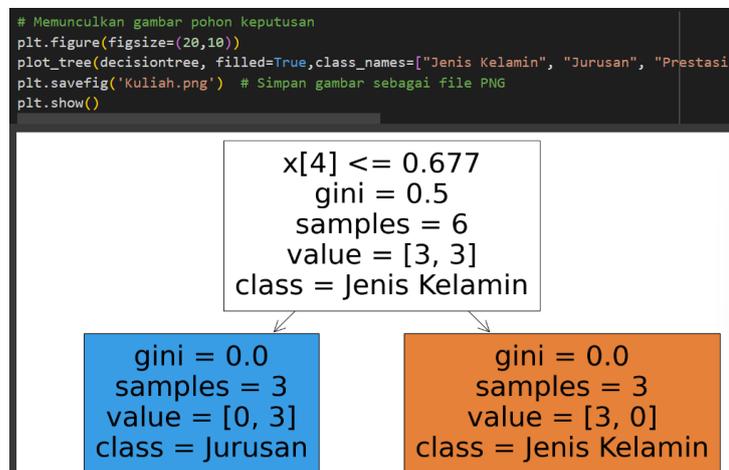
Gambar 13. Hasil Pohon Keputusan Bekerja



Gambar 14. Hasil Pohon Keputusan Kuliah



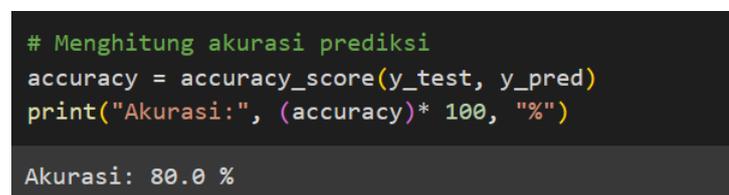
Gambar 15. Hasil Pohon Keputusan Menikah



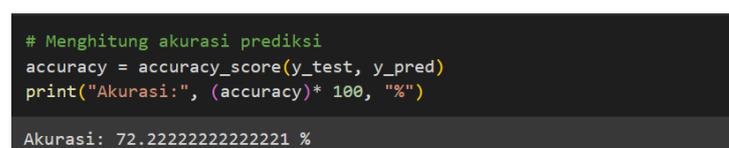
Gambar 16. Hasil Pohon Keputusan Kedinasan

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. Berikut adalah hasil akurasi yang didapat :



Gambar 17. Hasil Persentase Bekerja



Gambar 18. Hasil Persentase Kuliah

```
# Menghitung akurasi prediksi
accuracy = accuracy_score(y_test, y_pred)
print("Akurasi:", (accuracy)* 100, "%")
```

Akurasi: 50.0 %

Gambar 18. Hasil Persentase Menikah

```
# Menghitung akurasi prediksi
accuracy = accuracy_score(y_test, y_pred)
print("Akurasi:", (accuracy)* 100, "%")
```

Akurasi: 80.0 %

Gambar 19. Hasil Persentase Kedinasan

Table 3. Hasil Persentase Tingkat Akurasi

Data	Minat	Jumlah	Persentase
Karir			
Bekerja		12	80%
Kuliah		45	72,2%
Menikah		8	50%
Kedinasan		11	80%

Diketahui bahwa pada data diatas dilakukan implementasi *Google Colab* dengan algoritma Decision Tree C4.5 menggunakan program *Python* memiliki 76 data dengan 8 variabel yang dibagi menjadi 4, yaitu bekerja, kuliah, menikah, dan kedinasan.

IV. SIMPULAN

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 ketidak sesuaian karir sesuai minat dan bakat siswa.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada orang tua yang telah memberikan dukungan baik secara doa maupun materi dan kepada teman-teman yang telah memberikan dukungan dan nasihat sehingga dapat menyelesaikan tugas akhir.

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 Padi, "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.stkipyapisdampu.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)," *Borobudur Informatics Review*, vol. 2, no. 2, hlm. 91–102, Des 2022, doi: 10.31603/binr.6810.

Conflict of Interest Statement:

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.