BRAHMANA: Jurnal Penerapan Kecerdasan Buatan
Terakreditasi Nomor 204/E/KPT/2022 | Vol. 4, No. 2, Juni (2023), pp. 138-143

Penerapan Algoritma *Random Forest* Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo

Haiza Marlina[1], Elmayati[2], Antoni Zulius[3], Harma Oktafia Lingga Wijaya[4]
[1,2,4]Program Studi Sistem Informasi, Fakultas Komputer, Universitas Bina Insan, Lubuk Linggau, Indonesia
[3]Program Studi Rekayasa Komputer, Fakultas Komputer, Universitas Bina Insan, Lubuk Linggau, Indonesia
E-mail: [1]haizamarlina18@gmail.com, [2]elmayati@univbinainsan.ac.id, [3]antoni.zulius@univbinainsan.ac.id, [4]harmaoktafialingga@gmail.com

Abstract

The problem in this study is the difficulty in determining the majors at Tugumulyo State High School, for that a classification model is needed that can determine the right majors in student majors at Tugumulyo State High School. This research uses data collection methods, by observing and recording directly at the research site (observation), asking questions directly to the source (interview), and documentation by reading literature guidelines. The results showed that the majors classification model using the Random Forest algorithm. It can be concluded that the model built using the R programming language gets an accuracy value of 0.9314 and is included in the good classification category.

Keyword: *Classification, Alignment, Random Forest*

Abstrak

Masalah pada penelitian ini adalah kesulitan dalam penentuan penjurusan pada SMA Negeri Tugumulyo, untuk itu dibutuhkan sebuah model klasifikasi yang dapat menentukan penjurusan yang tepat pada penjurusan siswa di SMA Negeri Tugumulyo. Penelitian ini menggunakan metode pengumpulan data, dengan cara melakukkan pengamatan dan pencatatan langsung pada tempat penelitian (observasi), melakukan tanya jawab langsung pada sumber (interview), dan dokumentasi dengan cara membaca pedoman-pedoman literature. Hasil penelitian menunjukan bahwa model klasifikasi penjurusan dengan menggunakan algoritma Random Forest. Dapat disimpulkan bahwa model yang dibangun menggunakan bahasa pemrograman R mendapatkan nilai akurasi sebesar 0.9314 dan masuk dalam kategori good classification.

Kata Kunci: *Klasifikasi, Penjurusan, Random Forest*

1. Pendahuluan

Perkembangan teknologi saat ini sudah mengalami kemajuan yang sangat pesat, salah satunya perkembangan teknologi pembelajaran saat ini sudah cukup terjangkau dari pada tahun-tahun yang lalu. Sekolah SMA Negeri Tugumulyo merupakan salah satu sekolah menengah atas yang berlokasi di Trikoyo, Kecamatan Tugumulyo, Kabupaten Musi Rawas Sumatera Selatan. setiap tahunnya menerima siswa baru yang akan melanjutkan pendidikan ke Sekolah Menengah Atas (SMA). SMA Negeri Tugumulyo saat ini memiliki 2 jurusan yaitu Ilmu Pengetahuan Alam (IPA) dan Ilmu Pengetahuan Sosial (IPS). Penjurusan siswa yang dilakukan pada SMA Negeri Tugumulyo dilakukan sejak peserta didik mendaftar ke SMA memiliki tujuan untuk memberikan kesempatan kepada peserta didik mengembangkan kompetensi. Penjurusan siswa di SMA Negeri Tugumulyo ditentukan oleh pihak sekolah bedasarkan minat siswa dan nilai siswa. Penentuan jurusan yang dilakukan oleh pihak sekolah dengan cara melihat data nilai semua siswa terlebih

ISSN: 2715-9906
Copyright © BRAHMANA: Jurnal Penerapan Kecerdasan Buatan

dahulu kemudian dicocokkan dengan minat siswa tersebut. Hal tersebut digunakan untuk menentukan klasifikasi siswa-siswi dalam penentuan jurusan/ peminatnya, walaupun siswa-siswi mempunyai keinginan tersendiri namun pihak sekolah SMA Negeri Tugumulyo harus menyalurkan sesuai dengan kemampuan siswa-siswinya agar tidak kesulitan dalam proses belajar mengajar kedepannya.

2. Metodologi Penelitian

2.1. Penerapan

Penerapan adalah perbuatan menerapkan,sedangkan menurut para ahli penerapan adalah suatu perbuatan menerapkan, sedangkan menurut beberapa ahli, penerapapan adalah suatu perbuatan mempraktekkan suatu teori, metode, dan hal lain untuk mencapai tujuan tertentu dan untuk suatu kepentingan yang diinginkan oleh suatu kelompok atau golongan yang telah terancam dan terususun sebelumnya[2].

2.2. Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data, dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui. Dalam mencapai tujuan tersebut, proses klasifikasi membentuk suatu modal yang mampu membedakan data kedalam kelas-kelas yang berdasarkan aturan atau fungsi tertentu. Model itu sendiri bisa beruba aturan “jika-maka”, berupa pohon keputusan, atau formule matematis[4].

2.3. Data Mining

Data mining disebut *juga Knowledge Discovery in Database* (KDD) sebagai ekstraksi informasi potensial, implisit dan tidak dikenaldari sekumpulan data. Proses *Knowledge Discovery In Database* melibatkan hasil proses data mining (proses pengekstrak kecenderung suatu pola data) kemudian mengubah hasilnya secara akurat menjadi informasi yang mudah dipahami[5].

2.4. Algoritma *Random Forest*

Random Forest adalah pengembangan dari metode *Decision Tree* yang menggunakan beberapa *Decision Tree*, dimana setiap *Decision Tree* telah dilakukan pelatihan menggunakan sampel individu dan setiap atribut dipecah pada pohon yang dipilih antara atribut subset yang bersifat acak. *Random Forest* memiliki beberapa kelebihan, yaitu dapat meningkatkan hasil akurasi jika terdapat data yang hilang, dan untuk resisting outliers, serta efisien untuk penyimpanan sebuah data. Selain itu, *Random Forest* mempunyai proses seleksi fitur dimana mampu mengambil fitur terbaik sehingga dapat meningkatkan performa terhadap model klasifikasi. Dengan adanya seleksi fitur tentu *Random Forest* dapat bekerja pada big data dengan parameter yang kompleks secara efektif [5].

2.5. Metode Penelitian

Metode penelitian yang digunakan peneliti dalam menganalisa, merancang dan memahami permasalahan yang akan dibahas dalam penelitian ini menggunakan metode penelitian kualitatif yang bersifat deskriptif, serta menganalisis permasalahan pada tempat penelitian dalam memperoleh data dan informasi yang dibutuhkan. Penelitian deskriptif kualitatif adalah mendeskripsikan seluruh gejala atau keadaan yang ada, yaitu keadaan gejala menurut apa adanya pada saat penelitian dilakukan[14].

2.6. Metode Pengembangan Sistem

Pada penelitian ini penulis memilih proses model CRISP-DM (*Cross Industry standard processfor data mining*). Model ini menjadi dasar teori dalam metode penelitian ini. Adapun langkah-langkah dalam penelitian yaitu:

a. Fase Bussines Understanding

Fase Business Understanding, data nilai siswa yang diperoleh dari Siswa Sekolah Menengah Atas Negeri Tugumulyo, data ini diperlukan untuk menggali data tersebut dengan menggunakan Algoritma *Random Forest* untuk menemukan pola penjurusan di SMA Negeri Tugumulyo.

Dalam melakukan analisis data dalam menentukan penjurusan siswa SMAN Tugumulyo. Data yang didapat dari tempat penelitian dipergunakan untuk pembentukan data *training* dan data *testing.* Berikut data atribut yang digunakan dalam penelitian ini terlihat pada Tabel 1.

Tabel 1. Tabel Atribut

Atribut	Keterangan
Nilai_biologi	Nilai Biologi
Nilai_fisika	Nilai Fisika
Nilai kimia	Nilai Kimia
Nilai_sosiologi	Nilai Sosiologi
Nilai sejarah	Nilai Sejarah
Nilai ekonomi	Nilai Ekonomi
Nilai matematika	Nilai Matematika
Nilai b.indo	Nilai Bahasa Indonesia

b. Fase Data Understanding

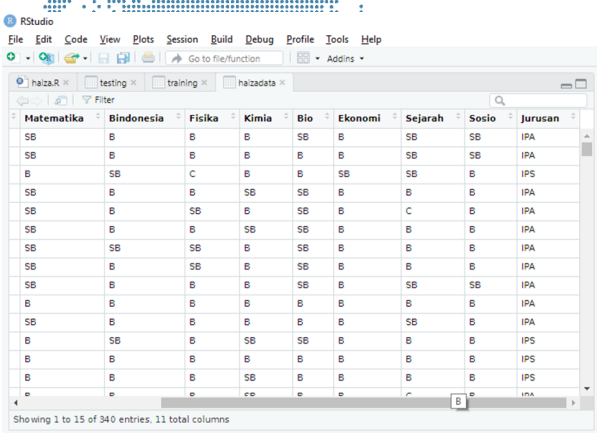
Fase Data Understanding, data yang akan diambil dari nilai siswa dari SMAN Tugumulyo antara lain nilai biologi, nilai kimia, nilai fisika, nilai sosiologi, nilai sejarah, nilai ekonomi, nilai matematika, nilai b.indonesia. Data siswa-siswi yang dikumpulkan merupakan data tes penjurusan yang telah dilaksanakan. Kriteria penjurusan terdiri dari nilai biologi, nilai nilai kimia, nilai nilai fisika, nilai sosiologi, nilai sejarah, nilai ekonomi, nilai matematika, dan nilai b.indonesia.

Tabel 2. Skala Penilaian

No	Skala Nilai	Nama Skala
1	90-100	Sangat Baik
2	70-89	Baik
3	50-69	Cukup
4	20-49	Kurang
5	0-19	Sangat Kurang

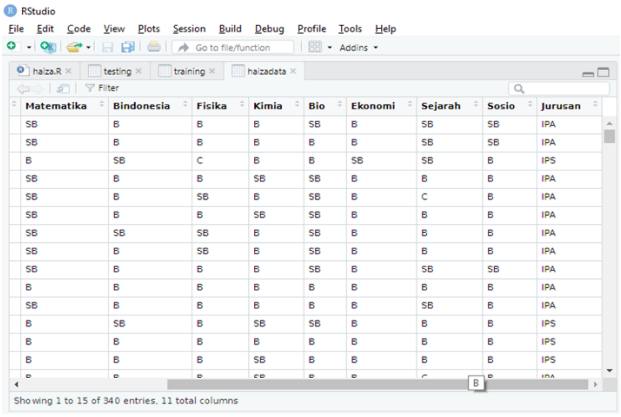
3. Hasil dan Pembahasan

Dari perancangan sistem maka hasil penelitian ini berupa sebuah model klasifikasi penjurusan pada SMA Negeri Tugumulyo menggunakan algoritma *Random Forest*, yang dapat membantu mengklasifikasi penjurusan pada SMA Negeri Tugumulyo yang akan di bahas dalam bab ini mengenai hasil klasifikasi untuk penentuan penjurusan pada SMA Negeri Tugumulyo. untuk dataset yang dibuat berjumlah 340 baris yang terdiri dari 11 kolom, yaitu kolom nomor, nama, jenis kelamin, matematika, bahasa Indonesia, fisika, kima, ekonomi, sejarah, sosiologi, dan kolom label penjurusan. Untuk lebih jelas bisa dilihat pada Gambar 1.



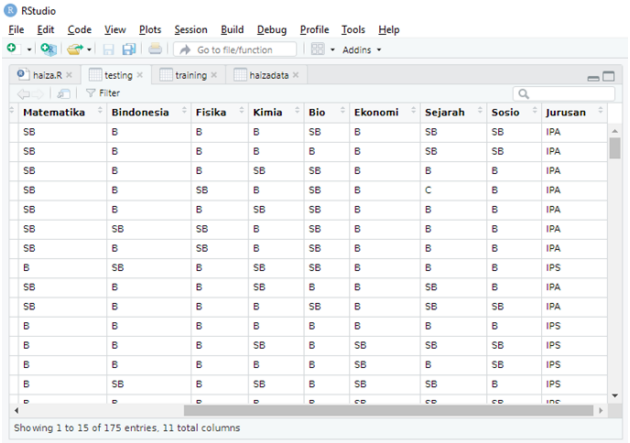
Gambar 1. Dataset Penjurusan

Pada Gambar 1 dapat dijelaskan bahwa menampilkan Dataset yang akan digunakan untuk penyusunan sebuah data training, dataset tersebut akan di *split* atau dibagi menjadi dua buah data yaitu data *training* dan data *testing*, pada penelitian ini akan digunakan data traning sebanyak 70% dan data *testing* sebanyak 30% dari keseluruhan dataset yang ada. Pada Gambar 2 ditampilkan data *traning* yang split sebanyak 70% dari seluruh dataset yang tersedia, untuk pembagian data ini akan diacak oleh sistem.



Gambar 2. Data Training

Selanjutnya Gambar 2 ditampilkan data testing yang telah displit dari dataset keseluruhannya menjadi 30 % dari dataset yaitu sebanyak 175 baris.



Gambar 3. Data Testing

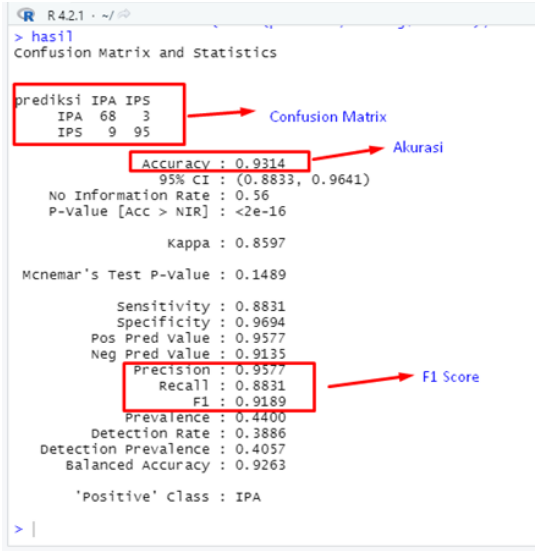
Untuk selanjutnya akan dilakukan prediksi terhadap model yang ada, pada Gambar 4 ditampilkan hasil prediksi dari model klasifikasi penjurusan pada SMA Negeri Tugumulyo menggunakan algoritma *Random Forest.*

```
               Accuracy : 0.9314
                 95% CI : (0.8833, 0.9641)
    No Information Rate : 0.56
    P-Value [Acc > NIR] : <2e-16

                  Kappa : 0.8597

 Mcnemar's Test P-Value : 0.1489

            Sensitivity : 0.8831
            Specificity : 0.9694
         Pos Pred Value : 0.9577
         Neg Pred Value : 0.9135
             Prevalence : 0.4400
         Detection Rate : 0.3886
   Detection Prevalence : 0.4057
      Balanced Accuracy : 0.9263

       'Positive' Class : IPA
```

Gambar 4. Hasil Klasifikasi

Pada Gambar 4 dapat dijelaskan bahwa hasil klasifikasi dari model yang dibangun mampu melakukan proses klasifikasi dengan tingkat akurasi tertinggi dengan percentage split sebesar 70% untuk data training dan 30 % untuk data testing dengan tingkat akurasi sebesar 0,9314 atau 93,14 %.

Pengujian yang dilakukan pada model klasifikasi penjurusan SMA Negeri Tugumulyo menggunakan confusion matrik. Hasil dari akurasi, *f1-score*, *recall* dapat dilihat pada Gambar 5.



Gambar 5. Pengujian Sistem

Dari pengujian ini didapat data akurasi sebesar 0.9314, F1 score sebesar 0,9189% dan recall 0,8831. klasifikasi dari sebuah metode dapat dianggap baik jika nilai akurasi dan nilai F1 memiliki nilai yang tinggi mendekati 1.

4. Kesimpulan

Berdasarkan hasil penelitian berupa model klasifikasi penjurusan pada SMAN Tugumulyo menggunakan algoritma *Random Forest*, maka dapat diambil kesimpulan bahwa dengan adanya model ini dapat membantu pihak sekolah didalam penentuan penjurusan pada SMA Negeri Tugumulyo. Tingkat akurasi dari model klasifikasi yang

telah dilakukan evaluasi dengan menggunakan Confusion Matrix didapat akurasi sebesar 93.14 % yang masuk dalam kategori *Good Classification*.

Dengan adanya penelitian ini, diharapkan adanya perbaikan maupun pengembangan pada model klasifikasi penjurusan pada SMAN Tugumulyo menggunakan algoritma *Random* seperti pengembangan Perlu ditambahkan lagi akurasi supaya lebih mendekati 100%. Dataset perlu ditambah untuk meningkatkan akurasi dari model yang akan dibagun kedepannya. Dataset perlu ditambah untuk meningkatkan akurasi dari model yang akan dibagun kedepannya.

Daftar Pustaka

[1] A. Z. Mafakhir And A. Solichin, "Penerapan Metode Naïve Bayes Classifier Untuk Penjurusan Siswa Pada Madrasah Aliyah Al-Falah Jakarta," *Fountain Informatics J.*, Vol. 5, No. 1, P. 21, 2020, Doi: 10.21111/Fij.V5i1.4007.

[2] A. Josi, "Penerapan Metode Prototyping Dalam Membangun Website Desa (Studi Kasus Desa Sugihan Kecamatan Rambang)," *Jti*, Vol. 9, No. 1, Pp. 50–57, 2017.

[3] C. Fadlan, S. Ningsih, And A. P. Windarto, "Penerapan Metode Naïve Bayes Dalam Klasifikasi Kelayakan Keluarga Penerima Beras Rastra," *J. Tek. Inform. Musirawas*, Vol. 3, No. 1, P. 1, 2018, Doi: 10.32767/Jutim.V3i1.286.

[4] 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.

[5] H. Leidiyana, "Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Resiko Kredit Kepemilikan Kendaraan Bermotor," *J. Penelit. Ilmu Komputer, Syst. Embed. Log.*, Vol. 1, No. 1, Pp. 65–76, 2013.

[6] R. Supriyadi, W. Gata, N. Maulidah, And A. Fauzi, "Penerapan Algoritma *Random Forest* Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. Dan Bisnis*, Vol. 13, No. 2, Pp. 67–75, 2020, Doi: 10.51903/E-Bisnis.V13i2.247.

[7] I. A. Ridlo, "Pedoman Pembuatan Flowchart," *Academia.Edu*, P. 27, 2017.

[8] M. S. Rossa A.S., *Rekayasa Perangkat Lunak*. 2018.

[9] U. Khultsum And A. Subekti, "Penerapan Algoritma *Random Forest* Dengan Kombinasi Ekstraksi Fitur Untuk Klasifikasi Penyakit Daun Tomat," *J. Media Inform. Budidarma*, Vol. 5, No. 1, P. 186, 2021, Doi: 10.30865/Mib.V5i1.2624.

[10] A. U. Zailani And N. L. Hanun, "Penerapan Algoritma Klasifikasi *Random Forest* Untuk Penentuan Kelayakan Pemberian Kredit Di Koperasi Mitra Sejahtera," *Infotech J. Technol. Inf.*, Vol. 6, No. 1, Pp. 7–14, 2020, Doi: 10.37365/Jti.V6i1.61.

[11] F. Mu'alim And R. Hiday, "Implementasi Metode *Random Forest* Untuk Penjurusan Siswa Di Madrasah Aliyah Negeri Sintang," *Jupiter*, Vol. 14, No. 1, Pp. 116–125, 2022, [Online]. Available: Https://Www.Neliti.Com/Publications/441871/Implementasi-Metode-Random-Forest-Untuk-Penjurusan-Siswa-Di-Madrasah-Aliyah-Nege#Cite.

[12] S. Mutrofin *Et Al.*, "Komparasi Kinerja Algoritma C4 . 5 , Gradient Boosting Trees , *Random Forest*s , Dan Deep Learning Pada Kasus Educational Data Performance Comparison Between Algorithms Of C4 . 5 , Gradient Boosting Trees , *Random Forest*s , And Deep Learning In," *J. Teknol. Inf. Dan Ilmu Komput.*, Vol. 7, No. 4, Pp. 807–814, 2020, Doi: 10.25126/Jtiik.2020732665.

[13] Bh. Hayadi, "Perbandingan Metode Naive Bayes Dan *Random Forest* Untuk Menentukan Prestasi Belajar Siswa Pada Jurusan Rpl (Studi Kasus Smk Swasta Siti Banun Sigambal)," *J. Comput. Sci. Inf. Technol. Progr. Stud. Teknol. Inf.*, No. 2, P. 2022, 2022, [Online]. Available: Http://Jurnal.Ulb.Ac.Id/Index.Php/Jcoint/Index.

[14] Mukhtar, *Metode Penelitian Deskriftif Kualitatif*. Jakarta: Gp Press Group, 2013.