

IMPLEMENTASI SELEKSI FITUR *INFORMATION GAIN RATIO* PADA ALGORITMA *RANDOM FOREST* UNTUK MODEL DATA KLASIFIKASI PEMBAYARAN KULIAH

Taghfirul Azhima Yoga Siswa, Wawan Joko Pranoto

Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, Universitas Muhammadiyah Kalimantan Timur

e-mail: tay758@umkt.ac.id, wjp337@umkt.ac.id

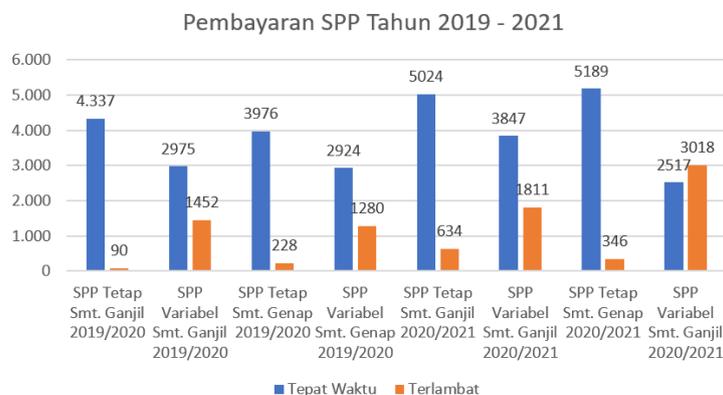
ABSTRAK

Perguruan tinggi swasta merupakan salah satu lembaga pendidikan dengan pembiayaan operasional utamanya dibebankan kepada mahasiswa melalui Sumbangan Pembangunan Pendidikan (SPP). Oleh karena itu, keterlambatan dari pembayaran SPP tentunya sangat berpengaruh signifikan bagi keberlangsungan aktivitas perguruan tinggi. Salah satu pendekatan data analisis yang bisa dilakukan adalah data mining dengan pemanfaatan algoritma *random forest*. *Random forest* merupakan metode yang biasa digunakan untuk menyelesaikan masalah yang berhubungan dengan klasifikasi dan regresi. Penelitian ini bertujuan mengimplementasikan dan mengevaluasi hasil kinerja algoritma *random forest* melalui pendekatan seleksi fitur *information gain ratio* dengan pembagian data training sebesar 90% dan data testing sebesar 10% dan dievaluasi menggunakan *confusion matrix* menggunakan rapidminer. Data yang digunakan dalam penelitian ini berupa data pembayaran SPP mahasiswa pada tahun 2019-2021 yang berasal dari Biro Administrasi Akademik dan Administrasi Keuangan yang berjumlah 12.408 *record* data. Hasil evaluasi kinerja seleksi fitur *information gain ratio* pada algoritma *random forest* menghasilkan nilai akurasi sebesar 78,47%. Atribut yang memiliki pengaruh dalam klasifikasi keterlambatan pembayaran SPP antara lain : fakultas, prodi, angkatan, gender penghasilan ayah, dan penghasilan ibu.

Kata Kunci: *data mining, klasifikasi, random forest, information gain ratio, pembayaran spp*

1. PENDAHULUAN

Universitas Muhammadiyah Kalimantan Timur (UMKT) merupakan salah satu amal usaha lembaga pendidikan yang dimiliki oleh persyarikatan muhammadiyah di provinsi Kalimantan Timur yang pembiayaan utama operasionalnya dibebankan kepada mahasiswa melalui Sumbangan Pembangunan Pendidikan (SPP). Oleh karena itu, keterlambatan dari pembayaran SPP tentunya sangat berpengaruh signifikan bagi keberlangsungan aktivitas perguruan tinggi. Berdasarkan data histori yang dikumpulkan, terjadi peningkatan keterlambatan pembayaran SPP yang signifikan di tahun 2020-2021 sebanyak 3.018 dari total keseluruhan 5.535 mahasiswa. Dampak dari keterlambatan pembayaran SPP ini akan menjadi kendala dalam mengalokasikan biaya operasional pembangunan perguruan tinggi seperti kegiatan akademik, pengajaran dosen dan karyawan dan lain-lain. Dari sisi mahasiswa, jika tidak melakukan pembayaran SPP sesuai waktu yang ditentukan, maka akibatnya mahasiswa tidak dapat mengambil mata kuliah (KRS), mencetak kartu ujian dan mengikuti proses perkuliahan semester selanjutnya.



Gambar 1. Pembayaran SPP di UMKT Tahun 2019-2021.

Dari masalah tersebut maka penting dilakukan sebuah analisis mendalam terkait analisis klasifikasi untuk dapat memberikan informasi atribut apa saja yang memiliki pengaruh pada label data pembayaran SPP. Salah satu teknik analisis yang dapat digunakan adalah *data mining*. Data mining merupakan aktivitas pengumpulan keterangan informasi dari dataset dengan menerapkan cara khusus dalam menentukan pengambilan keputusan [1]. Beberapa penelitian data mining sebelumnya terkait keterlambatan pembayaran SPP antara lain yaitu menggunakan algoritma C4.5 [2] [4]. Penelitian menggunakan metode *k-nearest neighbor* [3] dan penelitian yang menggunakan metode *naïve bayes* dengan *feature selection information gain* [5].

Algoritma *random forest* juga telah banyak digunakan dalam berbagai studi kasus salah satunya di bidang pendidikan antara lain tentang memprediksi jumlah mahasiswa baru menghasilkan tingkat akurasi sebesar 99,8% [6]. Penelitian tentang memprediksi ketidaktepatan waktu kelulusan mahasiswa Universitas Terbuka dengan hasil akurasi sebesar 93,23% [7]. Penelitian tentang perbandingan metode *decision tree*, *naïve bayes*, dan *random forest* untuk memprediksi prestasi siswa di SMKN 1 Gadingrejo Pringsewu Lampung dengan pengujian dengan metode *decision tree* menghasilkan akurasi sebesar 94,5%, *naïve bayes* sebesar 96,3%, dan *random forest* sebesar 97,3% [8]. Penelitian tentang memprediksi prestasi akademik mahasiswa menggunakan algoritma *random forest* dan C4.5 didapatkan pengujian metode C4.5 menghasilkan akurasi sebesar 87,1%, sedangkan metode *random forest* menghasilkan akurasi sebesar 92,4% [9]. Penelitian tentang memprediksi kelulusan mahasiswa teknik informatika UIN Suska Riau menghasilkan akurasi sebesar 92,59% [10]. Oleh karena itu dari penelitian sebelumnya, dapat disimpulkan bahwa algoritma *random forest* sangat mumpuni dalam melakukan prediksi di bidang pendidikan dengan memperoleh tingkat akurasi yang baik.

Penelitian keterlambatan pembayaran SPP pada dasarnya sudah banyak dilakukan dan metode yang digunakan berbagai macam seperti C4.5, *k-nearest neighbor*, dan *naïve bayes*. Namun untuk penelitian ini metode yang diusulkan adalah *random forest* dengan tambahan seleksi fitur *information gain ratio*. Beberapa penelitian sebelumnya telah menerapkan metode *information gain ratio* pada tahapan seleksi fitur pemodelannya memberikan peningkatan akurasi yang baik pada beberapa algoritma yang diterapkan. Pada penelitian diterapkan pendekatan heuristik menggunakan seleksi fitur *information gain ratio* untuk mendiagnosa penyakit jantung coroner [11]. Sementara pada penelitian [12] dilakukan penerapan seleksi fitur *information gain ratio* pada kasus klasifikasi penyakit ginjal kronis, dengan tujuan untuk meningkatkan akurasi pemodelan algoritma C4.5. Diharapkan dengan penerapan seleksi fitur *information gain ratio* ini nantinya dapat mengoptimasi akurasi algoritma *random forest* dalam menangani model kasus klasifikasi keterlambatan biaya kuliah pada penelitian ini.

2. TINJAUAN PUSTAKA

2.1. Sumbangan Pembinaan Pendidikan (SPP)

SPP adalah sumbangan pembinaan pendidikan (SPP) yang dibayarkan mahasiswa di perguruan tinggi swasta. Tujuan dari SPP adalah untuk mendanai kebutuhan pembelajaran sehingga kegiatan belajar mengajar dapat berlangsung dengan baik. SPP biasanya dibayarkan setiap semester oleh mahasiswa [13].

2.2. Data Mining

Menurut Hermawati [14], *data mining* adalah proses yang melakukan pekerjaan satu atau lebih teknik machine learning atau pembelajaran komputer yang dapat digunakan untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) yang dilakukan secara otomatis. Sementara pengertian data mining menurut Mardi [15] adalah sebuah proses yang memanfaatkan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* atau pembelajaran mesin untuk proses ekstraksi dan identifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database yang besar.

2.3. Klasifikasi

Menurut Hermawati [14], Klasifikasi merupakan proses pembelajaran suatu fungsi tujuan (target) f yang pada prosesnya memetakan dari setiap himpunan atribut x ke satu dari label kelas yang sebelumnya telah dilakukan proses pendefinisian. Klasifikasi adalah suatu cara untuk mengelompokkan benda berdasarkan ciri-ciri yang dimiliki oleh suatu objek klasifikasi, yang dalam prosesnya klasifikasi dapat dilakukan dengan berbagai cara baik secara manual maupun dengan bantuan teknologi [16].

2.4. Random Forest

Random Forest adalah salah satu algoritma dari *machine learning* untuk mengembangkan *decision tree*. *Random Forest* dapat dianggap sebagai kombinasi dari beberapa buah *decision tree* [17]. *Random Forest* merupakan salah satu bentuk yang berasal dari metode *ensemble* yang bertujuan untuk meningkatkan akurasi klasifikasi data dari sebuah pemilah tunggal yang tidak stabil melalui kombinasi dari

banyak jenis metode yang sama sebagai proses *majority voting* untuk menghasilkan prediksi tentang klasifikasi akhir [18]. Hasil prediksi dari algoritma *Random Forest* didapatkan melalui hasil terbanyak dari setiap individual *decision tree* (proses *voting* untuk klasifikasi dan rata-rata untuk regresi).

2.5. Confusion Matrix

Confusion Matrix adalah alat pengukuran yang melakukan pengukuran saat menganalisis klasifikasi. Pengklasifikasi baik dalam hal mengidentifikasi tupel dari kelas yang berbeda. Ketika mengklasifikasi dan memiliki data bernilai benar, maka nilai *True-Positive* dan *True-Negative* berfungsi untuk memberikan informasi tersebut. Sedangkan jika pengklasifikasi memiliki kesalahan saat mengklasifikasi data, maka nilai dari *False-Positive* dan *False-Negative* akan memberikan informasi tersebut [19].

Tabel 1. *Confusion Matrix*.

		True Values	
		True	False
Prediction	True	TP <i>Correct Result</i>	FP <i>Unexpected Result</i>
	False	FN <i>Missing Result</i>	TN <i>Correct absence of result</i>

Pada *confusion matrix* terdapat beberapa istilah yang digunakan pada kasus klasifikasi yaitu [20]:

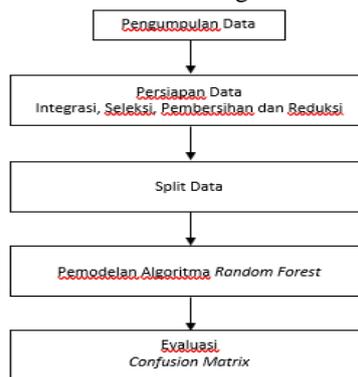
1. *True Positive (TP) confusion matrix* data positif yang terdeteksi benar
2. *False Positive (FP) confusion matrix* data negatif namun terdeteksi dengan benar
3. *False Negative (FN) confusion matrix* data positif yang terdeteksi sebagai data negatif
4. *True Negative (TN) confusion matrix* data negatif yang terdeteksi benar.

2.7. Rapidminer

RapidMiner merupakan perangkat lunak yang berfungsi sebagai alat pembelajaran dalam ilmu data mining. Aplikasi ini dikembangkan oleh perusahaan yang didedikasikan untuk semua langkah yang melibatkan sejumlah besar data dalam bisnis komersial, penelitian, pendidikan, pelatihan, dan pembelajaran. RapidMiner memiliki sekitar 100 solusi pembelajaran untuk pengelompokan, klasifikasi dan analisis regresi [21]. RapidMiner juga mendukung sekitar 22 format file, seperti .xls, .csv, dan sebagainya. RapidMiner dibangun untuk tim analisis, mengintegrasikan seluruh siklus ilmu data, dari persiapan data hingga pembelajaran mesin hingga penyebaran model prediksi.

3. METODE PENELITIAN

Alur atau tahapan dalam penelitian ini adalah sebagai berikut.



Gambar 2. Tahapan Penelitian.

3.1. Pengumpulan Data

Data yang dikumpulkan diperoleh dari dari Biro Administrasi Keuangan dan Bagian Administrasi Akademik pada tahun 2019-2021. Data dari Biro Administrasi Keuangan terdiri dari atribut nama, nim dan keterangan. Sedangkan data dari Biro Administrasi Akademik terdiri dari atribut nama, nim, gender, program studi, fakultas, angkatan, penghasilan ayah, penghasilan ibu, pendidikan ayah, dan pendidikan ibu.

Tabel 2. Atribut Data Mahasiswa dari Biro Administrasi Keuangan.

Atribut	Keterangan
---------	------------

Nama	Nama Mahasiswa
NIM	Nomor Induk Mahasiswa
Keterangan	Keterangan terlambat atau tepat waktu dalam membayar spp
Tabel 3. Atribut Data Mahasiswa dari Biro Administrasi Akademik.	
Atribut	Keterangan
Nama	Nama Mahasiswa
NIM	Nomor Induk Mahasiswa
Gender	Jenis Kelamin
Fakultas	Fakultas
Prodi	Program Studi
Angkatan	Tahun Angkatan
Penghasilan Ayah	Nominal Penghasilan Ayah
Penghasilan Ibu	Nominal Penghasilan Ibu
Pendidikan Ayah	Pendidikan Terakhir Ayah
Pendidikan Ibu	Pendidikan Terakhir Ibu

3.2. Persiapan Data

Dilakukan proses integrasi data, seleksi data, pembersihan dan reduksi data yang nantinya akan diperlukan dalam tahap permodelan dan evaluasi. Berikut ini adalah proses persiapan data yang akan dilakukan yaitu:

1. Integrasi Data

Menggabungkan 2 buah data dari Biro Administrasi Keuangan dan Administrasi Akademik yaitu dengan cara menyamakan nim sehingga menghasilkan atribut yang terdiri dari nama, nim, fakultas, prodi, angkatan, pendidikan ayah, pendidikan ibu, penghasilan ayah, penghasilan ibu dan keterangan.

2. Seleksi Data

Melakukan pemilihan atribut relevan yang akan digunakan dalam penelitian, atribut yang tidak relevan akan dihapus dan tidak akan digunakan. Dalam penelitian ini dilakukan seleksi fitur dengan *information gain ratio*.

3. Pembersihan Data

Menghapus data yang tidak lengkap atau kosong dan error pada atribut penghasilan ayah, penghasilan ibu, pendidikan ayah dan pendidikan ibu

4. Reduksi Data

Pengurangan data terhadap label kelas yang datanya tidak seimbang, dimana kelas tepat lebih banyak daripada kelas terlambat, sehingga perlu dilakukan reduksi dimensi data agar tidak terjadi klasifikasi lebih condong ke kelas mayoritas dibandingkan dengan kelas minoritas menggunakan teknik random under sampling.

3.3. Split Data

Dilakukan dengan membagi dataset menjadi dua bagian untuk data training sebesar 90% dan untuk data testing sebesar 10% yang nantinya dipakai untuk proses pengujian algoritma.

3.4. Pemodelan Algoritma *Random Forest*

Pemodelan dilakukan dengan memanfaatkan aplikasi RapidMiner. Proses diawali dengan mengambil data training yang telah disediakan sebelumnya dengan cara melakukan klik pada tab Import Data, dan pilih pada local folder yang menyimpan data tersebut. Kemudian lakukan pemodelan dengan memanggil fungsi algoritma *random forest* melalui fasilitas operator.

3.5. Evaluasi

Teknik *confusion matrix* akan dilakukan untuk mengukur keakuratan dan kualitas data latih (*training*) yang dilakukan oleh pemodelan dari rasio data latih yang diberikan. Adapun pengukuran evaluasi dari *confusion matrix* yang digunakan adalah *Accuracy*.

$$accuracy = \frac{TP + TN}{n}$$

4. HASIL DAN PEMBAHASAN

4.1. Data Penelitian

Data pembayaran SPP diperoleh dari Biro Administrasi Akademik dan Administrasi Keuangan pada tahun 2019-2021. Data yang didapatkan dari Biro Administrasi Keuangan memiliki 3 atribut yaitu nim, nama dan keterangan, dengan jumlah 8.833 data mahasiswa terlambat melakukan pembayaran SPP dan 30.811 data mahasiswa tepat waktu dalam pembayaran SPP. Sedangkan data yang didapatkan dari Bagian Administrasi Akademik berjumlah 10.959 data dengan 10 atribut yaitu nim, nama, fakultas, prodi, angkatan, gender, penghasilan ayah, penghasilan ibu, pendidikan ayah dan pendidikan ibu.

4.2. Data Preparation

Pada penelitian ini, tahap persiapan data yang dilakukan yaitu pembersihan data, integrasi data dan reduksi data.

1. Integrasi Data

Pada tahap ini menggabungkan data yang didapatkan dari bagian Biro Administrasi Akademik dan Administrasi Keuangan. Setelah digabungkan data yang diperoleh menghasilkan sebuah data baru yang terdiri dari nim, nama, fakultas, prodi, angkatan, gender, penghasilan ayah, penghasilan ibu, pendidikan ayah, pendidikan ibu dan keterangan.

2. Pembersihan Data

Sebelum data digunakan untuk analisis klasifikasi keterlambatan pembayaran SPP, maka perlu dilakukan pembersihan data terlebih dahulu. Data awal sebelum dibersihkan yaitu 39.644 data dan terdapat 10.099 data yang harus dibersihkan karena berisi data yang kosong dan inkonsisten pada atribut penghasilan ayah, penghasilan ibu, pendidikan ayah dan pendidikan ibu. Total jumlah data setelah dilakukan proses *cleaning* menjadi 29.545 data

3. Seleksi Data

Pada proses seleksi data dilakukan pembobotan pada dataset menggunakan *information gain ratio* dengan aplikasi *rapidminer*. Proses pembobotan dilakukan dengan menggunakan operator *Weight By Information Gain Ratio*.



Gambar 3. Skema Seleksi Fitur.

Hasil dari pembobotan didapatkan bahwa atribut Pendidikan Ibu dan Pendidikan Ayah merupakan atribut yang paling rendah dengan nilai 0.000 oleh karena itu atribut tersebut akan dihapus. Sehingga atribut yang digunakan pada penelitian ini adalah penghasilan ayah, penghasilan ibu, gender, fakultas, prodi dan angkatan. Berikut adalah hasil dari pembobotan dengan menggunakan *information gain ratio*.

Tabel 4. Hasil Pembobotan *Information Gain Ratio*.

Atribut	Gain Ratio
Pendidikan_Ibu	0.000
Pendidikan_Ayah	0.000
Penghasilan_Ayah	0.001
Penghasilan_Ibu	0.001
Gender	0.006
Fakultas	0.011
Prodi	0.012
Angkatan	0.133

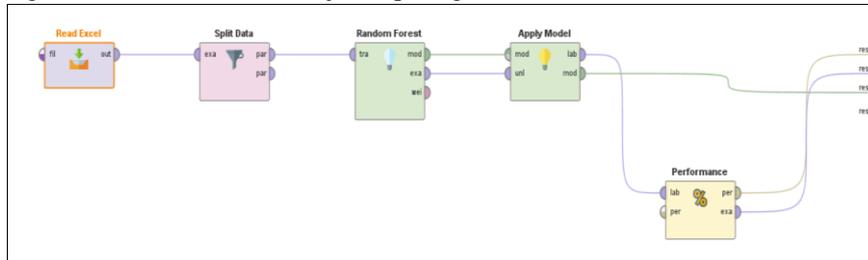
4. Reduksi Data

Data yang digunakan dalam penelitian ini dimensinya tidak seimbang, sehingga perlu dilakukan reduksi data karena kelas tepat berjumlah 23.341 data dan kelas terlambat berjumlah 6.204 data. Oleh karena itu, perlu dilakukan proses penyeimbang data dengan cara mengambil data mahasiswa

secara acak yang melakukan pembayaran SPP secara tepat waktu sebanyak 6.204 data dan terlambat sebanyak 6.204 data.

4.3. Pemodelan Random Forest

Algoritma yang diuji pada penelitian ini adalah algoritma random forest. Pengujian algoritma random forest menggunakan operator performance yang ada pada *rapidminer*. Adapun gambaran dari permodelan algoritma random forest ditunjukkan pada gambar 4 dibawah ini.



Gambar 4. Implementasi Algoritma *Random Forest* pada *Rapidminer*.

Pada gambar 4, operator *Read Excel* digunakan untuk membaca dataset penelitian dalam format *Excel*. Kemudian Operator *split data* digunakan untuk membagi dataset menjadi partisi *data training* dan *data testing* sebesar 90% dan 10%.

Operator *Random Forest* menghasilkan model klasifikasi dari algoritma *random forest* yang di implementasikan pada *data training* dengan presentase 90% dan *data testing* dengan presentase 10%. Selain itu operator *Apply Model* digunakan untuk menerapkan model yang telah dilatih sebelumnya. Operator *performance* digunakan untuk mengevaluasi kinerja model yang nantinya akan memberikan nilai kinerja berupa *accuracy*.

4.4. Evaluasi Performa Pemodelan Random Forest.

Berdasarkan pengujian menggunakan aplikasi *rapidminer* dengan algoritma *random forest*. Adapun hasil dari *performance* algoritma random forest dapat dilihat pada gambar 5 dibawah ini.

accuracy: 78.47%			
	true Terlambat	true Tepat	class precision
pred. Terlambat	527	174	75.18%
pred. Tepat	93	446	82.75%
class recall	85.00%	71.94%	

Gambar 5. Akurasi Pemodelan *Random Forest*.

Nilai *performance* yang dihasilkan oleh aplikasi *rapidminer* menggunakan model *confusion matrix* menghasilkan nilai akurasi sebesar 78,47%. Evaluasi dilakukan untuk mengetahui kinerja algoritma *random forest* dalam melakukan klasifikasi. Pada pengujian ini, data yang berjumlah 12.408 data dibagi menjadi 90% *data training* dan 10% *data testing*. Sehingga dapat disimpulkan hasil akurasi pada penelitian ini tergolong masuk pada kategori *fair classification*.

Tabel 5. Evaluasi *Confusion Matrix*.

<i>Confusion Matrix</i>	Prediksi	
	Terlambat	Tepat
Akurasi	527	174
	93	446

Pada tabel 5 diketahui nilai dari prediksi terlambat 527 yang merupakan nilai *true positive* dan nilai *true negative* mendapatkan nilai 174. Sedangkan nilai *true positive* pada prediksi tepat mendapatkan nilai 93 dan *true negative* 446. Dari tabel *confusion matrix*, didapatkan hasil perhitungan kinerja algoritma *random forest* sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN} * 100\%$$

$$\begin{aligned}
 &= \frac{527 + 446}{527 + 446 + 93 + 174} * 100\% \\
 &= \frac{973}{1.240} * 100 \\
 &= 78,47\%.
 \end{aligned}$$

5. KESIMPULAN

Dari hasil penelitian yang dilakukan menggunakan aplikasi rapidminer dengan menerapkan seleksi fitur *information gain ratio* pada algoritma *random forest* dalam klasifikasi keterlambatan pembayaran SPP, didapatkan kesimpulan bahwa Atribut yang memiliki pengaruh dalam kelas / label keterlambatan pembayaran SPP antara lain : fakultas, prodi, angkatan, gender penghasilan ayah, penghasilan ibu dan keterangan. Hasil implementasi algoritma *random forest* untuk prediksi keterlambatan pembayaran spp di Universitas Muhammadiyah Kalimantan Timur pada permodelan menghasilkan *true* prediksi terlambat sebesar 527 dan *true* prediksi tepat sebesar 93.

Hasil evaluasi algoritma *random forest* dengan seleksi fitur *information gain ratio* untuk model data klasifikasi keterlambatan pembayaran SPP di Universitas Muhammadiyah Kalimantan Timur dengan pembagian *data training* sebesar 90% dan *data testing* 10% menghasilkan akurasi sebesar 78,47%.

DAFTAR PUSTAKA

- [1] Santosa, B., & Umam, A., 2018, Data Mining Dan Big Data Analytics. Penebar Media Pustaka.
- [2] Apandi, T. H., Maulana, R. B., Piarna, R., & Vernanda, D. (2019). Menganalisis Kemungkinan Keterlambatan Pembayaran Spp Dengan Algoritma C4.5 (Studi Kasus Politeknik Tedc Bandung). Jurnal Techno Nusa Mandiri, 16(2), 93–98.
- [3] Abdullah, R. W., 2019, Sistem Prediksi Keterlambatan Pembayaran SPP Dengan Metode K-Nearest Neighbor (Studi Kasus: SMK AL-Islam Surakarta). Universitas Amikom Yogyakarta.
- [4] Istiana, S. R., 2018, Aplikasi Prediksi Pembayaran Bulanan Santri Dengan Menggunakan Algoritma C4.5 (Studi Kasus Pondok Pesantren Assalafi Al Fithrah Meteseh Semarang). Universitas Diponegoro.
- [5] Muqorobin, M., Kusriani, K., & Luthfi, E. T., 2019, Optimasi Metode Naive Bayes Dengan Feature Selection Information Gain Untuk Prediksi Keterlambatan Pembayaran Spp Sekolah. Jurnal Ilmiah SINUS, 17(1), 1.
- [6] Rianto, M., & Yunis, R., 2021, Analisis Runtun Waktu Untuk Memprediksi Jumlah Mahasiswa Baru Dengan Model Random Forest. Paradigma - Jurnal Komputer dan Informatika, 23(1), 70–74.
- [7] Suwardika, G., & Suniantara, I. ketup P. S., 2019, Analisis Random Forest Pada Klasifikasi CART Ketidaktepatan Waktu Kelulusan Mahasiswa Universitas Terbuka. Jurnal Ilmu Matematika dan Terapan, 13(3), 179–186.
- [8] Triwidianti, J., Alfian, F. Y. A., & Prasojo, M., 2021, Perbandingan Metode Data Mining Untuk Prediksi Prestasi Siswa Tingkat Pendidikan Menengah Kejuruan Pada Sekolah Menengah Kejuruan Negeri (Smkn 1) Gadingrejo Pringsewu Lampung. Seminar Nasional Hasil Penelitian dan Pengabdian Masyarakat, 126–133.
- [9] Linawati, S., Nurdiani, S., Handayani, K., & Latifah., 2020, Prediksi Prestasi Akademik Mahasiswa Menggunakan Random Forest Dan C4.5. JURNAL KHATULISTIWA INFORMATIKA, 8(1), 47–52.
- [10] Rosady, F., 2021, Penerapan Seleksi Fitur Information Gain Raiot dan Algoritma Random Forest Untuk Prediksi Kelulusan Mahasiswa Teknik Informatika UIN Suska Riau. Universitas Islam Negeri Sultan Syarif Kasim Riau.
- [11] Ghasemi, F., Neysiani, B. S., & Nematbakhsh, N., 2020, Feature selection in pre-diagnosis heart coronary artery disease detection: a heuristic approach for feature selection based on information gain ratio and Gini index. In 2020 6th International Conference on Web Research (ICWR) (pp. 27-32). IEEE.
- [12] Lestari, A., 2020, Increasing accuracy of C4. 5 algorithm using information gain ratio and adaboost for classification of chronic kidney disease. Journal of Soft Computing Exploration, 1(1), 32-38.
- [13] Fadlan., 2020. Notifikasi Keterlambatan Pembayaran Sumbangan Pembinaan Pendidikan (SPP) Berbasis Android Di SMK N 1 Baso. Institut Agama Islam (IAIN) Bukittinggi.
- [14] Hermawati, F.A., 2013, Data Mining. Yogyakarta: Penerbit Andi.
- [15] Mardi, Y., 2017, Data Mining: Klasifikasi Menggunakan Algoritma C4. 5. Edik Informatika, 2(2), 213-219.

- [16] Wibawa, A. P., 2018, Metode-metode Klasifikasi. In Prosiding SAKTI (Seminar Ilmu Komputer dan Teknologi Informasi) (Vol. 3, No. 1, pp. 134-138).
- [17] Primartha, R., 2021. Algoritma Machine Learning. Bandung: Informatika Bandung.
- [18] Putra, M.I. 2019. Sistem Rekomendasi Kelayakan Kredit Menggunakan Metode Random Forest pada BRI Kantor Cabang Pelaihari. Universitas Islam Negeri Sunan Ampel.
- [19] Liparas, D., Ha, Cohen-Kerner, Y., Moumtzidou, A., Vrochidis, S., & Kompatsiaris, I., 2014, News articles classification using random forests and weighted multimodal features. In Information Retrieval Facility Conference (pp. 63-75). Springer, Cham
- [20] Pulungan, A. F., 2019, Analisis Kinerja Bray Curtis Distance, Canberra Distance dan Euclidean Distance pada Algoritma K-Nearest Neighbour. Tersedia di: <https://repositori.usu.ac.id/handle/123456789/15051> [Diakses 28 Januari 2022].
- [21] D. Aprillia, D. A. Baskoro, L. Ambarwati dan I. W. S. Wicaksana, Belajar Data Mining dengan RapidMiner, Jakarta, 2013.