

**PREDIKSI *CHURN* SISWA-SISWI SMP PAHLAWAN
NASIONAL DENGAN MENGGUNAKAN ALGORITMA
*RANDOM FOREST***

SKRIPSI

DWI FITRI ARDAYANI SEMBIRING

198160075



**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA**

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

Document Accepted 23/1/25

1. Dilarang Mengutip sebagian atau seluruh dokumen ini tanpa mencantumkan sumber
2. Pengutipan hanya untuk keperluan pendidikan, penelitian dan penulisan karya ilmiah
3. Dilarang memperbanyak sebagian atau seluruh karya ini dalam bentuk apapun tanpa izin Universitas Medan Area

**PREDIKSI *CHURN* SISWA-SISWI SMP PAHLAWAN
NASIONAL DENGAN MENGGUNAKAN ALGORITMA
*RANDOM FOREST***

SKRIPSI

*Diajukan Untuk Memenuhi Sebagian Persyaratan Dalam Memperoleh
Gelar Sarjana Universitas Medan Area*



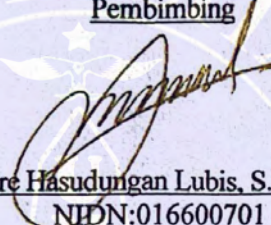
**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS MEDAN AREA
2024**

LEMBAR PENGESAHAN

Judul : Prediksi *Churn* Siswa-Siswi Smp Pahlawan Nasional
Dengan Menggunakan Algoritma *Random Forest*
Nama : Dwi Fitri Ardayani Sembiring
NPM : 198160075
Fakultas : Teknik
Prodi : Teknik Informatika

Disetujui Oleh Komisi Pembimbing

Pembimbing


Andre Hasudungan Lubis, S.TI, M.Sc
NIDN:016600701

Diketahui Oleh:

Dekan Fakultas Teknik

Ketua Prodi Teknik Informatika



Dwi Fitri Ardayani Sembiring, S.T, M.T
NIDN:016600701



Rizki Muliono, S.Kom, M.Kom
NIDN:0109038902

Tanggal lulus : 30 Agustus 2024

HALAMAN PERNYATAAN

Saya menyatakan bahwa skripsi yang saya susun, sebagai syarat memperoleh gelar sarjana merupakan hasil karya tulis saya sendiri. Adapun bagian-bagian tertentu dalam penulisan skripsi ini yang saya kutip dari hasil karya orang lain telah dituliskan sumbernya secara jelas sesuai dengan norma, kaidah, dan etika penulisan ilmiah.

Saya bersedia menerima sanksi pencabutan gelar akademik yang saya peroleh dan sanksisanksi lainnya dengan peraturan yang berlaku, apabila di kemudian hari ditemukan adanya plagiat dalam skripsi ini.

Medan,.....2024



Dwi Fitri Ardayani Sembiring
NPM: 198160075

**HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI
TUGAS AKHIR/SKRIPSI/TESIS UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademik Universitas Medan Area, saya yang bertanda tangan di bawah ini:

Nama : Dwi Fitri Ardayani Sembiring
NPM : 198160075
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area **Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalty-Free Right*)** atas karya ilmiah saya yang berjudul:

Prediksi *Churn* Siswa Siswi SMP Pahlawan Nasional Dengan Menggunakan Algoritma *Random Forest*

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Medan Area berhak menyimpan, mengalih media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir/skripsi/tesis saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Medan, 10 September 2024
Yang menyatakan

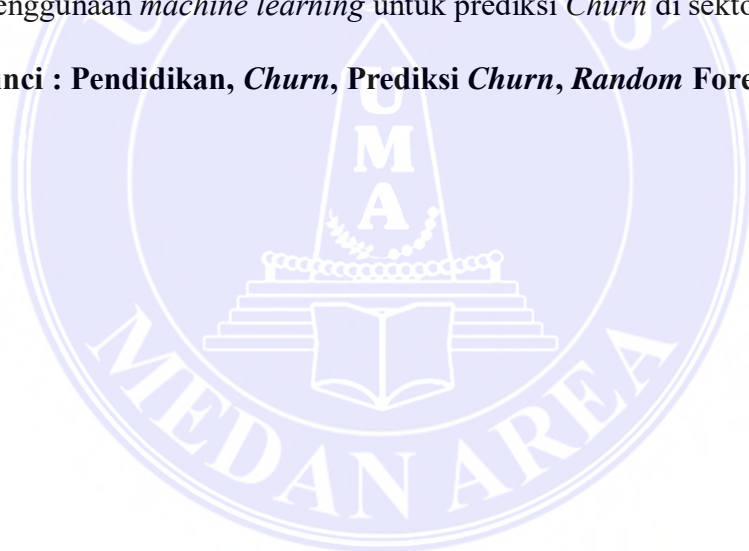


Dwi Fitri Ardayani Sembiring
NPM:198160075

ABSTRAK

Pendidikan merupakan pondasi penting bagi pembangunan suatu negara dan kehidupan individu. Namun, salah satu tantangan utama dalam dunia pendidikan saat ini adalah tingginya angka siswa yang mengalami *Churn*, yaitu siswa yang berhenti sekolah atau berpindah ke lembaga pendidikan lain. Masalah ini memberikan dampak signifikan bagi individu dan institusi pendidikan. Penelitian ini berfokus pada permasalahan *Churn* di SMP Pahlawan Nasional dan bertujuan untuk membangun model prediksi guna mengidentifikasi siswa yang berpotensi *Churn*. Untuk mencapai tujuan tersebut, algoritma *Random Forest* diterapkan menggunakan data akademik, absensi, dan informasi sosioekonomi. Model dibangun melalui proses pengumpulan data, pra-pemrosesan, pembentukan model, dan evaluasi menggunakan metrik kinerja seperti akurasi, *precision*, *recall*, dan *F1-score*. Hasil evaluasi menunjukkan bahwa model memiliki akurasi sebesar 85,7%, dengan *precision* untuk kelas *Churn* sebesar 89,6% dan *recall* sebesar 81,2%. *F1-score* model mencapai 85,2%, menunjukkan keseimbangan yang baik antara *precision* dan *recall*. Implementasi model ini diharapkan dapat membantu sekolah dalam mengambil langkah preventif untuk mengurangi *Churn* dan meningkatkan retensi siswa. Penelitian ini juga memberikan kontribusi ilmiah dalam penggunaan *machine learning* untuk prediksi *Churn* di sektor pendidikan.

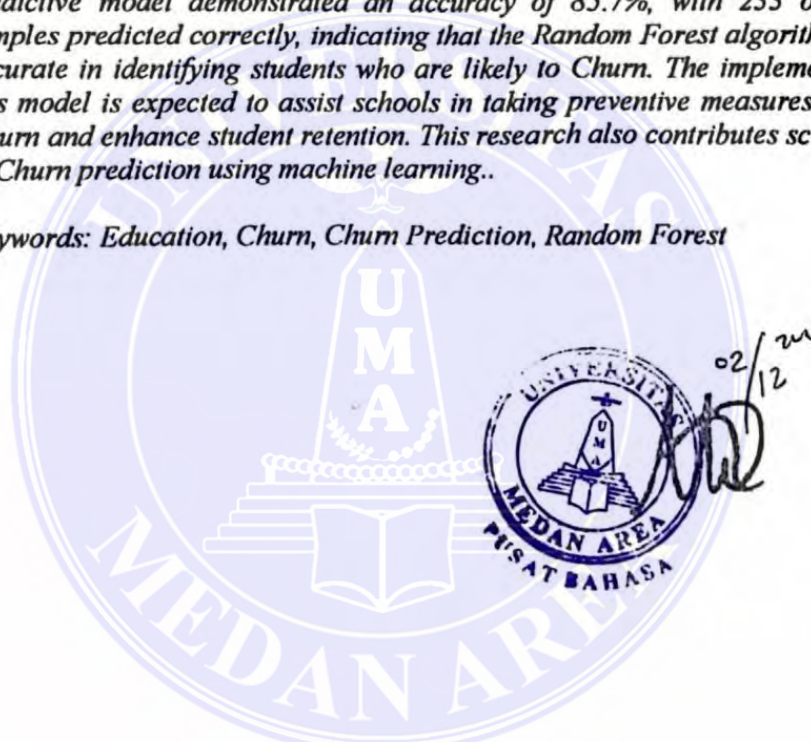
Kata kunci : Pendidikan, *Churn*, Prediksi *Churn*, *Random Forest*



ABSTRACT

Education is the foundation for the development of a country and is essential for every individual to achieve a more meaningful life. However, one of the major challenges in the education sector today is the high rate of student Churn, which refers to students who drop out or transfer to other educational institutions. This issue has significant impacts on both individuals and educational institutions. The Churn problem at SMP Pahlawan Nasional indicates the need for a predictive model to identify students at risk of Churn. This is crucial for educational institutions to take appropriate preventive measures. This research aimed to predict student Churn by applying the Random Forest algorithm, utilizing academic data, attendance, and socioeconomic information. The model was constructed through data collection, preprocessing, model building, and evaluation using Mean Squared Error (MSE) and Root Mean Squared Error (RMSE), with results showing an MSE of 0.22 and an RMSE of 0.47. The predictive model demonstrated an accuracy of 85.7%, with 233 out of 272 samples predicted correctly, indicating that the Random Forest algorithm is quite accurate in identifying students who are likely to Churn. The implementation of this model is expected to assist schools in taking preventive measures to reduce Churn and enhance student retention. This research also contributes scientifically to Churn prediction using machine learning..

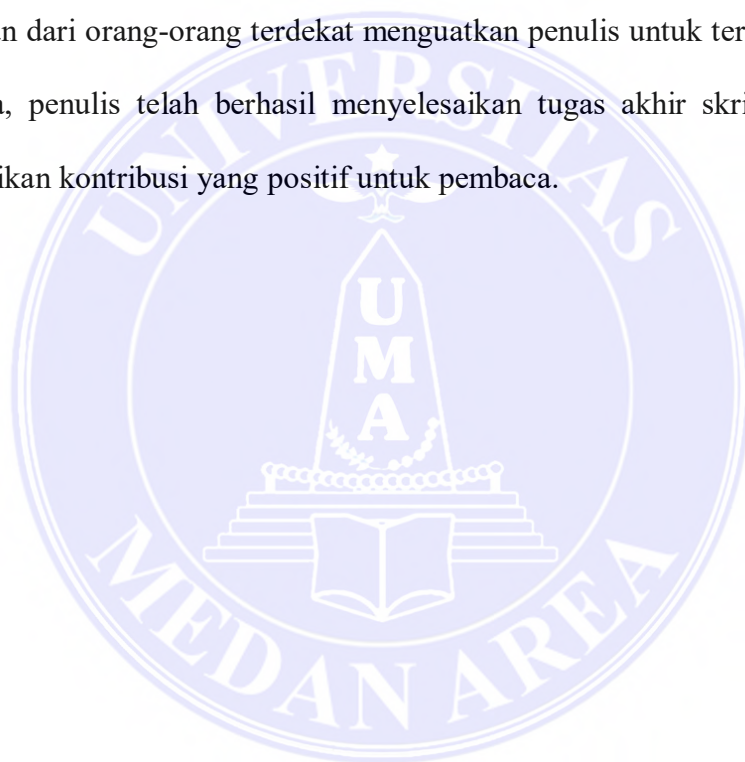
Keywords: Education, Churn, Churn Prediction, Random Forest



RIWAYAT HIDUP

Penulis dilahirkan di Kota Pematang Siantar pada tanggal 20 Januari 2001, anak dari Bapak Alm. Swadaya Sembiring dan Ibu Megawani. Penulis merupakan anak ke dua dari tiga bersaudara. Pada Tahun 2018 Penulis lulus dari SMA Yayasan Perguruan Sultan Agung Pematang Siantar dan pada tahun 2019 terdaftar sebagai mahasiswa Fakultas Teknik Universitas Medan Area.

Berkat Izin dari Allah SWT serta usaha, perjuangan, kesabaran dan dukungan dari orang-orang terdekat menguatkan penulis untuk terus berjuang dan berusaha, penulis telah berhasil menyelesaikan tugas akhir skripsi ini mampu memberikan kontribusi yang positif untuk pembaca.



KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah Swt, berkat limpahan rahmat dan karunia yang telah diberikan sehingga penulis berhasil menyelesaikan skripsi ini dengan judul “Prediksi *Churn* Siswa Siswi SMP Pahlawan Nasional Dengan Menggunakan Algoritma *Random Forest*”

Skripsi ini merupakan salah satu syarat untuk menyelesaikan pendidikan untuk mencapai gelar sarjana di Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area. Pada kesempatan ini peneliti mengucapkan terimakasih kepada:

1. Bapak Prof. Dr. Dadan Ramdan, M.Eng, M.Sc. selaku Rektor Universitas Medan Area.
2. Bapak Dr.Eng Supriatno S.T, M.T sebagai Dekan Fakultas Teknik Universitas Medan Area.
3. Bapak Rizki Muliono, S.Kom., M.Kom selaku Kepala Program Studi Teknik Informatika Universitas Medan Area
4. Bapak Andre Hasudungan Lubis, S.TI, M.Sc selaku Dosen pembimbing yang telah membantu penulis dari segi materi dan dukungan sehingga penulis dapat menyelesaikan skripsi ini.
5. Kepada Ayahanda Alm. Swadaya Sembiring terimakasih selama hidupmu senantiasa menjadi sosok ayah yang tegas, baik dan menyenangkan. Terimakasih telah senantiasa mencintai dan menyayangi penulis sepanjang hidupmu.

6. Kepada Ibunda Megawani, terimakasih sudah melahirkan, membesarkan, memberi kasih sayang serta mendidik penulis sampai pada hari ini, terimakasih sudah panjang umur dan sehat sehingga dapat menemani penulis sampai hari ini. Terimakasih atas dukungan moral maupun moril yang telah di berikan, serta doa yang senantiasa dilangitkan. Semoga Allah SWT selalu menjaga ibu dalam kebaikan dan kemudahan Aamiin.
7. Kepada Almh. Eyang Uti, kesayangan penulis, terimakasih selama hidupmu sudah menemani dan mendoakan penulis. Terimakasih telah senantiasa memberikan dukungan moral maupun moril serta senantiasa mencintai dan menyayangi penulis sepanjang hidupmu. Terimakasih karna sudah menjadi alasan penulis untuk menyelesaikan skripsi ini.
8. Kepada ke dua saudara-saudara penulis Rizky Ardayani Sembiring, A.Md dan Tegar Ananta Sembiring, S.Kom. Terimakasih senantiasa mendoakan serta memberikan dukungan dan senantiasa percaya bahwa penulis bisa menyelesaikan skripsi ini dengan baik.
9. Seluruh Dosen dan Staf Program Studi Teknik Informatika Universitas Medan Area.
10. Seluruh teman-teman yang sudah memberikan dukungannya selama penulisan skripsi ini, khususnya Lala Oktaviantri yang sudah menemani dan mendukung penulis dalam segala hal dan teman-teman Teknik Informatika angkatan 2019.

Penulis menyadari bahwa penelitian ini masih memiliki kekurangan, oleh karena itu kritik dan saran yang bersifat membangun sangat penulis harapkan demi kesempurnaan penelitian ini. Penulis berharap tugas penelitian ini dapat bermanfaat baik kalangan pendidikan maupun masyarakat. Akhir kata penulis ucapkan terima kasih.

Medan, 10 September 2024



Dwi Fitri Ardayani Sembiring
NPM:198160075



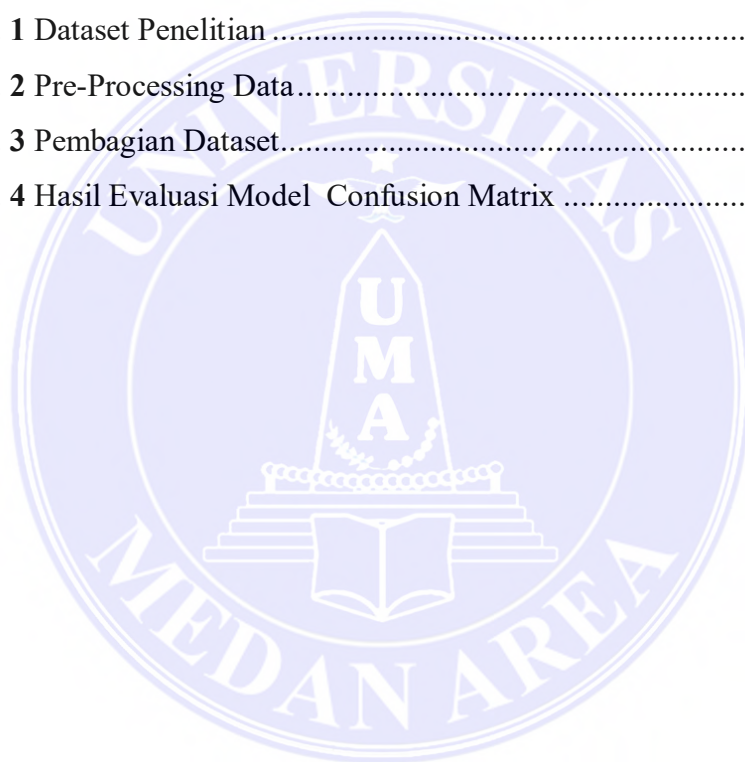
DAFTAR ISI

LEMBAR PENGESAHAN	Error! Bookmark not defined.
HALAMAN PERNYATAAN	iii
HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI	Error!
Bookmark not defined.	
ABSTRAK	vi
ABSTRACT	Error! Bookmark not defined.
RIWAYAT HIDUP	viii
KATA PENGANTAR	ix
DAFTAR TABEL	xiv
DAFTAR GAMBAR	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian	5
1.4 Manfaat Penelitian	5
1.5 Batasan Masalah	6
BAB II LANDASAN TEORI	7
2.1 <i>Churn</i>	7
2.2 Variabel- Variabel Yang Mempengaruhi <i>Churn</i>	8
2.2.1 Kemampuan Akademik	8
2.2.2 Absensi	9
2.2.3 Status Ekonomi Keluarga	10
2.3 <i>Random Forest</i>	11
2.4 Klasifikasi.....	13
2.5 Prediksi dalam <i>Machine Learning</i>	14
2.6 Evaluasi Model	15
2.6.1 <i>Confusion Matrix</i>	15
2.6.2 <i>Area Under the Curve (AUC)</i>	18

2.7 Penelitian Terdahulu	19
BAB III METODOLOGI PENELITIAN	22
3.1 Tahapan Penelitian	22
3.2 Pengumpulan Data	23
3.3 Pra-Pemrosesan Data.....	25
3.4 Pembentukan Model.....	26
3.4.1 Analisis Menggunakan Random Forest	27
3.5 Evaluasi Model	34
3.6 Alat dan Bahan Penelitian	34
BAB IV HASIL DAN PEMBAHASAN.....	36
4.1 Hasil	36
4.1.1 Evaluasi Model.....	38
4.1.2 Visualisasi Hasil.....	43
4.2 Pembahasan	46
BAB V KESIMPULAN DAN SARAN.....	48
5.1 Kesimpulan.....	48
5.2 Saran.....	48
DAFTAR PUSTAKA.....	50
LAMPIRAN	55

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu	19
Tabel 3. 2 Contoh Dataset	25
Tabel 3. 3 Tabel Mengubah Atribut Menjadi Jenis Kategorikal	27
Tabel 3. 4 Tabel Node 1	28
Tabel 3. 5 Tabel Node 1.1 Nilai Rata – Rata.....	30
Tabel 3. 6 Tabel Node 1.2 Nilai Rata – Rata.....	32
Tabel 3. 7 Kebutuhan Perangkat Keras	35
Tabel 3. 8 Kebutuhan Perangkat Lunak	35
Tabel 4. 1 Dataset Penelitian	36
Tabel 4. 2 Pre-Processing Data.....	37
Tabel 4. 3 Pembagian Dataset.....	38
Tabel 4. 4 Hasil Evaluasi Model Confusion Matrix	40



DAFTAR GAMBAR

Gambar 2. 1 Proses Kerja Algoritma Random Forest Sumber; (Allaam,2023) ..	12
Gambar 2. 2 Ilustrasi Gambar Confussion Matrix	16
Gambar 3. 1 Flowchart Tahapan Penelitian.....	23
Gambar 3. 2 Visualisasi Pohon Keputusan.....	33
Gambar 4. 1 Evaluasi Model Menggunakan Confussion Matrix.....	39
Gambar 4. 2 Hasil Evaluasi Model Menggunakan AUC	42
Gambar 4. 3 Analisis Scatter Plot	43
Gambar 4. 4 Feature Importance.....	45



BAB I

PENDAHULUAN

1.1 Latar Belakang

Pendidikan merupakan fondasi bagi pembangunan suatu negara. Pendidikan penting bagi setiap individu untuk mencapai kehidupan yang lebih bermakna, dimana setiap individu berhak menerima pendidikan dari orang tua dan lembaga pendidikan formal seperti sekolah dan perguruan tinggi (Nurhuda, 2022). Lembaga pendidikan sebagai salah satu komponen penting dalam struktur sosial budaya saat ini menimbulkan berbagai tantangan yang kompleks. Saat ini, lembaga pendidikan dihadapkan pada arus perubahan yang cepat akibat globalisasi, yang menghasilkan persaingan dalam pengelolaannya, baik lembaga pendidikan negeri maupun swasta. Sehingga tantangan dalam dunia pendidikan tidak bisa dipandang sebelah mata (Hastari, 2023).

Salah satu tantangan utama yang dihadapi oleh lembaga pendidikan adalah besarnya jumlah siswa yang melakukan tindakan *Churn* setiap tahun nya. *Churn* adalah individu yang meninggalkan atau berhenti berlayanan pada suatu perusahaan karena berbagai alasan. Sehingga pada konteks pendidikan, *Churn* di definisikan sebagai individu yang pindah atau berhenti sekolah karena berbagai alasan. Sehingga penting bagi lembaga pendidik untuk mampu memprediksi siswa-siswi guna mengurangi kasus *Churn* yang semakin tinggi (Mathai, 2020). Permasalahan siswa-siswi putus sekolah di Indonesia adalah masalah yang *multidimensional*. Putus sekolah merupakan persoalan pendidikan yang perlu dicari akar penyebabnya. Terdapat berbagai unsur dasar alasan seorang anak bisa putus sekolah, antara lain faktor ekonomi karena orang tua tidak mampu

membayai anaknya untuk bersekolah, juga karena faktor membantu usaha orang tua, atau faktor siswa sendiri yang kurang mengerti akan pentingnya pendidikan. Selain itu, faktor lingkungan dan teman-teman siswa juga turut berperan dalam keputusan siswa untuk putus sekolah (Fatma & Rochmawati, 2024). Tindakan *Churn* yang dilakukan siswa memiliki dampak yang signifikan, baik bagi individu maupun institusi pendidikan, yang meliputi kerugian sumber daya, waktu, dan finansial. Untuk mengatasi masalah ini, diperlukan upaya untuk memahami penyebab dan pola perilaku serta mengembangkan strategi untuk mengurangi tingkat *Churn* (Fauszt dkk, 2021).

Berdasarkan permasalahan tersebut menunjukkan pentingnya memahami dan mengatasi masalah siswa yang berhenti sekolah dengan lebih efisien. Salah satunya Di SMP Pahlawan Nasional, pemahaman yang mendalam mengenai faktor-faktor yang mempengaruhi keputusan siswa untuk tetap melanjutkan atau pindah dari sekolah menjadi sangat penting. Pada saat ini, kemajuan teknologi informasi dalam komputasi telah berkembang dengan cepat dan berhasil.

Salah satu bidang yang sedang berkembang dalam teknologi informasi adalah *Machine Learning*. *Machine Learning* dibagi menjadi beberapa bagian, salah satunya adalah regresi. Regresi adalah jenis analisis regresi yang digunakan untuk menjelaskan hubungan antara satu variabel dependen (*variabel respon*) dengan dua atau lebih variabel independen (*variabel prediktor*) (Ghisyani dkk., 2023). Metode regresi, seperti regresi linier atau regresi logistik, adalah alat statistik yang digunakan untuk memahami hubungan antara variabel dan membuat perkiraan berdasarkan data historis. Dalam regresi linier, kita mencari hubungan linier

antara variabel independen dan dependen, yang digambarkan dengan garis lurus, sementara regresi logistik digunakan ketika variabel dependen adalah biner, membantu kita memprediksi probabilitas suatu peristiwa berdasarkan variabel independen. Baik regresi linier maupun regresi logistik berguna untuk membuat estimasi dan memahami pengaruh faktor-faktor tertentu terhadap hasil yang ingin diprediksi (Rahayu dkk., 2024). Salah satu algoritma yang dapat diterapkan pada model regresi yaitu algoritma *Random Forest*.

Random Forest (RF) merupakan evolusi dari metode *bagging* yang bertujuan untuk mengurangi variasi model statistik dan menangani variabilitas data dengan cara secara acak mengekstraksi sampel *bootstrap* dari set pelatihan dan melakukan prediksi secara *agregat* untuk data baru. *Random Forest* memiliki kemampuan yang sangat baik untuk meningkatkan prediksi dari banyak metode *supervised* lainnya, terutama pohon keputusan (Kurniawan dkk., 2023).

Teori tersebut dibuktikan melalui beberapa penelitian yang sudah dilakukan, di antaranya penelitian yang sudah dilakukan oleh (Makurumidze dkk., 2022) dengan judul penelitian “*Implementing Random Forest to Predict Churn*” Penelitian ini bertujuan untuk memprediksi klien bank yang berpotensi untuk melakukan *Churn*. penelitian ini menggunakan empat algoritma *machine learning*, yaitu *Gradient Boost*, *Random Forest*, *Adaboost*, dan *Decision Tree*. Hasil penelitian ini menunjukkan bahwa *Random Forest* dan *Gradient Boosting* memiliki kinerja yang lebih baik dibandingkan dengan metode lain seperti *Adaboost* dan *Decision Tree*. Evaluasi model menunjukkan bahwa *Random Forest* memiliki akurasi yang tinggi, dengan nilai presisi 93%, recall 84%, F1-score 88%

dan akurasi 96%. Hasil tersebut menggambarkan bahwa random forest memiliki kemampuan yang baik dalam mempelajari pola data dan memprediksi *churn* dengan akurat.

Kemudian penelitian yang dilakukan oleh (Malikireddy, 2021) dengan judul “*Customer Churns Prediction Model Based on Machine Learning Techniques: A Systematic Review*” penelitian ini melakukan *review* terhadap penelitian untuk memprediksi *Churn* pelanggan dengan menggunakan beberapa algoritma *machine learning*. Hasil penelitian menemukan bahwa Random Forest, LSTM, dan CNN memiliki kinerja yang berbeda-beda dalam memprediksi *Churn* pelanggan. Namun algoritma Random Forest memiliki keunggulan dalam menganalisis fitur-fitur data yang lebih baik dibandingkan algoritma yang lain.

Kemudian penelitian yang dilakukan oleh (Behr dkk., 2020) dengan judul penelitian “*Early Prediction of University Dropouts – A Random Forest Approach*” Tujuan penelitian ini yaitu untuk memprediksi apakah seorang mahasiswa akan lulus atau *drop-out* dari universitas. Hasil penelitian menunjukkan bahwa model yang digunakan mampu memberikan prediksi yang baik terhadap mahasiswa yang berhenti kuliah, dengan hasil AUC yang sangat tinggi, berkisar antara 0,83 hingga 0,88. Nilai recall 81% dan 89%, menunjukkan bahwa lebih dari 80% mahasiswa yang *drop-out* dapat diidentifikasi dengan benar oleh model-model yang dibuat.

Setelah melakukan studi literatur terhadap algoritma *Random Forest* dapat disimpulkan bahwa algoritma RF dapat melakukan proses prediksi dengan baik

serta dapat menghasilkan akurasi yang baik. Berdasarkan masalah yang sudah dipaparkan, maka pada tugas akhir ini akan dilakukan penelitian dengan melakukan model prediksi dan mengimplementasikan algoritma *Random Forest* untuk memprediksi tindakan *Churn* siswa-siswi di SMP Pahlawan Nasional sehingga tugas akhir ini diberi judul “Prediksi *Churn* Siswa-Siswi SMP Pahlawan Nasional Dengan Menggunakan Algoritma *Random Forest*”.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas maka dapat dirumuskan rumusan masalah pada penelitian yaitu:

Bagaimana cara implementasi algoritma *Random Forest* untuk memprediksi tindakan *Churn* Siswa-Siswi SMP Pahlawan Nasional?

1.3 Tujuan Penelitian

Adapun tujuan dari penelitian ini yaitu, menerapkan algoritma *Random Forest* untuk memprediksi tindakan *Churn* siswa-siswi di SMP Pahlawan Nasional.

1.4 Manfaat Penelitian

Adapun manfaat yang akan didapatkan dari penelitian ini yaitu:

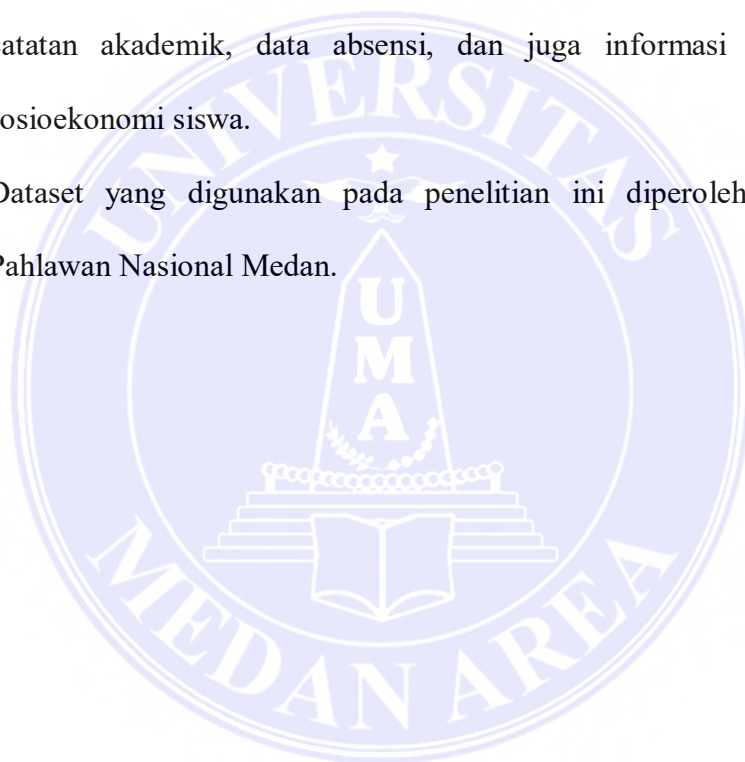
1. Hasil penelitian ini diharapkan dapat memberikan pengetahuan ilmiah terkhusus dalam proses Prediksi menggunakan algoritma *Random Forest*
2. Menjadi landasan referensi kepada peneliti selanjutnya apabila akan melakukan penelitian terkait dengan algoritma *Random Forest*.

3. Diharapkan hasil dari penelitian ini dapat memberikan kontribusi dalam peningkatan efektivitas dalam mengatasi masalah *Churn* siswa-siswi di SMP Pahlawan Nasional.

1.5 Batasan Masalah

Adapun batasan masalah pada penelitian ini:

1. Data yang digunakan pada penelitian ini merupakan informasi seperti catatan akademik, data absensi, dan juga informasi terkait kondisi sosioekonomi siswa.
2. Dataset yang digunakan pada penelitian ini diperoleh melalui SMP Pahlawan Nasional Medan.



BAB II

LANDASAN TEORI

2.1 *Churn*

Churn merujuk pada fenomena ketika pelanggan beralih dari satu penyedia layanan ke penyedia lain, yang dapat menyebabkan kehilangan pelanggan secara berkala dalam suatu organisasi, dan ini dapat mengakibatkan penurunan pendapatan yang signifikan bagi sebuah perusahaan atau organisasi. Hal ini merupakan masalah serius sehingga untuk mempertahankan pelanggan, sebuah organisasi atau lembaga harus meningkatkan fasilitas dan layanannya, serta mengidentifikasi pelanggan yang memiliki kecenderungan untuk meninggalkan layanan (Wicaksono dkk., 2021).

Pada umumnya tindakan *Churn* memiliki tiga jenis yaitu, *Active/Deliberate*, dimana seseorang secara aktif memutuskan untuk berhenti berlangganan dan beralih ke penyedia layanan atau jasa lainnya. Kedua, *Rotational/Incidental*: yaitu seseorang yang memutuskan untuk berhenti berlangganan tanpa beralih ke penyedia layanan atau jasa lainnya. Yang ketiga *Passive/Non Voluntary: Churn* jenis ini terjadi ketika pemutusan layanan dilakukan oleh perusahaan tanpa keinginan dari pelanggan (Atthariq, 2020). Sehingga pada penelitian ini tindakan *Churn* yang dimaksud yaitu, siswa yang memutuskan untuk berhenti sekolah atau beralih ke lembaga pendidikan atau sekolah lainnya.

Prediksi *Churn* merupakan metode untuk mengidentifikasi siswa-siswi yang akan berpindah atau berhenti sebelum mereka benar-benar melakukannya. Ini dilakukan dengan menganalisis data dan menemukan pola yang dapat

memberikan petunjuk tentang perilaku *Churn*. Dengan demikian, lembaga pendidikan dapat mengambil langkah-langkah pencegahan atau retensi yang sesuai untuk mempertahankan siswi-siswi tersebut (Wicaksono dkk., 2021)

2.2 Variabel- Variabel Yang Mempengaruhi *Churn*

Peningkatan dalam sektor pendidikan merupakan aspek yang sangat signifikan dalam meningkatkan kualitas Sumber Daya Manusia. Oleh karena itu, pemerintah telah menegaskan pentingnya program wajib belajar 12 tahun untuk menciptakan Sumber Daya Manusia yang berkualitas. Namun, perlu ditingkatkan partisipasi sekolah pada tingkat sekolah menengah agar dapat mengurangi angka tindakan *Churn* siswa (Hakim, 2020). Ada beberapa variabel-variabel yang mempengaruhi tindakan *Churn* siswa yaitu sebagai berikut:

2.2.1 Kemampuan Akademik

Kemampuan akademik secara erat berkaitan dengan kesuksesan akademik siswa. Salah satu penunjuk keberhasilan akademik adalah tingkat kemampuan siswa yang harus senantiasa ditingkatkan, seperti kemampuan berpikir kritis, penalaran tinggi, dan keterampilan dalam memecahkan masalah, karena hal ini berkaitan dengan kemampuan akademik siswa secara keseluruhan. Jika kemampuan akademik siswa terhambat, maka proses kognitifnya juga akan terhambat. Namun, pada kenyataannya, tidak semua siswa mampu mengatasi tekanan akademik yang mereka hadapi di lingkungan sekolah (Rahmadani & Daulay, 2023).

Faktor-faktor dalam lingkungan sekolah meliputi beban kurikulum yang berlebihan, fokus pada nilai, kecemasan menghadapi ujian, pengajaran guru yang tidak menarik, hukuman yang tidak memberikan pembelajaran, mata pelajaran tertentu yang dianggap menakutkan, kurangnya fasilitas yang mendukung bakat dan minat siswa, serta lingkungan sosial yang dapat memicu stres pada peserta didik. Begitu juga dengan atmosfer sekolah, metode pengajaran guru/dosen, materi yang dianggap sulit, dan beban tugas dapat menyebabkan tekanan dan stres yang berdampak pada proses belajar berdampak pada hasil belajar siswa sehingga menghasilkan rendahnya nilai akademik yang di peroleh sehingga hal tersebut menjadi pemicu terjadinya tindakan *Churn* pada peserta didik (Rahmawati dkk., 2022).

2.2.2 Absensi

Peserta didik belajar lebih baik ketika mereka menghadiri sekolah secara teratur dan cenderung kesulitan belajar apabila seringkali tidak hadir sekolah atau yang sering disebut absen. Jika seorang siswa seringkali absen, mereka mungkin akan mengalami kesulitan dalam memahami pelajaran, mendapat nilai yang lebih rendah, dan bahkan mungkin putus sekolah lebih awal (Keppens, 2023). Permasalahan membolos atau tindakan kebiasaan absen di sekolah tanpa izin, seringkali menjadi penyebab rendahnya pencapaian belajar. Kebiasaan ini terutama terjadi di kalangan siswa sekolah menengah dan memiliki dampak yang meluas pada masyarakat. Dampaknya seperti hasil akademis yang buruk, peningkatan risiko putus sekolah, penyalahgunaan narkoba, perilaku kenakalan, dan kehamilan usia remaja. Dampak pada jangka panjang, membolos bisa

berdampak masalah di masa dewasa, seperti kemiskinan, kekerasan, ketidakstabilan dalam hubungan, ketidakstabilan pekerjaan, kriminalitas, dan bahkan penahanan. Ketidakhadiran yang tinggi di sekolah seringkali berujung pada tindakan *Churn*, yang menjadi konsekuensi yang paling jelas dari masalah tersebut (Setiawati, 2020).

2.2.3 Status Ekonomi Keluarga

Peserta didik berasal dari latar belakang sosial-ekonomi yang berbeda. Peserta didik memiliki beragam latar belakang, belajar dan berinteraksi bersama dalam proses pembelajaran. Meskipun demikian, perbedaan ini seharusnya tidak menghambat proses belajar mengajar. Namun, terkadang status sosial-ekonomi ini dapat menjadi halangan bagi peserta didik (Guru dan Tenaga Kependidikan Pendidikan Dasar, 2020). Penyebab utama anak putus sekolah adalah kesulitan ekonomi atau ketidakmampuan orang tua untuk membiayai pendidikan anak-anak mereka. Oleh karena itu, anak-anak terpaksa berhenti sekolah karena keluarganya tidak mampu memenuhi biaya pendidikan mereka. masalah ekonomi menjadi penyebab utama.

Hal ini terlihat dari banyaknya kasus putus sekolah yang disebabkan oleh keterbatasan ekonomi keluarga. Kedua, rendahnya kesadaran orang tua akan pentingnya pendidikan juga berkontribusi terhadap fenomena ini. Hal ini tercermin dari masih adanya pandangan di masyarakat bahwa pendidikan anak perempuan tidak penting atau tidak perlu sampai jenjang yang tinggi. Selain itu, seringkali orang tua meminta anak mereka berhenti sekolah karena mereka

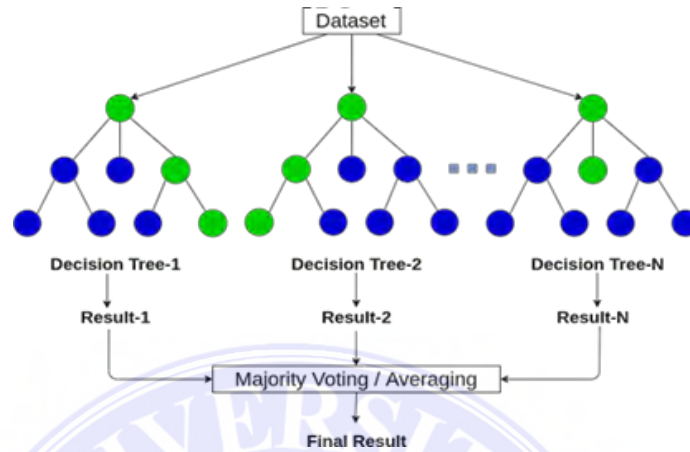
membutuhkan bantuan anak untuk pekerjaan keluarga, terutama di daerah perkotaan di mana anak-anak sering bekerja di pabrik untuk membantu ekonomi keluarga. Di daerah pedesaan, selain di sektor pertanian dan perkebunan, anak-anak putus sekolah juga sering bekerja di sektor industri kecil, sektor informal, dan perdagangan tradisional. Sehingga fenomena tersebut seringkali menjadi pemicu tindakan *Churn* dari siswa (Budi Lestari dkk., 2020)

2.3 *Random Forest*

Random Forest merupakan sebuah algoritma *Machine Learning* yang digunakan untuk tugas-tugas klasifikasi dan regresi. Dimana algoritma *Random Forest* menggabungkan hasil dari beberapa pohon keputusan (*Decision Tree*). Seperti namanya, "*forest*" mengacu pada sekumpulan pohon yang diperoleh melalui proses *bagging* atau *bootstrap aggregating*. Proses *bagging* ini bertujuan untuk membuat sejumlah besar model yang mungkin cenderung *noisy*, tetapi tidak memiliki bias yang signifikan, dan untuk mengurangi variansnya. Setiap pohon keputusan dibangun secara independen, menggunakan subset acak dari data pelatihan dan fitur yang dipilih secara acak. Setiap pohon dalam *Random Forest* memberikan suara (*vote*) terhadap prediksi, dan prediksi akhir didasarkan pada mayoritas suara dari semua pohon tersebut (Allaam, 2023)

Pada saat memprediksi, setiap pohon memberikan suara atau estimasi, dan prediksi akhir adalah hasil mayoritas suara dari semua pohon tersebut. Keunggulan utama dari *Random Forest* adalah kemampuannya untuk mengatasi *overfitting*, karena penggunaan banyak pohon yang masing-masing hanya melihat

sebagian kecil dari data. Ini membuatnya menjadi salah satu algoritma yang umum digunakan dalam berbagai aplikasi *Machine Learning* (Sandag, 2020).



Gambar 2. 1 Proses Kerja Algoritma *Random Forest* Sumber; (Allaam,2023)

Dalam menentukan nilai *Decision Tree* pada *Random Forest* akan dilakukan perhitungan nilai *entropy*, dimana rumus yang digunakan untuk menghitung nilai *entropy* adalah sebagai berikut:

$$Entropy(Y) = \sum p(c|Y) \log_2 p(c|Y) \quad (2.1)$$

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2.2)$$

Dengan keterangan dimana Y adalah himpunan kasus dan p(c|Y) merupakan proporsi nilai Y terhadap kelas c.

Pada gambar 2.1 menunjukkan alur proses dari algoritma *Random Forest* dengan detail penjelasan sebagai berikut:

1. Pemilihan Sampel Acak: Algoritma memilih sampel data acak dengan penggantian dari data pelatihan.
2. Pembentukan *Decision Tree*: *Decision Tree* dibangun pada setiap sampel data acak, memberikan prediksi untuk setiap observasi.
3. Pemilihan Nilai Prediksi: Nilai prediksi dipilih berdasarkan mayoritas suara (untuk klasifikasi) atau rata-rata (untuk regresi) dari semua *Decision Tree*.
4. Pemilihan Hasil Terbaik: Hasil prediksi terbaik dipilih berdasarkan kondisi yang ada, untuk menghasilkan prediksi akhir yang optimal. Dengan menggabungkan prediksi dari beberapa pohon keputusan, *Random Forest* dapat menghasilkan prediksi yang lebih stabil dan akurat untuk data baru (Rahmi dkk., 2023)

2.4 Klasifikasi

Klasifikasi merupakan salah satu teknik dalam *machine learning* yang termasuk dalam kelompok pembelajaran terawasi (*supervised learning*), di mana tujuan utamanya adalah memprediksi kelas atau kategori dari data input. Sebelum prediksi dapat dilakukan, model perlu dilatih menggunakan data khusus yang disebut data latih. Setelah model dilatih, data uji digunakan untuk memverifikasi keakuratan prediksi (M. Reza Faisal & Dodon T. Nugrahadi, 2020)

Dalam konteks *Machine Learning* Klasifikasi adalah proses menempatkan objek ke dalam kelas tertentu berdasarkan atribut-atribut yang dimilikinya.

Metode ini melibatkan pembuatan model menggunakan data latih yang sudah ada, kemudian model tersebut digunakan untuk mengklasifikasikan data baru. Menurut (Faisal & Abadi, 2020), klasifikasi juga dapat digunakan untuk menemukan atau menentukan sebuah model dan fungsi yang mendefinisikan suatu kelas atau membedakan antar kelas, dengan tujuan memprediksi kelas yang belum diketahui dari suatu objek.

2.5 Prediksi dalam *Machine Learning*

Dalam konteks *Machine Learning*, prediksi adalah proses di mana model belajar dari data yang ada untuk membuat perkiraan atau estimasi tentang hasil yang akan datang berdasarkan pola yang teridentifikasi dalam data tersebut. Model *Machine Learning* menggunakan teknik statistik dan algoritma untuk mempelajari hubungan antara variabel *input* dan *output* dari data latih, sehingga dapat menghasilkan prediksi yang akurat untuk data baru yang belum pernah dilihat sebelumnya (Kafil, 2020).

Prediksi adalah proses sistematis untuk memperkirakan kejadian di masa depan berdasarkan informasi dari masa lalu dan saat ini, dengan tujuan mengurangi kesalahan antara hasil prediksi dan kenyataan. Prediksi membantu memberikan pandangan yang lebih jelas tentang masa depan dan menghubungkan data besar untuk mempermudah pengambilan keputusan. Meskipun prediksi tidak memberikan jawaban pasti, namun berusaha untuk memberikan perkiraan yang semakin mendekati kejadian yang akan terjadi (Afrizal dkk., 2023). Pada umumnya, model prediksi sangat bermanfaat bagi sebuah perusahaan atau organisasi dalam mengantisipasi kebutuhan pasar dan mengambil keputusan yang

lebih baik. Sehingga hasil prediksi digunakan untuk meningkatkan fasilitas atau layanan yang lebih baik sehingga meningkatkan kualitas sebuah organisasi atau lembaga. Dengan memiliki pemahaman yang lebih baik tentang preferensi seseorang, maka sebuah organisasi dapat mengoptimalkan kinerjanya (Hutabarat, 2020).

2.6 Evaluasi Model

Evaluasi model merupakan proses penilaian kinerja dan keefektifan model statistik atau algoritma pembelajaran mesin dalam memprediksi atau mengklasifikasikan data. Tujuannya adalah untuk memahami seberapa baik model dapat melakukan tugasnya berdasarkan data yang diberikan. Evaluasi ini merupakan langkah penting dalam proses pengembangan model karena membantu kita memahami kinerja model dan potensinya dalam menangani data baru (Daqiqil, 2021).

2.6.1 *Confusion Matrix*

Confusion Matrix adalah metode yang digunakan dalam *machine learning* untuk mengevaluasi kinerja model klasifikasi. Metode ini sangat penting karena memberikan gambaran yang jelas tentang bagaimana model melakukan prediksi yang benar maupun salah. *Confusion Matrix* menyajikan data dalam bentuk tabel, yang berisi informasi terkait hasil prediksi model pada data uji (Karimi, 2021)

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Gambar 2. 2 Ilustrasi Gambar *Confussion Matrix*

Tabel *Confusion Matrix* berisi empat elemen utama:

1. **True Positive (TP)**: Data yang seharusnya diklasifikasikan sebagai kelas positif (misalnya, spam), dan model berhasil mengklasifikasikannya dengan benar.
2. **False Positive (FP)**: Data yang seharusnya diklasifikasikan sebagai kelas negatif (bukan spam), namun model salah mengklasifikasikannya sebagai kelas positif.
3. **False Negative (FN)**: Data yang seharusnya diklasifikasikan sebagai kelas positif, tetapi model salah menebaknya sebagai kelas negatif.
4. **True Negative (TN)**: Data yang seharusnya diklasifikasikan sebagai kelas negatif, dan model berhasil mengklasifikasikannya dengan benar.

Dengan menggunakan *Confusion Matrix*, kita bisa menghitung beberapa metrik evaluasi yang penting, di antaranya:

Akurasi (*Accuracy*): Mengukur seberapa sering model membuat prediksi yang benar. Akurasi dihitung dengan membagi jumlah prediksi yang benar (TP + TN) dengan total keseluruhan prediksi (TP + TN + FP + FN). Rumusnya:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (2.3)$$

Precision adalah hasil dari rasio antara jumlah prediksi positif yang benar terhadap keseluruhan prediksi positif yang dihasilkan oleh model, dengan rumus sebagai berikut::

$$Precision = \frac{TP}{TP + FP} \quad (2.4)$$

Recall mengukur sejauh mana model berhasil mengidentifikasi data yang benar, sehingga recall dihitung sebagai rasio antara jumlah prediksi positif yang benar terhadap keseluruhan data yang seharusnya bernilai positif.

$$Recall = \frac{TP}{TP + FN} \quad (2.5)$$

F1-Score yaitu hubungan yang memiliki timbal balik pada nilai *precision* dan *recall*. Persamaan untuk menghitung nilai *F1-score*:

$$f1 - score = \frac{2 * (Precision * Recall)}{Precision + Recall} \quad (2.6)$$

2.6.2 Area Under the Curve (AUC)

Area Under the Curve (AUC) dalam konteks *machine learning* adalah ukuran performa yang digunakan untuk mengevaluasi model klasifikasi, terutama yang digunakan bersama kurva ROC (*Receiver Operating Characteristic*). AUC mengukur kemampuan model untuk membedakan antara kelas positif dan negatif secara keseluruhan. AUC banyak digunakan ketika model klasifikasi memiliki output probabilitas, karena memberikan pemahaman yang lebih menyeluruh dibandingkan dengan metrik lain seperti akurasi (Tuah dkk., 2022)

AUC adalah ukuran numerik yang menunjukkan luas area di bawah kurva ROC, yang digunakan untuk mengevaluasi kinerja model klasifikasi. Nilai AUC berkisar dari 0 hingga 1, di mana AUC = 1 menunjukkan bahwa model tersebut sempurna dan mampu memisahkan kelas positif dan negatif dengan sangat baik. Sebaliknya, AUC = 0.5 menunjukkan bahwa model tidak lebih baik daripada tebakan acak. Jika nilai AUC kurang dari 0.5, ini berarti model memiliki performa yang lebih buruk daripada tebakan acak, yang biasanya menandakan adanya kesalahan serius dalam prediksi kelas. Semakin besar nilai AUC (mendekati 1), semakin baik kemampuan model untuk membedakan antara dua kelas, sehingga AUC sering digunakan sebagai indikator kinerja model yang efektif. Ini menggambarkan rata-rata dari nilai absolut dari selisih antara prediksi model dan nilai aktual yang diamati, di mana semua perbedaan individu diberikan bobot yang sama (Annisa & Ulama, 2022).

$$AUC = \frac{1 + TPrate - FPrate}{2} \quad 2.7$$

2.7 Penelitian Terdahulu

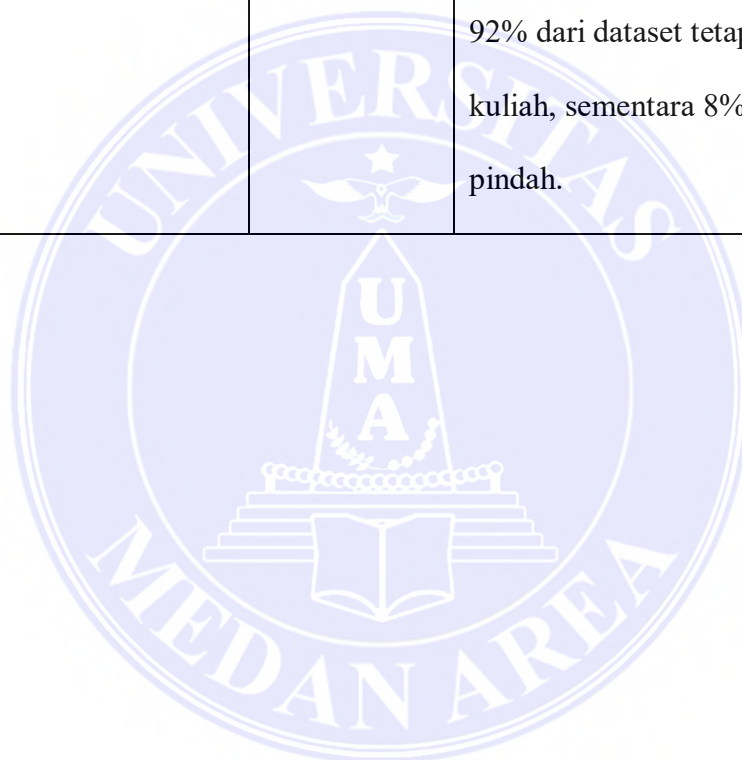
Tabel 2. 1 Penelitian Terdahulu

No.	Judul & Peneliti	Metode	Kesimpulan
1.	<i>Customer Churn Prediction on Credit Card Services using Random Forest Method</i> (Miao & Wang, 2022)	<i>Random Forest, Regresi Linier, dan K-Nearest Neighbor (KNN)</i>	Tujuan penelitian ini untuk memprediksi pelanggan <i>Churn</i> kartu kredit menggunakan metode <i>machine learning</i> . Penelitian ini menggunakan tiga model, yaitu <i>Random Forest, Regresi Linier, dan K-Nearest Neighbor (KNN)</i> . Hasil penelitian menunjukkan bahwa model <i>Random Forest</i> memiliki kinerja terbaik di bandingkan dua model lainnya dengan akurasi mencapai 96.25%.
2.	<i>Implementing Random Forest to Predict Churn</i> (Makurumidze dkk, 2022)	<i>Gradient Boost, Random Forest, Adaboost, dan Decision Tree</i>	Penelitian ini bertujuan untuk memprediksi klien bank yang berpotensi untuk melakukan <i>Churn</i> . penelitian ini menggunakan empat algoritma <i>machine learning</i> , yaitu <i>Gradient Boost, Random Forest, Adaboost, dan Decision Tree</i> . Hasil

			<p>penelitian ini menunjukkan bahwa <i>Random Forest</i> dan <i>Gradient Boosting</i> memiliki kinerja yang lebih baik dibandingkan dengan metode lain seperti <i>Adaboost</i> dan <i>Decision Tree</i>. Evaluasi model menunjukkan bahwa <i>Random Forest</i> memiliki akurasi yang tinggi yaitu 96%, dengan nilai presisi 93%, <i>recall</i> 84%, <i>F1-score</i> 88%.</p>
3.	<p><i>Using Machine Learning Methods to Predict Subscriber Churn of a Web-Based Drug Information Platform</i> (Theodoridis & Tsadiras, 2021)</p>	<p><i>Support Vector Machine, Random Forest Extreme Gradient Boosting, Light Gradient Boosting Machine, Neural</i></p>	<p>Tujuan dari penelitian ini untuk memprediksi pelanggan yang cenderung berpindah (<i>Churn</i>) pada platform informasi obat berbasis web, dengan menggunakan beberapa algoritma <i>machine learning</i> yaitu, <i>Support Vector Machine, Random Forest Extreme Gradient Boosting, Light Gradient Boosting Machine, Neural Network</i>. Hasil penelitian ini menunjukkan bahwa <i>Random Forest</i> lebih unggul dalam mengidentifikasi</p>

		<i>Network</i>	pelanggan yang berpotensi <i>Churn</i> sehingga sangat baik di gunakan dalam dunia nyata.
4.	<i>Predicting Student Churn From the University of Washington</i> (Velagapudi dkk., 2020)	<i>Logistic Regression, Random Forests, K-Nearest Neighbors.</i>	Penelitian ini bertujuan untuk memprediksi mahasiswa yang berisiko <i>Churn</i> atau <i>drop out</i> dari <i>University of Washington</i> . Hasil penelitian menunjukkan bahwa penggunaan model <i>Machine Learning</i> dapat mencapai tingkat akurasi yang cukup baik. Dengan masing masing Regresi Logistik mencapai akurasi sebesar 66.59% dengan AUC 0.729, <i>K-Nearest Neighbors</i> mencapai akurasi sebesar 64.60% dengan AUC 0.660, dan <i>Random Forest</i> mencapai akurasi sebesar 62.24% dengan AUC 0.694.
5.	<i>Churn Prediction on Higher Education Data with Fuzzy Logic</i>	<i>Fuzzy Logic</i>	Penelitian ini bertujuan untuk memprediksi apakah mahasiswa akan pindah atau tetap lanjut berkuliah, data yang digunakan yaitu

	<p><i>Algorithm</i></p> <p>(Supangat dkk., 2021)</p>		<p>data masiswa sejak tahun 2014-2020 dengan menggunakan variabel yaitu, absensi mahasiswa, waktu terakhir mahasiswa membayar pendaftaran, periode pembayaran terakhir, dan total pembayaran mahasiswa.</p> <p>Penelitian ini menunjukkan bahwa 92% dari dataset tetap melanjutkan kuliah, sementara 8% diprediksi akan pindah.</p>
--	--	--	---

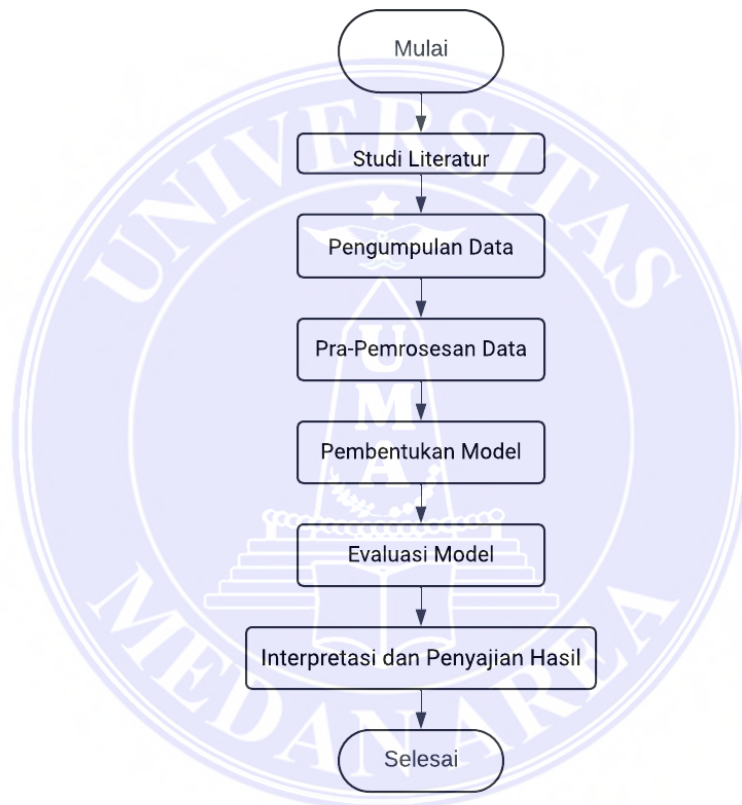


BAB III

METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

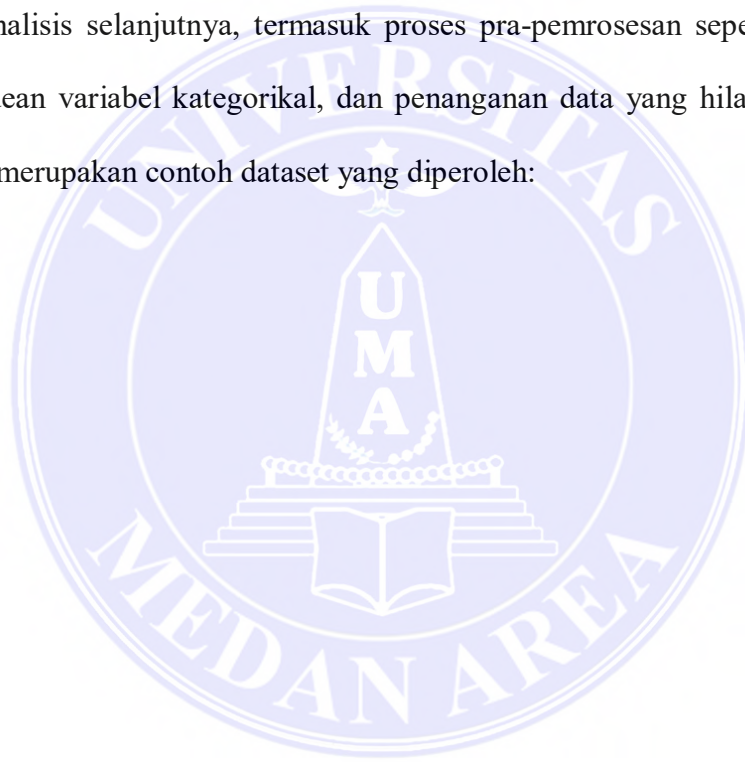
Penelitian ini berfokus pada pengembangan model prediksi untuk mengidentifikasi siswa yang akan meninggalkan sekolah (*Churn*) dan siswa yang kemungkinan akan tetap tinggal. Dengan demikian, tujuan utama adalah memahami faktor-faktor yang mempengaruhi keputusan siswa, sehingga dapat diambil tindakan preventif atau retentif yang sesuai untuk mempertahankan siswa di sekolah. Berikut tahapan penelitian yang akan dilakukan, yaitu sebagai berikut:



Gambar 3. 1 *Flowchart* Tahapan Penelitian

3.2 Pengumpulan Data

Pada tahapan ini data di peroleh melalui tahapan observasi dan wawancara di SMP Pahlawan Nasional untuk mengidentifikasi variabel-variabel yang relevan untuk proses penelitian ini. Data yang dikumpulkan mencakup berbagai variabel yang dapat mempengaruhi keputusan siswa untuk tetap tinggal atau meninggalkan sekolah (*churn*). Dataset yang digunakan dalam analisis ini berisi informasi tentang siswa, dengan fitur-fitur seperti Nilai Rata-Rata, Jumlah Absen, Ekonomi Keluarga, dan Status *Churn*. Data yang telah terkumpul akan disiapkan untuk analisis selanjutnya, termasuk proses pra-pemrosesan seperti normalisasi, pengkodean variabel kategorikal, dan penanganan data yang hilang atau outlier. Berikut merupakan contoh dataset yang diperoleh:



Tabel 3. 1 Contoh Dataset

Nama	Nilai Rata- Rata	Jumlah Absen	Ekonomi Keluarga	Status <i>Churn</i>
Azizah Permata	85	3	Rendah	Lanjut
Agustio	78	5	Sedang	Keluar
Juanda Pratama	92	1	Tinggi	Lanjut
Riani Anjarsari	80	2	Rendah	Lanjut
Eva Manurung	88	4	Sedang	Keluar

3.3 Pra-Pemrosesan Data

Pada tahap pra-pemrosesan data, data yang telah dikumpulkan akan disiapkan dan dipersiapkan untuk analisis lebih lanjut. Langkah-langkah yang dilakukan dalam tahap ini yaitu:

- 1 Data yang dikumpulkan akan diperiksa untuk mengidentifikasi dan menangani nilai-nilai yang hilang atau tidak valid. Ini dapat mencakup penghapusan baris data yang tidak lengkap, atau penggantian nilai yang hilang dengan estimasi atau nilai yang sesuai.

- 2 Kemudian Mengubah nilai kolom Status *Churn*, Ekonomi Keluarga, menjadi numerik dan kolom Status *Churn* Menjadi Numerik.

Kolom "Status *Churn* dalam dataset berisi informasi tentang status *Churn* siswa, dengan nilai "Lanjut" untuk siswa yang tetap berlanjut dan "Keluar" untuk siswa yang keluar. Namun, model machine learning bekerja lebih baik dengan data numerik. Oleh karena itu, kita perlu mengubah nilai kolom ini menjadi bentuk numerik. Dimana tahapan ini akan

Mengganti nilai "Lanjut" dengan 0.

Mengganti nilai "Keluar" dengan 1.

3.4 Pembentukan Model

Pada tahapan ini merupakan tahapan pembentukan model untuk melatih model prediksi *Churn* siswa menggunakan data pelatihan. Dimana pada penelitian ini menggunakan algoritma Random Forest untuk mengembangkan model prediksi. Proses pelatihan dilakukan dengan memasukkan data pelatihan ke dalam algoritma, yang kemudian mengidentifikasi pola dan hubungan antara fitur-fitur dalam data tersebut untuk membuat model yang dapat digunakan untuk memprediksi *Churn* siswa di masa depan. Selama proses pelatihan, parameter-parameter model dapat disesuaikan dan dioptimalkan untuk meningkatkan kinerja model. Setelah melalui proses pelatihan, model siap digunakan untuk melakukan prediksi pada data pengujian dan evaluasi kinerjanya.

3.4.1 Analisis Menggunakan Random Forest

Pada tahapan ini dilakukan analisis menggunakan rumus fungsi dari *Random Forest*.

Mengubah atribut **nilai rata – rata** dan **jumlah absen** menjadi jenis Kategorikal :
 Nilai rata – rata : **Low** (≤ 80) dan **High** (> 80).

Jumlah abses : **Sedikit** (< 3) dan **Banyak** (≥ 3).

Tabel 3. 2 Tabel Mengubah Atribut Menjadi Jenis Kategorikal

Nama	Nilai Rata – Rata	Jumlah Absen	Ekonomi Keluarga	Status
Azizah Permata	<i>High</i>	Banyak	Rendah	Keluar
Agustio	<i>Low</i>	Banyak	Sedang	Keluar
Juanda Pratama	<i>High</i>	Sedikit	Tinggi	Lanjut
Riani Anjarsari	<i>Low</i>	Sedikit	Rendah	Lanjut
Eva Manurung	<i>High</i>	Banyak	Sedang	Keluar

Adapun Rumus yang digunakan yaitu rumus pada (2.1) dan (2.2)

NODE 1

Tabel 3. 3 Tabel Node 1

Atribut	Jumlah	Churn	Keluar	Lanjut	Entropy	Gain
Total	5	1	2	2	0,07	
Nilai Rata - Rata						-0,008
<i>High</i>	3	1	1	1	0,03	
<i>Low</i>	2	0	1	1	0,15	
Jumlah Absen						-0,11
Banyak	3	1	2	0	0,10	
Sedikit	2	0	0	2	0,30	

$$\begin{aligned} \text{Entropy Total} &: \left(-\frac{1}{5} \times \log_2\left(\frac{1}{5}\right)\right) + \left(\frac{2}{5} \times \log_2\left(\frac{2}{5}\right)\right) + \left(\frac{2}{5} \times \log_2\left(\frac{2}{5}\right)\right) \\ &= 0,07 \end{aligned}$$

Menghitung Entropy Nilai Rata – Rata :

$$\text{High} : \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) = 0,03$$

$$\text{Low} : \left(-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) + \left(\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 0,15$$

Menghitung Gain Nilai Rata – Rata :

$$\text{Gain(Nilai Rata - Rata)} = 0,07 - \left(\frac{3}{5}(0,03) + \frac{2}{5}(0,15)\right) = -0,008$$

Menghitung Entropy Jumlah absen :

$$\mathbf{Banyak} : \left(-\frac{1}{3} \times \log_2 \left(\frac{1}{3} \right) \right) + \left(\frac{2}{3} \times \log_2 \left(\frac{2}{3} \right) \right) + \left(\frac{0}{3} \times \log_2 \left(\frac{0}{3} \right) \right) = 0,10$$

$$\mathbf{Sedikit} : \left(-\frac{0}{2} \times \log_2 \left(\frac{0}{2} \right) \right) + \left(\frac{0}{2} \times \log_2 \left(\frac{0}{2} \right) \right) + \left(\frac{2}{2} \times \log_2 \left(\frac{2}{2} \right) \right) = 0,30$$

Menghitung Gain Jumlah absen :

$$\mathbf{Gain(Jumlah absen)} = 0,07 - \left(\frac{3}{5} (0,10) + \frac{2}{5} (0,30) \right) = -0,11$$

Sehingga yang menjadi node 1 yaitu nilai rata – rata. Atribut yang dimiliki adalah high dan low. Kemudian perhitungan dilakukan lebih lanjut sebab nilai tersebut belum bisa diklasifikasikan menjadi satu.

NODE 1.1 (Nilai Rata – Rata == High)

Nama	Nilai Rata – Rata	Jumlah Absen	Ekonomi Keluarga	Status
Azizah Permata	High	Banyak	Rendah	Keluar
Juanda Pratama	High	Sedikit	Tinggi	Lanjut
Eva Manurung	High	Banyak	Sedang	Keluar

Tabel 3. 4 Tabel Node 1.1 Nilai Rata – Rata

Node	Atribut	Jumlah	Churn	Keluar	Lanjut	Entropy	Gain
1.1	Total	3	1	1	1	0,03	
	Jumlah Absen						-0,07
	Banyak	2	1	1	0	0	
	Sedikit	1	0	0	1	0,30	
	Ekonomi Keluarga						-0,27
	Tinggi	1	0	0	1	0,30	
	Sedang	1	0	1	0	0,30	
	Rendah	1	1	0	0	-0,30	

Menghitung *Entropy* Total :

$$\begin{aligned} \text{Entropy Total} &: \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) \\ &= 0,03 \end{aligned}$$

Menghitung *Entropy* Jumlah absen :

$$\text{Banyak} : \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) = 0$$

$$\text{Sedikit} : \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0,30$$

Menghitung *Gain* Jumlah absen:

$$\text{Gain(Jumlah absen)} = 0,03 - \left(\frac{2}{3}(0) + \frac{1}{3}(0,30)\right) = -0,07$$

Menghitung *Entropy* Ekonomi Keluarga :

$$\text{Tinggi} : \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0,30$$

$$\text{Sedang} : \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0,30$$

$$\text{Rendah} : \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = -0,30$$

Menghitung *Gain* Ekonomi keluarga :

$$\begin{aligned} \text{Gain(ekonomi keluarga)} &= 0,03 - \left(\frac{1}{3}(0,30) + \frac{1}{3}(0,30) + \frac{1}{3}(0,30)\right) \\ &= -0,27 \end{aligned}$$

Kemudian jumlah absen akan di jadikan node 1.1. Atribut yang dimiliki adalah banyak dan sedikit. Dilanjutkan melakukan perhitungan dikarenakan nilai tersebut masih belum diklasifikasikan menjadi satu.

NODE 1.2 (Nilai Rata – Rata == *Low*)

Nama	Nilai Rata – Rata	Jumlah Absen	Ekonomi Keluarga	Status
Agustio	<i>Low</i>	Banyak	Sedang	Keluar
Riani Anjarsari	<i>Low</i>	Sedikit	Rendah	Lanjut

Tabel 3. 5 Tabel Node 1.2 Nilai Rata – Rata

Atribut		Jumlah	Churn	Keluar	Lanjut	Entropy	Gain
Total		2	1	1	1	0,07	
Ekonomi Keluarga							-0,23
	Tinggi	0	0	0	0	0	
	Sedang	1	0	1	0	0,30	
	Rendah	1	0	0	1	0,30	
Nilai Rata – Rata							-0,12
	High	0	0	0	0	0	
	Low	2	0	1	1	0,15	

$$\begin{aligned} \text{Entropy Total} &: \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) \\ &= 0,07 \end{aligned}$$

Menghitung *Entropy* Ekonomi Keluarga :

$$\text{Tinggi} : \left(-\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) + \left(\frac{0}{0} \times \log_2\left(\frac{0}{0}\right)\right) = 0$$

$$\text{Sedang} : \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0,30$$

$$\text{Rendah} : \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0,30$$

Menghitung *Gain* Ekonomi Keluarga :

$$\text{Gain(ekonomi keluarga)} = 0,07 - \left(\frac{0}{2}(0) + \frac{1}{2}(0,30) + \frac{1}{2}(0,30)\right) = -0,23$$

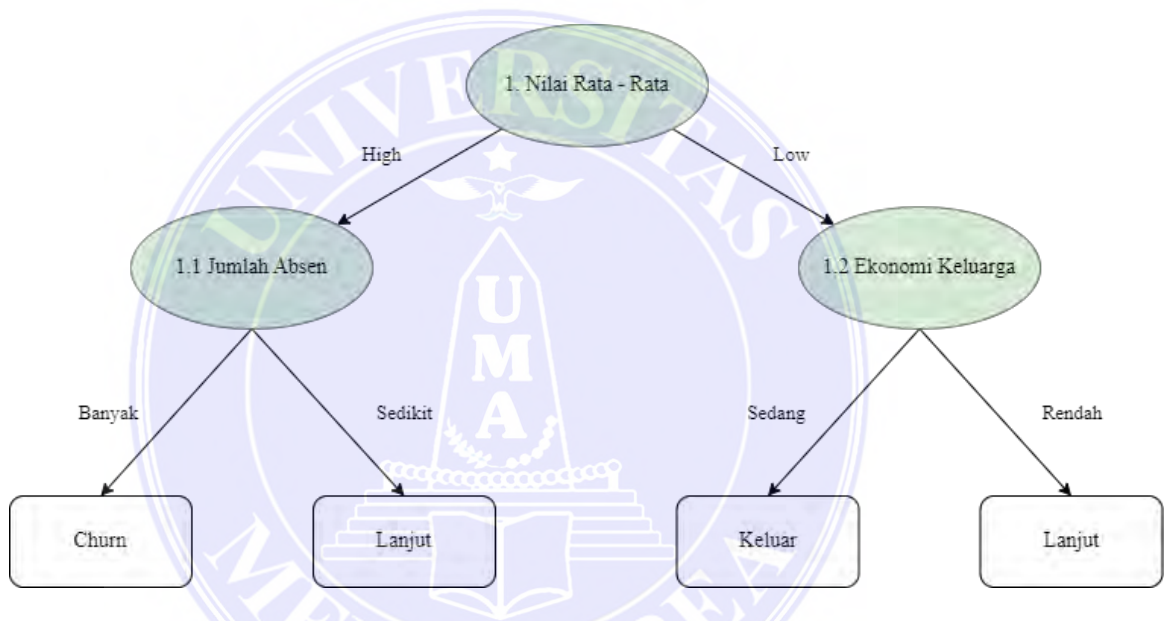
Menghitung *Entropy* Nilai Rata - Rata :

$$\mathbf{High} : \left(-\frac{0}{0} \times \log_2 \left(\frac{0}{0} \right) \right) + \left(\frac{0}{0} \times \log_2 \left(\frac{0}{0} \right) \right) + \left(\frac{1}{0} \times \log_2 \left(\frac{1}{0} \right) \right) = 0$$

$$\mathbf{Low} : \left(-\frac{0}{2} \times \log_2 \left(\frac{0}{2} \right) \right) + \left(\frac{1}{2} \times \log_2 \left(\frac{1}{2} \right) \right) + \left(\frac{1}{2} \times \log_2 \left(\frac{1}{2} \right) \right) = 0,15$$

Menghitung *Gain* Nilai Rata - Rata :

$$\mathbf{Gain}(\mathbf{Nilai\ Rata - Rata}) = 0,03 - \left(\frac{0}{2}(0) + \frac{2}{2}(0,15) \right) = -0,12$$



Gambar 3. 2 Visualisasi Pohon Keputusan

Setelah dilakukan analisis menggunakan rumus *Random Forest* maka didapatkan pohon keputusan pada Gambar (3.2) yang mana yaitu:

1. Jika nilai rata – rata nya *low* (kurang atau sama dengan 80) dan memiliki ekonomi keluarga rendah maka statusnya lanjut.

2. Jika nilai rata – rata nya *low* (kurang atau sama dengan 80) dan memiliki ekonomi keluarga sedang maka statusnya keluar.
3. Jika nilai rata – rata nya *high* (lebih dari 80) dan memiliki jumlah absen sedikit (kurang dari 3) maka statusnya lanjut.
4. Jika nilai rata – rata nya *high* (lebih dari 80) dan memiliki jumlah absen banyak (lebih atau sama dengan 3) maka statusnya *Churn*.

3.5 Evaluasi Model

Pada tahapan ini dilakukan evaluasi pada model yang telah dibentuk, model prediksi *Churn* siswa akan di periksa lebih lanjut untuk melihat seberapa baik kinerja model sebelum diterapkan di lingkungan nyata. Evaluasi model dilakukan menggunakan menggunakan *Confussion matrix* dan AUC. Evaluasi model dilakukan menggunakan data uji yang tidak digunakan selama pelatihan. Hal tersebut dilakukan untuk memastikan bahwa model dapat menggeneralisasi pola dari data yang belum pernah dilihat sebelumnya dan bukan hanya menghafal data pelatihan.

3.6 Alat dan Bahan Penelitian

Adapun alat dan bahan yang akan digunakan pada penelitian ini yaitu terdiri dari perangkat keras dan perangkat lunak yang ditunjukkan pada tabel dibawah ini:

a. Perangkat Keras

Adapun kebutuhan perangkat keras yang digunakan dalam penelitian ini yaitu:

Tabel 3. 6 Kebutuhan Perangkat Keras

No	Nama Perangkat	Spesifikasi
1.	Device	LAPTOP HP 14s-769BU8BD
2.	Processor	Intel(R) Core(TM) i3-10110U CPU @ 2.10GHz 2.59 GHz
3.	SSD	239 GB
4.	RAM	4.00 GB

b. Perangkat Lunak

Adapun kebutuhan perangkat lunak yang digunakan pada penelitian ini yaitu sebagai berikut:

Tabel 3. 7 Kebutuhan Perangkat Lunak

No	Nama Perangkat	Spesifikasi
1	<i>Operating System</i> Laptop	Windows 11
2	Bahasa Pemograman	<i>Python</i>
3	<i>Cloud Computing</i>	Google Colab
4	Penyimpanan Data	Google Drive

BAB V KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil evaluasi dan pembahasan yang telah dilakukan, dapat disimpulkan bahwa:

Berdasarkan hasil penelitian ini, model *Random Forest* yang diterapkan untuk memprediksi *Churn* siswa di SMP Pahlawan Nasional menunjukkan performa yang baik dengan akurasi sebesar 85.7%. Model ini juga memiliki *precision* 89.6% dan *recall* 81.2%, yang menunjukkan bahwa model cukup efektif dalam mengidentifikasi siswa yang berpotensi *Churn*. Nilai *F1-Score* sebesar 85.2% mencerminkan keseimbangan yang baik antara kemampuan model dalam mendeteksi churn dan meminimalkan kesalahan klasifikasi. Dengan demikian, dapat disimpulkan bahwa model *Random Forest* merupakan alat yang berguna untuk membantu sekolah dalam mengidentifikasi siswa yang mungkin akan keluar, sehingga langkah-langkah pencegahan dapat diambil untuk meningkatkan retensi siswa.

5.2 Saran

1. Saran untuk Penelitian Selanjutnya dapat melakukan penambahan Dataset dengan mengumpulkan lebih banyak data siswa untuk meningkatkan ukuran dataset. Hal ini akan membantu model dalam mengidentifikasi pola yang lebih kompleks dan menggeneralisasi lebih baik pada data baru.

2. Disarankan untuk melakukan penyempurnaan model dengan menambahkan fitur-fitur lain yang mungkin relevan, seperti data perilaku siswa atau faktor psikososial yang dapat mempengaruhi status *Churn*.
3. Untuk penelitian selanjutnya untuk dapat melakukan penelitian lebih lanjut dengan berbagai kombinasi *hyperparameter* untuk meningkatkan akurasi model. Metode lain seperti *Random Search* atau *Bayesian Optimization* dapat digunakan untuk tuning *hyperparameter*.
4. Penelitian selanjutnya dapat melakukan evaluasi dengan menggunakan model yang telah dibangun sebagai alat bantu untuk mengidentifikasi siswa yang berisiko *Churn*. Dengan demikian, sekolah dapat merancang dan melaksanakan intervensi yang tepat waktu dan tepat sasaran untuk membantu siswa melanjutkan studi.

DAFTAR PUSTAKA

- Afrizal, M., Siswanto, & Sudarsono, A. (2023). *Implementation of Forecasting Sales of Electronic Goods Using the Semi Average Method at Alex Electronics Store*. 2(2), 319–330.
- Allaam, A. (2023). Prediksi Churn Konsumen Menggunakan Algoritma *Random Forest* dengan *Fuzzy C-Means* untuk Meningkatkan Produktivitas Penjualan Bisnis Prediksi Churn Konsumen Menggunakan Algoritma *Random Forest* dengan *Fuzzy C-Means* untuk Meningkatkan Produktivitas Penjualan.
- Annisa, Z., & Ulama, B. S. S. (2022). Analisis Sentimen Data Ulasan Pengguna Aplikasi “ PeduliLindungi ” pada Google *Play Store* Menggunakan Metode *Naïve Bayes Classifier* Model Multinomial. 11(6).
- Atthariq, A. S. (2020). *Klasifikasi Customer Churn Berdasarkan Segmentasi Pelanggan Menggunakan Algoritma Naïve Bayes* (Studi Kasus : Esl Express Tasikmalaya).
- Behr, A., Giese, M., K, H. D. T., & Theune, K. (2020). *Early Prediction of University Dropouts – A Random Forest Approach*.
- Budi Lestari, A. Y., Kurniawan, F., & Bayu Ardi, R. (2020). Penyebab Tingginya Angka Anak Putus Sekolah Jenjang Sekolah Dasar (SD). *Jurnal Ilmiah Sekolah Dasar*, 4(2), 299. <https://doi.org/10.23887/jisd.v4i2.24470>
- Daqiqil, I. (2021). *Machine Learning : Teori , Studi Kasus dan Implementasi Menggunakan Python*. July. <https://doi.org/10.5281/zenodo.5113507>

- Faisal, M. R., & Abadi, F. (2020). *Pemrograman Web Dasar I* (Nomor August).
- Fatma, Y. L., & Rochmawati, N. (2024). Prediksi Siswa Putus Sekolah Menggunakan Algoritma *Decision Tree C4.5*. *Journal of Informatics and Computer Science*, *05*, 486–493.
- Fauszt, T., Dóra, K. E., Bognár, D. L., & Kovács, E. (2021). *Design of a Machine Learning Model to Predict Student Attrition. More than one article*, *18*(November 2021), 164–187.
- Ghisyani, A., Putri, A. Z., Farhanna, K. P., & Indrayatna, F. (2023). Prediksi Premi Asuransi dengan Pendekatan Regresi Linear dalam Konteks *Machine Learning*. Seminar Nasional Statistika Aktuaria II (2023).
- Guru dan Tenaga Kependidikan Pendidikan Dasar. (2020). Mmodul Belajar Mandiri Calon Guru Pegawai Pemerintah dengan Perjanjian Kerja (PPPK). In *Mmodul Belajar Mandiri Calon Guru Pegawai Pemerintah dengan Perjanjian Kerja (PPPK)*.
- Hakim, A. (2020). *FAKTOR PENYEBAB ANAK PUTUS SEKOLAH*.
- Hastari, R. (2023). Manajemen Strategik Dalam Meningkatkan Daya Saing Di Smp Ma ' Arif Nu 1 Ajibarang. *Tesis UIN Profesor Kiai Haji Saifudin Zuhri Purwokerto*. https://eprints.uinsaizu.ac.id/19450/1/Rina_Hastari_Manajemen_Strategik_Dalam_Meningkatkan_Daya_Saing_di_SMP_Ma%27arif_NU_1_Ajibarang_Kabupaten_Banyumas.pdf
- Hutabarat, C. (2020). *Penerapan Data Mining Untuk Memeprediksi Permintaan*

Produk Kartu Perdana Internet Menggunakan- ALGORITMA C5.0 (Studi Kasus: Vidha Ponsel). 6(April), 419–424.

Kafil, M. (2020). *Penerapan Metode K-NEAREST NEIGHBORS Untuk Prediksi Penjualan Berbasis Web Pada Boutiq Dealove Bondowoso. 3(2), 59–66.*

Karimi, Z. (2021). *Confusion Matrix. October.*

Keppens, G. (2023). *School absenteeism and academic achievement : Does the timing of the absence matter ? 86(March).*

Kurniawan, I., Buani, D. C. P., Abdussomad, A., Apriliah, W., & Saputra, R. A. (2023). Implementasi Algoritma Random Forest Untuk Menentukan Penerima Bantuan Raskin. *Jurnal Teknologi Informasi dan Ilmu Komputer, 10(2), 421–428.* <https://doi.org/10.25126/jtiik.20231026225>

M. Reza Faisal, & Dodon T. Nugrahadi. (2020). Data Science Klasifikasi dengan Bahasa Pemrograman R. In *Patra Widya: Seri Penerbitan Penelitian Sejarah dan Budaya*. (1 ed., Vol. 21, Nomor 3). Scripta Cendekia.

Makurumidze, L., Manjoro, W. S., & Makondo, W. (2022). Implementing Random Forest to Predict Churn. *International Journal of Computer Science and Mobile Computing, 11(2), 75–84.* <https://doi.org/10.47760/ijcsmc.2022.v11i02.009>

Malikireddy, V. P., & Kasa, M. (2021). Customer Churns Prediction Model Based on Machine Learning Techniques: A Systematic Review. *Proceedings of the 3rd International Conference on Integrated Intelligent Computing*

Communication & Security (ICIIC 2021), 4(Iciic), 167–174.

<https://doi.org/10.2991/ahis.k.210913.021>

Mathai, P. P. (2020). *Customer Churn Prediction : A Survey*. August.

Miao, X., & Wang, H. (2022). Customer Churn Prediction on Credit Card Services using Random Forest Method. *Proceedings of the 2022 7th International Conference on Financial Innovation and Economic Development (ICFIED 2022)*, 648(Icfied), 649–656.
<https://doi.org/10.2991/aebmr.k.220307.104>

Nurhuda. (2022). *Landasan Pendidikan Nurhuda*. www.ahlmediapress.com

Rahayu, P. W., Sudipa, I. G. I., Suryani, Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, M. N., Slamet, I., Harlina, S., & Maysanjaya, I. M. D. (2024). *Buku ajar data mining* (Efitra, I. K. Sari, & Y. Agusdi (ed.); Nomor January). PT. Sonpedia Publishing Indonesia.

Rahmadani, A., & Daulay, N. (2023). *Analisis Faktor Penyebab Menurunnya Resiliensi Akademik Pada Siswa MTsN*. 9623, 417–427.

Rahmawati, S., Indriayu, M., & Sabandi, M. (2022). *Pengaruh Tekanan Akademik Terhadap Prestasi Akademik Mahasiswa Pendidikan Ekonomi Fakultas Keguguran Dan Ilmu Pendidikan Universitas Sebelas Maret*.

Rahmi, I. A., Afendi, F. M., & Kurnia, A. (2023). *Metode AdaBoost dan Random Forest untuk Prediksi Peserta JKN-KIS yang Menunggak*. 5(1), 83–94.

Sandag, G. A. (2020). *Prediksi Rating Aplikasi App Store Menggunakan*

Algoritma Random Forest Application Rating Prediction on App Store using Random Forest Algorithm. 6(2), 167–178.

Setiawati, S. M. (2020). Perilaku Membolos: Penyebab, Dampak, dan Solusi. *Proceeding Seminar dan Lokakarya Nasional Bimbingan dan Konseling 2021, 99–108.*

Supangat, Saringat, M. Z. Bin, Kusnanto, G., & Andrianto, A. (2021). Churn Prediction on Higher Education Data with Fuzzy Logic Algorithm. *Sisforma, 8(1), 22–29.* <https://doi.org/10.24167/sisforma.v8i1.3025>

Theodoridis, G., & Tsadiras, A. (2021). Using Machine Learning Methods to Predict Subscriber Churn of a Web-Based Drug Information Platform. *HAL Open Science.*

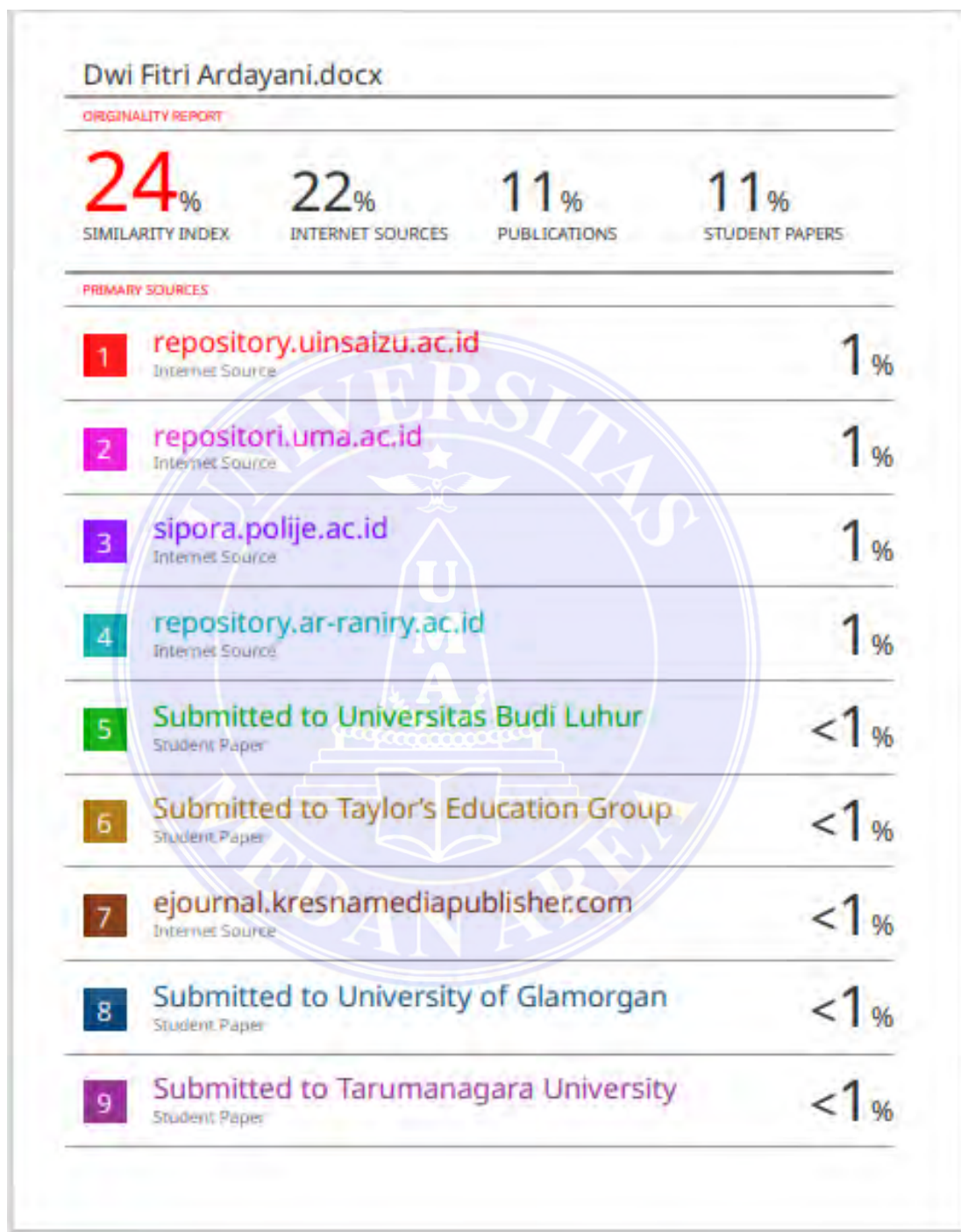
Tuah, A. N., Ibrahim, A. B., Dzulkifly, S., Mohammad Yusof, F., Awang Nor, R., & Ariffin, R. (2022). Analysis of the Area Under a Curve (AUC) using C-programming: Trapezium and Simpson rules techniques. *Journal of ICT in Education, 9(1), 143.*

Velagapudi, N., Aulck, L., Blumenstock, J., & West, J. (2020). *Predicting Student Churn From the University of Washington Predicting Length of Stay. 729.*

Wicaksono, A., Anita, & Padilah, T. N. (2021). *Uji Performa Teknik Klasifikasi untuk Memprediksi Customer Churn. 9(1).*

LAMPIRAN

Hasil Plagiasi





Nomor : 928/FT/01.10/III/2024

1 April 2024

Lampiran :-

Hal : Pembimbing Tugas Akhir

Yth. Pembimbing Tugas Akhir

ANDRE HASUDUNGAN LBS S.Ti, M.Sc (Sebagai Pembimbing)

di Tempat

Dengan hormat, sehubungan telah dipenuhinya persyaratan untuk memperoleh Tugas Akhir dari mahasiswa atas :

Nama : DWI FITRI ARDAYANI SEMBIRING

NIM : 198160075

Jurusan : TEKNIK INFORMATIKA

Maka dengan hormat kami mengharapkan kesediaan saudara :

ANDRE HASUDUNGAN LBS S.Ti, M.Sc (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :

"Prediksi Churn Siswa Siswi SMP Pahlawan Nasional dengan Menggunakan Algoritma Random Forest"

SK Pembimbing ini berlaku selama enam bulan terhitung sejak SK ini diterbitkan. Jika proses pembimbing melebihi batas waktu yang telah ditetapkan, SK ini dapat ditinjau ulang.

Demikian kami sampaikan, atas kesediaan saudara diucapkan terima kasih.

Dekan,



Dr Eng. Supriatno.ST, MT.





UNIVERSITAS MEDAN AREA

FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PBSI Nomor 1 ☎ (061) 7366878, 7360168, 7364348, 7366781, Fax (061) 7366998 Medan 20223
 Kampus II : Jalan Setimbuk Nomor 79 / Jalan Sei Serayu Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122
 Website www.teknik.uma.ac.id E-mail unw_medanarea@uma.ac.id

Nomor : 249 /FT.6/01.10/V/2024 30 Mei 2024
 Lamp : -
 Hal : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Kepala Sekolah SMP Pahlawan Nasional
 Jln. Durung No. 205
 Di
 Medan

Dengan hormat,
 Kami mohon kesediaan Bapak/Ibu berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	NAMA	NPM	PRODI
1	Dwi Fitri Ardayani Sembiring	198160075	Teknik informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir pada perusahaan/Instansi yang Bapak/Ibu Pimpin.

Perlu kami jelaskan bahwa Pengambilan Data tersebut adalah semata-mata untuk tujuan ilmiah dan Skripsi yang merupakan salah satu syarat bagi mahasiswa tersebut untuk mengikuti ujian sarjana pada Fakultas Teknik Universitas Medan Area dan tidak untuk dipublikasikan, dengan judul penelitian :

Prediksi Churn Siswa Siswi SMP Pahlawan Nasional dengan Menggunakan Algoritma Random Forest

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.

Dekan,

 Dwi Fitri Ardayani Sembiring, ST, MT
 FAKULTAS TEKNIK

Tembusan :
 1. Ka. BPMPP
 2. Mahasiswa
 3. File





SURAT KETERANGAN

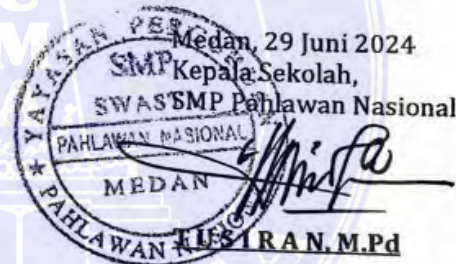
No.: 283 / SMP PAHNAS / VI/2024

Sehubungan dengan surat dari Fakultas Teknik Nomor : 249/FT.6/01.10/V/2024, Hal : Penelitian dan Pengambilan Data Tugas Akhir pada tanggal 30 Mei 2024, maka Kepala Sekolah SMP Pahlawan Nasional dengan ini menerangkan nama mahasiswa dibawah ini :

Nama : Dwi Fitri Ardayani Sembiring
NIM : 198160075
Program Studi : Teknik Informatika
Jenjang : S1

Benar telah selesai melakukan penelitian di SMP pahlawan Nasional pada tanggal 31 Mei s.d 29 Juni 2024 guna melengkapi data pada penyusunan Skripsi (Karya Ilmiah) yang berjudul : **"Prediksi Churn Siswa Siswi SMP Pahlawan Nasional dengan Menggunakan Algoritma Random Forest"**.

Demikian Surat Keterangan diperbuat untuk dapat dipergunakan seperlunya.



```

1s # Import library yang dibutuhkan
import pandas as pd # Untuk membaca dan memanipulasi data dalam format DataFrame
import io # Untuk operasi input/output, terutama untuk menangani file
from google.colab import files # Untuk mengunggah file dari komputer pengguna ke Google Colab
from IPython.display import display # Untuk menampilkan DataFrame dalam format yang rapi
from sklearn.model_selection import train_test_split # Untuk membagi data menjadi set pelatihan dan pengujian
from sklearn.preprocessing import LabelEncoder # Untuk mengonversi fitur kategoris menjadi numerik
from sklearn.ensemble import RandomForestClassifier # Untuk membuat model Random Forest
from sklearn.metrics import classification_report, accuracy_score, ConfusionMatrixDisplay # Untuk menghitung metrik evaluasi
import matplotlib.pyplot as plt # Untuk visualisasi data
import seaborn as sns # Untuk visualisasi data yang lebih interaktif
from sklearn.metrics import roc_curve, roc_auc_score # Untuk menghitung ROC curve dan AUC score
from sklearn.tree import plot_tree # Untuk visualisasi pohon keputusan (Decision Tree)
from sklearn.tree import DecisionTreeClassifier

21s # Fungsi untuk mengunggah dan menampilkan file CSV
def upload_and_display_csv():
    # Mengupload file
    uploaded = files.upload()

    # Membaca file CSV ke dalam DataFrame
    for filename in uploaded.keys():
        df = pd.read_csv(io.BytesIO(uploaded[filename])) # Mengonversi file CSV yang diunggah menjadi DataFrame

    # Menampilkan seluruh data dalam DataFrame
    display(df)
    return df

# Memanggil fungsi untuk mengunggah dan menampilkan DataFrame
df = upload_and_display_csv()

```

Choose Files dataset_balanced (1).csv

- dataset_balanced (1).csv(text/csv) - 51264 bytes, last modified: 11/22/2024 - 100% done

Saving dataset_balanced (1).csv to dataset_balanced (1) (1).csv

	Nama	Nilai_Rata-Rata	Jumlah_Absen	Ekonomi_Keluarga	Status_Churn
0	Heryanto Utama	98.91	20	Rendah	Lanjut
1	Nardi Padmasari	76.34	5	Sedang	Lanjut
2	Paiman Palastri	96.67	8	Tinggi	Lanjut
3	Atma Waluyo	94.04	15	Tinggi	Lanjut
4	Shakila Maryadi	90.39	5	Tinggi	Lanjut
...
1355	Malda Nashiruddin	61.25	16	Tinggi	Keluar
1356	Marwata Siregar	95.60	13	Sedang	Keluar
1357	Saadat Sitorus	66.15	17	Sedang	Keluar
1358	Rizki Farida	63.44	17	Rendah	Keluar
1359	Mustika Sitompul	67.57	8	Sedang	Keluar

1360 rows x 5 columns

```

0s [10] # Hapus kolom 'Nama' yang tidak diperlukan (jika ada)

```

```

[10] # Hapus kolom 'Nama' yang tidak diperlukan (jika ada)
#Menghapus kolom Nama dari df karena dianggap tidak diperlukan dalam analisis.
data_cleaned = df.drop(columns=['Nama'], errors='ignore')

#Menggunakan LabelEncoder untuk mengubah data kategoris pada kolom Ekonomi_Keluarga dan Status_Churn menjadi numerik.
# Encode fitur kategori: 'Ekonomi_Keluarga' dan 'Status_Churn'
label_encoder_ekonomi = LabelEncoder()
label_encoder_churn = LabelEncoder()

data_cleaned['Ekonomi_Keluarga'] = label_encoder_ekonomi.fit_transform(data_cleaned['Ekonomi_Keluarga'])
data_cleaned['Status_Churn'] = label_encoder_churn.fit_transform(data_cleaned['Status_Churn'])

[11] #Memisahkan Fitur dan Target, serta Split Data
X = data_cleaned.drop(columns=['Status_Churn']) # X merupakan fitur (semua kolom kecuali Status_Churn)
y = data_cleaned['Status_Churn'] #sedangkan y adalah target yang akan diprediksi (Status_Churn).

# Pisahkan data menjadi set pelatihan dan pengujian (80% pelatihan, 20% pengujian)
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)

[12] #Pelatihan Model Random Forest
# Inisialisasi model Random Forest
rf_classifier = RandomForestClassifier(n_estimators=100, random_state=42)

#inisialisasi model Random Forest dengan n_estimators=100.
rf_classifier.fit(X_train, y_train) #Melatih model (rf_classifier.fit) dengan set pelatihan.

# Prediksi pada set pengujian
y_pred = rf_classifier.predict(X_test) #Memprediksi hasil pada set pengujian (y_pred).

[12] # Prediksi pada set pengujian
y_pred = rf_classifier.predict(X_test) #Memprediksi hasil pada set pengujian (y_pred).

# Akurasi model
accuracy = accuracy_score(y_test, y_pred) #Menghitung akurasi model dengan accuracy_score().
print(f"Model Random Forest memiliki akurasi sebesar {accuracy * 100:.2f}%")

Model Random Forest memiliki akurasi sebesar 85.66%

[13] #Evaluasi Model
# Laporan klasifikasi (precision, recall, F1-score)
report = classification_report(y_test, y_pred, target_names=['Non-Churn (0)', 'Churn (1)'])
print("Detail Kinerja Model:\n")
print(report)
#Menghasilkan laporan klasifikasi yang berisi precision, recall, F1-score, dan akurasi masing-masing kelas

Detail Kinerja Model:

              precision    recall  f1-score   support

Non-Churn (0)       0.82      0.90      0.86         134
 Churn (1)          0.90      0.81      0.85         138

 accuracy
macro avg          0.86      0.86      0.86         272
weighted avg       0.86      0.86      0.86         272

[14] #Visualisasi
    
```

✓ 1s completed at 11:52 PM