

**KLASIFIKASI DATA TRANSFUSI DARAH MELALUI
PENDEKATAN ALGORITMA *RANDOM FOREST*
DAN *SUPPORT VECTOR MACHINE* (SVM)**

SKRIPSI

Oleh:

DWI KURNIA PUTRA

198160073



PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

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

Document Accepted 1/7/24

Access From (repository.uma.ac.id)1/7/24

**KLASIFIKASI DATA TRANSFUSI DARAH MELALUI
PENDEKATAN ALGORITMA *RANDOM FOREST*
DAN *SUPPORT VECTOR MACHINE* (SVM)**

SKRIPSI

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh

Gelar Sarjana di Fakultas Teknik

Universitas Medan Area



Oleh:

DWI KURNIA PUTRA

198160073

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS MEDAN AREA

MEDAN

2024

UNIVERSITAS MEDAN AREA

© Hak Cipta Di Lindungi Undang-Undang

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

i
Document Accepted 1/7/24

HALAMAN PENGESAHAN

Judul Skripsi : "Klasifikasi Data Trasfusi Darah Melalui Pendekatan Algoritma
Radom Forest dan Support Vector Machine (SVM)".

Nama : Dwi Kurnia Putra
NPM : 198160073
Fakultas : Teknik
Prodi : Teknik Informatika

Disetujui Oleh
Komisi Pembimbing

Dr. Ir. Rahmad Syah., M.Kom., IPM., Asean Eng., APEC Eng

Mengetahui


Dr. Eng. Supratno, ST., MT
Dekan Fakultas Teknik


S.Kom., M.Kom
Ketua Program Studi

Tanggal Lulus: 3 April 2024

HALAMAN PERNYATAAN

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

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



**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 Kurnia Putra
NPM : 198160073
Program Studi : Teknik Informatika
Fakultas : Teknik
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Medan Area Hak Bebas Royalti (*Non-Exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul : “Klasifikasi Data Transfusi Darah Melalui Pendekatan Algoritma *Random Forest* dan *Support Vector Machine (SVM)*”. Dengan Hak Bebas Royalti Non Eksklusif ini Universitas Medan Area berhak menyimpan, mengalih meida/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan skripsi saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta. Demikian pernyataan ini saya buat dengan sebenarnya.

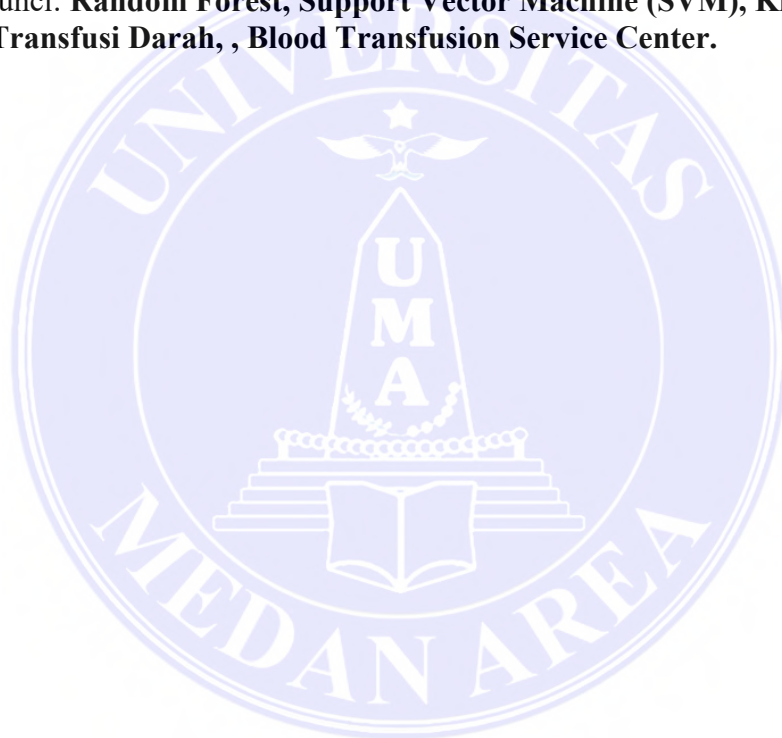
Dibuat di : Medan
Pada tanggal : 3 April 2024
Yang menyatakan


(Dwi Kurnia Putra)

ABSTRAK

Penelitian ini bertujuan untuk mengevaluasi dan membandingkan efektivitas algoritma Random Forest dan Support Vector Machine (SVM) dalam mengklasifikasikan data transfusi darah. Tujuan utama adalah untuk meningkatkan pemahaman tentang kebutuhan transfusi darah serta mengoptimalkan alokasi sumber daya dan identifikasi pasien berisiko tinggi. Studi ini menggunakan dataset Blood Transfusion Service Center, yang terdiri dari sampel yang diklasifikasikan sebagai pendonor reguler dan pendonor tidak reguler. Data dibagi menjadi 80% untuk pelatihan dan 20% untuk pengujian. Random Forest menunjukkan akurasi sebesar 72%, dan SVM mencapai akurasi 76%. Algoritma Random Forest menunjukkan kinerja yang lebih seimbang antara kedua kelas, dan SVM menunjukkan presisi yang tinggi untuk pendonor reguler tetapi kekurangan dalam mengenali pendonor tidak reguler.

Kata Kunci: Random Forest, Support Vector Machine (SVM), Klasifikasi Data, Transfusi Darah, , Blood Transfusion Service Center.



ABSTRACT

This research aimed to evaluate and compare the effectiveness of Random Forest and Support Vector Machine (SVM) algorithms in classifying blood transfusion data. The main objective was to improve understanding of blood transfusion needs as well as optimise resource allocation and identification of high-risk patients. This research used the Blood Transfusion Service Centre dataset, which consists of samples classified as regular donors and irregular donors. The data was divided into 80% for training and 20% for testing. Random Forest showed an accuracy of 72%, and SVM achieved an accuracy of 76%. Random Forest algorithm showed a more balanced performance between the two classes, and SVM showed high precision for regular donors but lacked in recognising irregular donors.

Keywords: *Random Forest, Support Vector Machine (SVM), Data Classification, Blood Transfusion, Blood Transfusion Service Centre.*



KATA PENGANTAR

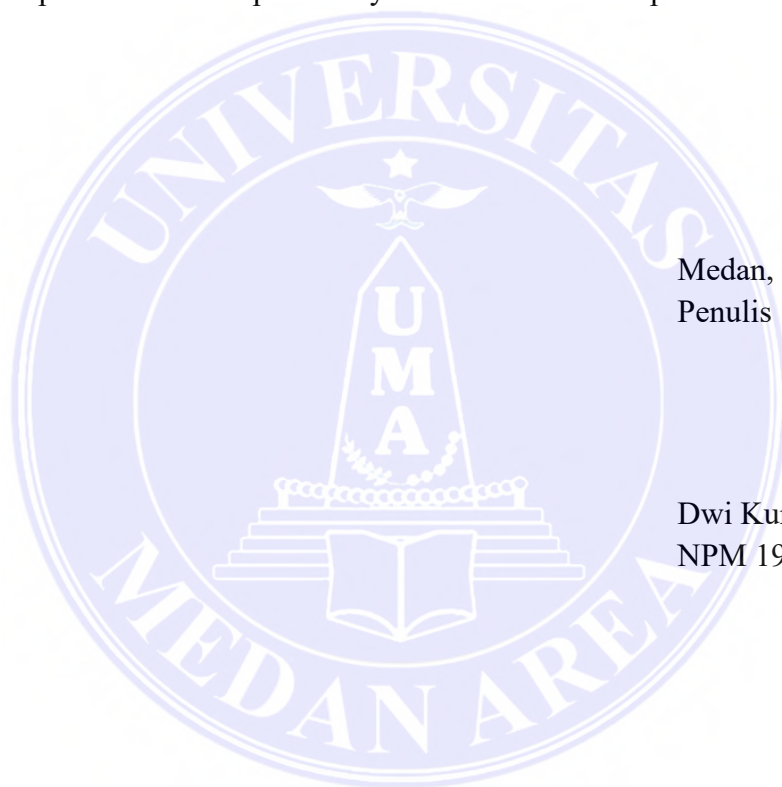
Puji Syukur kepada Tuhan Yang Maha Esa, atas berkat dan cinta kasih sehingga penulis mampu menyelesaikan Tugas Akhir yang berjudul “Klasifikasi Data Transfusi Darah Melalui Pendekatan Algoritma *Random Forest* dan *Support Vector Machine* (SVM)”. Tugas Akhir ini adalah salah satu persyaratan yang wajib dipenuhi untuk dapat lulus Sarjana Strata Satu (S1) pada Program Studi Teknik Informatika Universitas Medan Area.

Dalam menyelesaikan Tugas Akhir ini tidak lepas dari bantuan dan dukungan, baik secara moril maupun materil dari berbagai pihak. Oleh karena itu, pada kesempatan ini penulis ingin mengucapkan terima kasih kepada :

1. Tuhan Yang Maha Esa, yang telah memberikan kesehatandan keselamatan selama melakukan studi.
2. Bapak Prof. Dr. Dadan Ramdan, M. Eng, M.Sc, selaku Rektor Universitas Medan Area.
3. Bapak Dr. Eng. Supriatno, ST., MT selaku Dekan Fakultas Teknik Universitas Medan Area.
4. Bapak Rizki Muliono, S.Kom, M.Kom selaku Kepala Program Studi.
5. Bapak Dr. Ir. Rahmad Syah., M.Kom., IPM., Asean Eng., APEC Eng selaku Dosen Pembimbing.
6. Orang tua penulis Bapak Suratno dan Ibu Janiah,S.Pd yang telah mendoakan tiada henti dan memberikan semangat serta membantu penulis dalam segi materi dan moril sehingga penulis dapat menyelesaikan skripsi dengan sebaik-baiknya.

7. Teman-teman Prodi Teknik Informatika Stambuk 2019.
8. Seluruh pihak yang tidak dapat disebutkan satu persatu yang membantu dalam menyelesaikan skripsi ini.

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



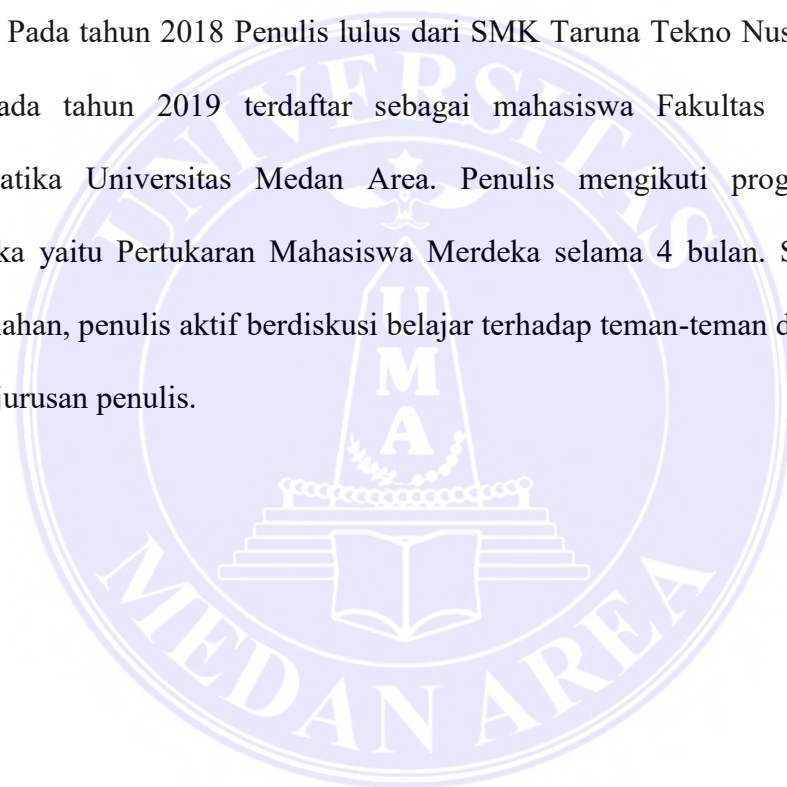
Medan, 3 April 2024
Penulis

Dwi Kurnia Putra
NPM 198160073

RIWAYAT HIDUP

Penulis dilahirkan di Kota Medan Provinsi Sumatera Utara pada tanggal 09 Februari 2000, dari Ayah Suratno dan Ibu Janiah, dan penulis merupakan anak ke 2 dari 2 bersaudara. Penulis pertama sekali sekolah di SDN 106158 Medan pada tahun 2006-2012 dan kemudian melanjutkan Sekolah Menengah Pertama di SMPN 45 Medan lulus pada tahun 2015.

Pada tahun 2018 Penulis lulus dari SMK Taruna Tekno Nusantara Medan dan pada tahun 2019 terdaftar sebagai mahasiswa Fakultas Teknik Prodi Informatika Universitas Medan Area. Penulis mengikuti program Kampus Merdeka yaitu Pertukaran Mahasiswa Merdeka selama 4 bulan. Selama proses perkuliahan, penulis aktif berdiskusi belajar terhadap teman-teman dan mahasiswa selain jurusan penulis.



DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN	Error! Bookmark not defined.
ABSTRAK	v
ABSTRACT	vii
KATA PENGANTAR	viii
RIWAYAT HIDUP	ix
DAFTAR ISI	x
DAFTAR GAMBAR	xi
DAFTAR TABEL	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Tujuan Penelitian.....	3
1.4. Manfaat Penelitian	3
1.5. Batasan Masalah	4
1.6. Sistematika Penulisan	4
BAB II TINJAUAN PUSTAKA	6
2.1. Data Mining.....	6
2.1.1. Deskripsi.....	7
2.1.2. Prediksi.....	7
2.1.3. Estimasi	8
2.1.4. Klasifikasi.....	8
2.1.5. Asosiasi.....	10
2.2. Random Forest.....	10
2.3. Support Vector Machine (SVM)	14
2.4. Klasifikasi.....	21
2.5. Penelitian Terdahulu	22
BAB III METODOLOGI PENELITIAN	26
3.1. Alat dan Bahan Penelitian	26

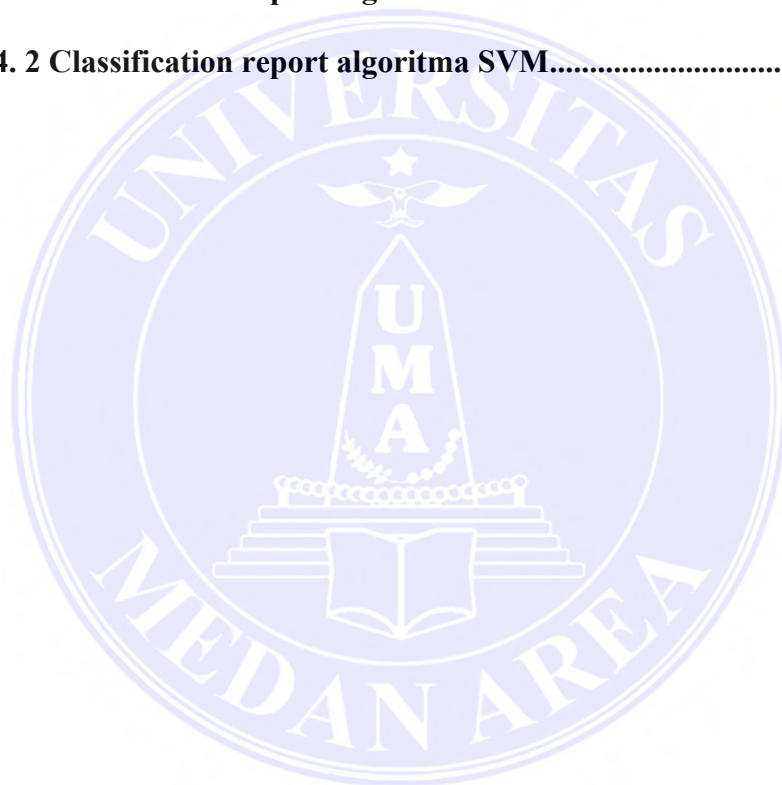
3.1.1. Perangkat Keras.....	26
3.1.2. Perangkat Lunak.....	26
3.2. Tahapan Penelitian	27
3.3. Dataset.....	28
3.4. Data Analis.....	29
3.6. Metode Evaluasi.....	30
3.6.1. <i>Confusion matrix</i>	30
3.6.2. <i>Performa Measure</i>	31
BAB IV HASIL DAN PEMBAHASAN.....	33
4.1. Analisis distribusi data donor darah	33
4.1.1. Analisis data recency.....	33
4.1.2. Analisis Data Frequency	34
4.1.3. Analisis data monetary	35
4.1.4. Analisis data time	36
4.2. Analisis pola donasi darah: frequency, recency, monetary, dan time	37
4.2.1. <i>Analisis frequency dan recency dalam donasi darah</i>	37
4.2.2. <i>Analisis recency dan monetary dalam donasi darah</i>	38
4.2.3. <i>Analisis monetary dan time</i>	39
4.2.4. <i>Analisis frequency dan monetary</i>	40
4.3. Matriks hubungan variabel donasi darah.....	41
4.4. Hasil Klasifikasi Menggunakan Algoritma Random Forest.....	42
4.5. Hasil klasifikasi menggunakan algoritma SVM.....	44
4.6. Pembahasan	45
BAB V KESIMPULAN DAN SARAN	47
5.1. Kesimpulan.....	47
5.2. Saran	47
DAFTAR PUSTAKA	48
LAMPIRAN.....	53

DAFTAR GAMBAR

Gambar 2. 1	Hyperlane Class Negatif dan Positif	20
Gambar 3. 1	Kerangka Kerja Penelitian	27
Gambar 3. 2	Confusion Matrix	30
Gambar 4. 1	Analisis data regency	34
Gambar 4. 2	Analisis data frequency	35
Gambar 4. 3	Analisis data monetary	36
Gambar 4. 4	Analisis data time	37
Gambar 4. 5	Analisis frequency dan recency	38
Gambar 4. 6	Analisis recency dan monetary dalam donasi darah	39
Gambar 4. 7	Analisis monetary dan time	40
Gambar 4. 8	Analisis frequency dan monetary	41
Gambar 4. 9	Matrix hubungan variabel donasi darah	42
Gambar 4. 10	Confusion matrix random forest	43
Gambar 4. 11	Confusion matrix algoritma SVM	44

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu.....	22
Tabel 3. 1 Perangkat Keras yang digunakan.....	26
Tabel 3. 2 Perangkat Lunak yang digunakan	27
Tabel 3. 3 Deskripsi Variabel Dataset	28
Tabel 3. 4 Sampel Data Transfusi Data	29
Tabel 4. 1 Classification report algoritma random forest	43
Tabel 4. 2 Classification report algoritma SVM.....	45



BAB I

PENDAHULUAN

1.1. Latar Belakang

Pengolahan dan analisis data merupakan aspek penting dalam berbagai bidang, termasuk dalam dunia kesehatan. Dalam konteks transfusi darah, manajemen data yang efisien sangatlah vital untuk memastikan keberhasilan dan keamanan prosedur tersebut. Oleh karena itu, penelitian ini akan mengeksplorasi perbandingan antara dua algoritma klasifikasi yang umum digunakan, yaitu Random Forest dan Support Vector Machine (SVM), dalam konteks klasifikasi data transfusi darah. (Ahmad & Fauzi R, 2020).

Dalam konteks sistem informasi kesehatan, klasifikasi data transfusi darah menjadi topik yang penting untuk dikaji. Dengan adanya algoritma klasifikasi yang efektif, proses pengelolaan data dan pengambilan keputusan di dalam sistem transfusi darah dapat ditingkatkan. Oleh karena itu, penelitian ini akan fokus pada mencari perbandingan performa antara algoritma Random Forest dan SVM dalam melakukan klasifikasi data transfusi darah. (Putri & Siti A, 2019).

Fokus utama dari penelitian ini adalah untuk membandingkan akurasi dan performa antara algoritma Random Forest dan SVM dalam melakukan klasifikasi data transfusi darah. Dengan mengetahui algoritma mana yang lebih efektif, dapat membantu meningkatkan efisiensi dan efektivitas sistem transfusi darah, serta memberikan kontribusi terhadap pengembangan sistem informasi kesehatan yang lebih baik. (Hasan & Adi M, 2021).

Berbagai penelitian sebelumnya telah mengkaji tentang penggunaan algoritma klasifikasi dalam berbagai konteks, termasuk dalam klasifikasi data

medis. Random Forest dan SVM merupakan dua algoritma yang sering digunakan dalam berbagai aplikasi klasifikasi, namun belum banyak penelitian yang membandingkan performa keduanya dalam konteks klasifikasi data transfusi darah. (Budi P Siregar, 2020).

Penelitian ini akan menggunakan pendekatan menggunakan data sekunder, yang terdiri dari data transfusi darah yang telah terdokumentasi sebelumnya. Data tersebut akan dibagi menjadi dua bagian: data training dan data testing. Selanjutnya, algoritma Random Forest dan SVM akan diimplementasikan untuk melakukan klasifikasi pada kedua set data tersebut. Performa keduanya akan dievaluasi berdasarkan metrik-metrik evaluasi yang relevan, seperti akurasi, presisi, recall, dan F1-score. (Wijaya & Anton K, 2022).

Hasil analisis akan menunjukkan performa masing-masing algoritma dalam melakukan klasifikasi data transfusi darah. Dengan membandingkan hasil dari kedua algoritma, akan dapat disimpulkan algoritma mana yang lebih efektif dalam konteks ini. Analisis juga akan membahas faktor-faktor yang memengaruhi performa kedua algoritma tersebut. Diskusi akan mengulas implikasi dari hasil penelitian ini dalam konteks pengembangan sistem informasi kesehatan, khususnya dalam manajemen data transfusi darah. Selain itu, akan dibahas juga potensi penerapan hasil penelitian ini dalam penelitian lanjutan atau pengembangan aplikasi praktis di bidang kesehatan.

Penelitian ini memiliki signifikansi penting dalam pengembangan sistem informasi kesehatan, dengan memberikan pemahaman lebih dalam mengenai performa dua algoritma klasifikasi utama, yaitu Random Forest dan SVM, dalam konteks klasifikasi data transfusi darah. Dengan hasil yang diperoleh, diharapkan

dapat memberikan panduan bagi praktisi dan peneliti di bidang kesehatan dalam memilih algoritma yang tepat untuk keperluan klasifikasi data transfusi darah. (Susilo & Bambang T, 2021).

Berdasarkan permasalahan di atas maka peneliti tertarik untuk melakukan penelitian dengan judul “Klasifikasi Data Trasfusi Darah Melalui Pendekatan Algoritma Radom Forest dan Support Vector Machine (SVM)”.

1.2. Rumusan Masalah

Berdasarkan permasalahan di atas yang telah diuraikan, maka rumusuan masalah dalam penelitian yang dilakukan oleh peneliti dapat disimpulkan yakni bagaimana mengklasifikasi data transfusi darah melalui pendekatan algoritma *Random Forest* dan *Support Vector Machine* (SVM).

1.3. Tujuan Penelitian

Adapun tujuan dilakukannya penelitian ini ialah untuk mengetahui penerapan algoritma *Random Forest* dan *Support Vector Machine* dalam mengklasifikasikan data transfusi darah serta meningkatkan pemahaman tentang kebutuhan transfusi darah, mengoptimalkan alokasi sumber daya, atau mengidentifikasi pasien yang berisiko tinggi dan membutuhkan perhatian lebih.

1.4. Manfaat Penelitian

Berdasarkan Tujuan diatas, penelitian ini diharapkan akan membantu meningkatkan pemahaman tentang algoritma serta klasifikasi data Random Forest dan SVM dalam konteks pengolahan dan analisis. Dengan demikian, penelitian ini dapat memberikan wawasan baru dalam bidang ilmu komputer dan teknologi informasi.

1.5. Batasan Masalah

Adapun batasan masalah dalam penelitian adalah sebagai berikut:

1. Proses penelitian ini menggunakan perbandingan algoritma *Random Forest* dan *Support Vector Machine* (SVM).
2. Penelitian dilakukan menggunakan dataset *Blood Transfusion Service Center Data Set*.
3. Data yang diperoleh memberikan hasil yang sebanding
4. Sistem dibangun dengan bahasa pemrograman *python*.
5. Software yang dipakai dalam sistem ini yaitu software *Google Coolab*.

1.6. Sistematika Penulisan

Dalam penulisan skripsi ini penulis menjelaskan tentang Klasifikasi Data Transfusi Darah Melalui Pendekatan Algoritma *Radom Forest* dan *Support Vector Machine* (SVM) yang terbagi menjadi 5 bab di antaranya:

BAB I: PENDAHULUAN

Dalam bab ini akan di uraikan tentang latar belakang dari rumusan-rumusan masalah, tujuan penelitian, manfaat penelitian, dan sistematika penulisan.

BAB II: TINJAUAN PUSTAKA

Pada bab ini akan diuraikan mengenai landasan teori-teori yang digunakan dalam pembahasan penulisan skripsi ini beserta sumbernya.

BAB III: METODE PENELITIAN

Pada bab ini berisi tentang metode penelitian yang akan dilakukan beserta langkah-langkah yang digunakan beserta kaitannya dengan penelitian yang dilakukan.

BAB IV: HASIL DAN PEMBAHASAN

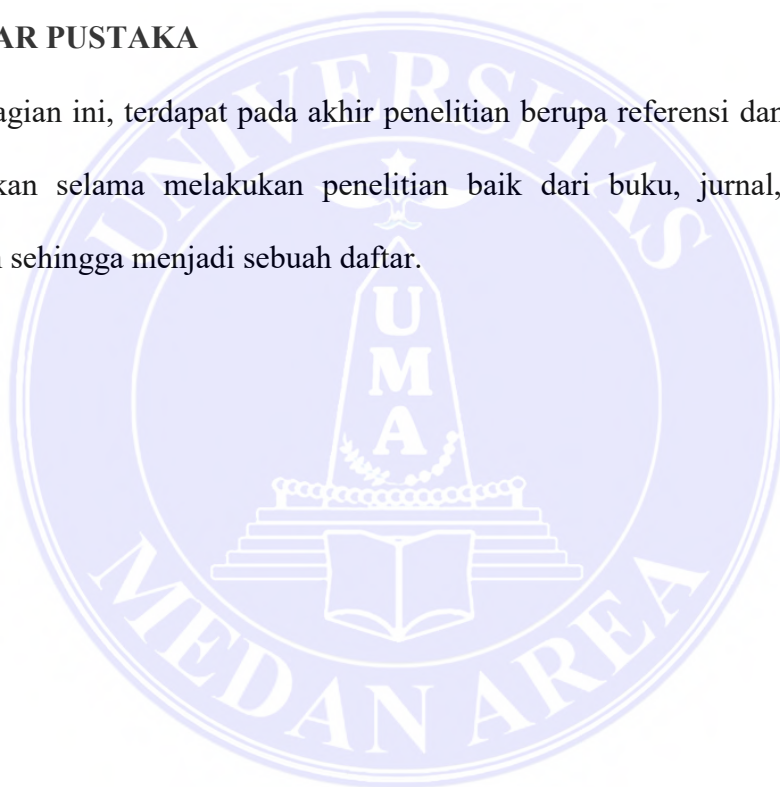
Pada bab ini berisi tentang hasil dan pembahasan dari perancangan program yang meliputi hasil dan implementasi program.

BAB V: KESIMPULAN DAN SARAN

Pada bab ini merupakan akhir dari penulisan skripsi, yang berisikan uraian yang telah dibahas akan dituliskan berupa kesimpulan akhir dari penelitian dan saran untuk penelitian selanjutnya.

DAFTAR PUSTAKA

Pada bagian ini, terdapat pada akhir penelitian berupa referensi dan sumber yang digunakan selama melakukan penelitian baik dari buku, jurnal, artikel yang disusun sehingga menjadi sebuah daftar.



BAB II

TINJAUN PUSTAKA

2.1. Data Mining

Menurut salah satu para ahli megemukakan bahwasanya data mining merupakan salah satu proses pengkajian serta analisis data guna mengidentifikasi pola, hubungan, dan informasi berguna yang tersembunyi di dalam kumpulan data besar. Tujuan utama dari data mining itu sendiri adalah untuk memindahkan data mentah menjadi pengetahuan yang dapat diambil manfaatnya (Osman, A. S., 2019) Metode-metode data mining mencakup eksplorasi data, pemodelan prediktif, pengklasifikasian, dan klasifikasi data. Pada tahap eksplorasi data, teknik seperti pengelompokan (*clustering*) dan pencarian asosiasi (*association*) digunakan untuk mengidentifikasi kelompok- kelompok data yang sama atau hubungan antara atribut yang berbeda.

Dalam teknik data mining yang sudah cukup banyak dikembangkan serta diuji coba dengan menggunakan asumsi kelinearan yang membuat suatu algoritma dapat menghasilkan sesuatu namun terbatas dalam kasus-kasus linear. Sedangkan untuk SVM sendiri, dapat dipergunakan dan dipakai yakni pada data non-linear yaitu dengan memakai pendekatan kernel di fitur data awal yaitu kumpulan data. Adapun tujuan penggunaan kernel ini ialah agar dapat melihat kumpulan data awal, yakni dimana dimensi yang lebih rendah ke dimensi yang relatif lebih tinggi (Prabowo, 2021).

2.1.1. Deskripsi

Deskripsi ini dimaksudkan mampu mengidentifikasi pola-pola berulang di dalam sebuah data serta bertujuan menjadi aturan dan kriteria utama yang mudah dipahami oleh para ahli di bidang aplikasi. Biasanya aturan yang dihasilkan haruslah mudah dipahami agar atinya dapat secara efektif meningkatkan pengetahuan dalam system data minning. Deskripsi data merupakan bagian dari teknik eksplorasi data yang dimana pada umumnya digunakan dalam post-processing sebagai pembuktiann serta dapat menjelaskan hasil penambangan data. Post-processing adalah proses yang digunakan dalam memastikan bahwa hasil yang diperoleh valid dan bermanfaat sehingga nantinya dapat digunakan untuk kepentingan khalayak umum.

2.1.2 Prediksi

Prediksi yang dimaksudkan ialah perkiraan yang bisa dilakukan baik secara kualitatif maupun kuantitatif dalam sebuah penelitian. Menurut Heizer dan Render (2021) mengemukakan bahwa dalam melakukan analisis data pada masa yang akan datang maka dapat dilakukan dengan menggunakan data-data yang ada seperti data historis. Selain itu, bisa juga dilakukan dengan melakukan proses kalkulasi dan model-model yang sudah dirancang oleh pihak manajemen. Data yang diklasifikasikan biasanya memiliki nilai yang diharapkan nantinya akan terjadi di masa yang akan datang.

Dalam data prediksi sendiri lebih cenderung menggunakan teknik kombinasi dari beberapa teknik data minning yang sering dipergunakan dan bisanya akan dipakai dalam kepentingan lainnya seperti ketika menganalisis perkiraan peristiwa yang akan datang. Salah satu contohnya adalah dalam

memprediksi penurunan jumlah pelanggan pada jangka waktu dekat maupun ketika memprediksi harga saham dalam tiga bulan ke depannya.

2.1.3. Estimasi

Estimasi di sini dimaksudkan sebagai suatu variabel yang berperan penting pada nilai-nilai. Pada tahap estimasi ini dilakukan fokus variabel- variabel untuk mencapai target yang lebih menitikberatkan pada nilai numerik dibandingkan dengan kategori. Adapun dalam pembuatan model yang dilakukan yakni dengan menggunakan data yang selengkap-lengkapnya agar nantinya bisa memberikan nilai pada variabel-variabel target dengan tujuan variabel tersebut nantinya akan menjadi nilai prediktif dalam variabel target. Kemudian pada tahap berikutnya evaluasi yang dilakukan yakni menyesuaikan nilai variabel target apakah sudah sesuai dengan variabel prediksi yang sudah ditentukan dalam tahap sebelumnya.

Salah satu contohnya, kita mampu memperkirakan tekanan darah sistolik pada pasien rumah sakit berdasarkan faktor-faktor seperti usia pasien, jenis kelamin, berat badan, serta kadar natrium dalam tubuh manusia. Dalam tekanan darah sistolik dengan nilai variabel prediksi selama proses pembelajaran akan menghasilkan model estimasi yang memiliki hubungan diantara keduanya.

2.1.4. Klasifikasi

Klasifikasi dalam SVM ini melibatkan proses model serta fungsi yang menggambarkan dan juga yang membedakan data ke dalam kelas yang sudah ditentukan. Dalam proses klasifikasi, karakteristik objek diperiksa dan objek ditempatkan pada salah satu kelas yang telah ditentukan. Klasifikasi bisa juga diartikan sebagai suatu pekerjaan yang melihat penilaian terhadap suatu objek data

untuk masuk dalam suatu kelas tertentu dari sejumlah kelas yang telah tersedia (Prasetyo, 2022). Klasifikasi terbagi atas berbagai macam metode diantaranya yaitu, CART (Classification And Regression Trees), CHAID (Chi-Square Automatic Interaction Detection), Random Forest, SVM (Support Vector Machine) dan metode klasifikasi lainnya.

Dalam metode SVM lebih dikenal sebagai metode klasifikasi yang memiliki nilai akurasi yang tinggi. Menurut (Prasetyo, 2022) konsep klasifikasi dengan SVM bisa diasumsikan secara sederhana sebagai suatu usaha mencari hyperplane terbaik dimana berfungsi sebagai pemisah diantara kedua data pada input space. Metode klasifikasi nonparametrik yang juga sering digunakan adalah Classification and Regresssion Trees (CART). Metode Random Forest adalah pengembangan dari metode CART, yaitu dengan menerapkan metode bootstrap aggregating (bagging) dan random feature selection (Breiman, 2021). Sedangkan dalam Random Forest teknik dasar yang biasa dipakai yakni "*Decision Tree*". Biasanya Random Forest memberikan akurasi yang bagus dalam klasifikasi, sehingga mampu menangani data training yang jumlahnya cukup besar serta efektif dalam mengatasi data yang belum lengkap. Akurasi klasifikasi merupakan ukuran ketepatan klasifikasi yang menunjukkan performansi teknik klasifikasi secara keseluruhan (Nugroho, 2023). Ketika akurasi klasifikasinya semakin meningkat maka akan mendapatkan hasil yang semakin baik pula. Pada umumnya ketika mengukur kinerja klasifikasi menggunakan matriks konfusi dimana matriks ini mengandung informasi yang bisa membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

2.1.5. Asosiasi

Dalam data minning ada juga yang disebut sebagai asosiasi atau biasa disebut dengan “*Association Rule*”. Dalam data minning ini, asosiasi merupakan aturan dalam mendukung pengolahan data-data serta mampu mengidentifikasi korelasi. Selain itu, asosiasi ini juga bisa mempertimbangkan bagaimana urutan dalam mengambil materi ataupun data- data yang diperlukan sebagai suatu aturan dalam pengolahan data. Adapun tugas utama domain asosiasi data mining ini yakni berkaitan dengan menemukan atribut yang sering muncul secara bersamaan. Dalam pandangan dunia global dan bisnis, ini lebih sering dikenal sebagai analisis keranjang belanja (*market basket analysis*). Selain itu, tugas asosiasi ialah mengidentifikasi aturan yang bias mengukur hubungan antara dua atau lebih atribut dalam data minning.

2.2. Random Forest

Pada tahun 1995, Tin Kam Ho dari Bell Labs merupakan orang pertama yang memperkenalkan Random Forest. Ia berpendapat bahwasanya, Random Forest merupakan sebuah algoritma pembelajaran ensemble yang menggunakan tahapan data mining yakni mampu mengklasifikasi serta regresi. Random Forest juga ada yang memakai teknik bagging dimana teknik ini dipakai untuk membangun ensemble decision tree. Dalam penelitiannya, Breiman mengungkapkan ada terdapat beberapa kelebihan menggunakan Random Forest, salah satu diantaranya yaitu Random Forest ini biasa memberikan hasil akhir klasifikasi yang baik sehingga tidak ada kemungkinan adanya kerusakan atau permasalahan. Selain itu, Random Forest ini juga mampu mengatasi jumlah data yang cukup banyak dan efektif dalam permasalahan missing data (Breiman, 2021).

Algoritma Random Forest dalam data terdiri dari n amatan dan terdiri atas q variabel independen. Adapun tahapan penyusunan dan pendugaan menggunakan Random Forest dijelaskan dalam beberapa tahapan berikut ini. (Breiman & Cutler, 2023).

1. Pengambilan atau penarikan sampel dilakukan secara acak dengan ukuran n dimana replacement (pengembalian) disesuaikan dengan penarikan sampel (bootstrap).
2. Memilih variabel secara acak dari variabel lainya (q) dimana biasanya ukuran pada variabel (m) merupakan variabel yang paling terbaik diantara variabel lainya. variabel (m) ini dipilih dari aproksimasi dimana akar kuadrat dari total variabel (q) kemudian setengahnya merupakan nilai akar kuadrat dari total jumlah (q) variabel pada (m).
3. Setelah menentukan pemilihan variabel (m) secara acak, maka selanjutnya masuk pada tahap pemangkasan. Lalu lakukan tahap random feature selection yaitu pada setiap node (simpul), selanjutnya dilakukan pemecahan simpul terbaik dalam suatu pohon dengan ukuran Gini Index. Penghitungan Gini Index dilakukan pada m variabel terpilih. Apabila node D dipisah menjadi dua partisi node $D1$ dan $D2$, maka Gini Index dari node yang dipisah dihitung menggunakan persamaan yaitu probabilitas dari kelas l pada node D , $l = 1,2,\dots,k$ dan N adalah jumlah total sampel data pada node D .
4. Langkah berikutnya ialah ulangi langkah 1 dan 2 sebanyak n kali, sehingga membentuk forest yang terdiri atas n - tree (terbentuk pohon sebanyak " n ").

Salah satu algoritma yang bisa kita pakai dan mudah digunakan dalam mengambil keputusan ialah algoritma grid search. Algoritma grid search merupakan algoritma yang paling sering diaplikasikan dalam melakukan hyperparameter optimization karena pengaplikasiannya yang cenderung mudah (Bergstra & Bengio, 2022). Perbedaan antara grid search dan random search yakni pada grid search ketika mencari parameter terbaik algoritma ini mencoba semua kombinasi yang ada sedangkan pada random search hanya mengambil beberapa kombinasi secara acak.

Random Forest ini diciptakan oleh Breiman bertujuan agar membangun beberapa data dan bias berkembang semakin baik lagi. Hasil akhir dari metode ini ialah dapat menghasilkan beberapa keputusan yang diturunkan dari modus masing-masing pohon yang ada. (Nisa, 2020). *Random Forest* (RF) yakni metode yang mampu meningkatkan keakuratan data menggunakan pendekatan acak dalam pembentukan anak simpul pada setiap simpul pada pohon keputusan. Metode ini digunakan untuk membangun pohon keputusan yang terdiri dari simpul akar (*Root node*), simpul internal (*internal node*), dan simpul daun (*leaf node*) dengan memilih atribut dan data secara acak berdasarkan aturan yang telah ditentukan sebelumnya. Simpul akar merupakan simpul teratas dalam pohon dan juga dikenal sebagai akar pohon. Simpul internal merupakan simpul percabangan yang memiliki minimal dua cabang keluar dan hanya satu cabang masuk. Sedangkan simpul daun atau simpul terminal merupakan simpul terakhir yang hanya memiliki satu cabang masuk dan tidak memiliki cabang keluar. Pembentukan pohon keputusan dimulai dengan menghitung nilai entropi untuk menentukan tingkat ketercampuran atribut dan nilai information gain. (Sibirian,

2019). Untuk menghitung nilai *entropy* maka rumus yang digunakan seperti pada persamaan 1, sedangkan nilai *information gain* menggunakan persamaan 2

$$Entropy (Y) = -\sum_i p (c| Y) \log_2 p (c|Y) \dots\dots(2.1)$$

Keterangan :

Y : himpunan kasus

P (c|Y) : proporsi nilai Y terhadap kelas c

$$Information Gain (Y,a) = Entropy (Y) - \sum Entropy (Y_v) \quad (2.2)$$

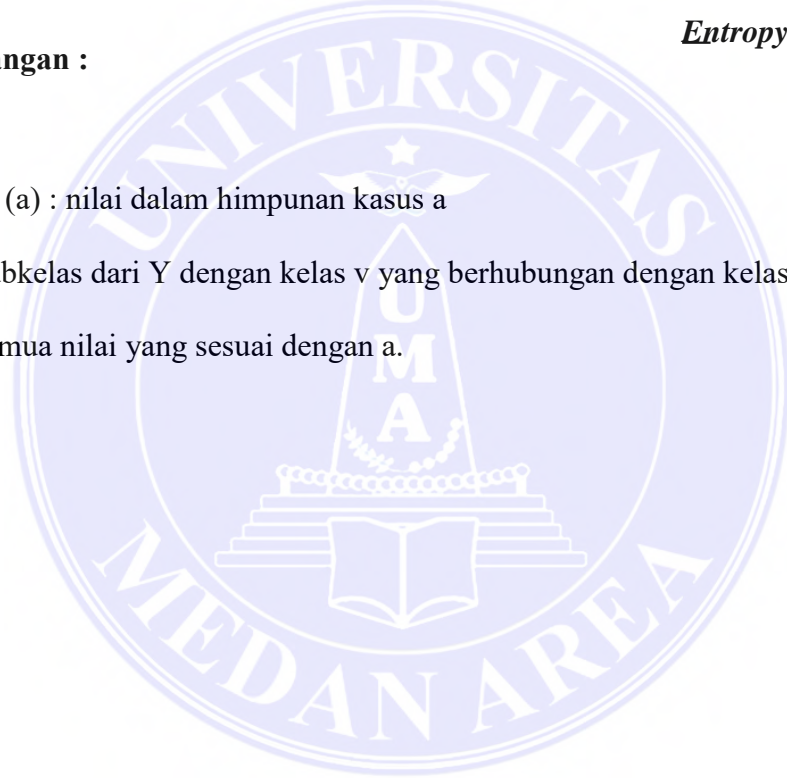
$$Entropy (Y_v) \quad (2.2)$$

Keterangan :

Values (a) : nilai dalam himpunan kasus a

Y_v : subkelas dari Y dengan kelas v yang berhubungan dengan kelas a

Y_a : semua nilai yang sesuai dengan a.



2.3. Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu algoritma hyperplane dimana berfungsi sebagai pemisah data set sebanyak dua dari dua kelas yang tidak sama. Secara umum, SVM memiliki prinsipnya tersendiri baik secara dasar yang berkaitan langsung terhadap linear classifier dimana linear itu sendiri bisa dipisahkan dari linear classifier. Dalam perkembangan SVM saat ini bisa lebih optimal bekerja dan masalah- masalah yang akan dikembanagkan dalam SVM berupa masalah non-linear. Masalah non-linear yang dimaksudkan ini yaitu berupa pengelolaan atau dapat menginput kernel dimana kernel tersebut memiliki dimensi yang cukup tinggi dan besar sehingga pada akhirnya algoritma dapat menemukan hyperlane. Pada hasil akhir hyperlane ini nantinya akan dipakai sebagai pemaksimalan data diantara jarak antar kelasnya.

SVM (*Support Vector Machine*) ada pertama kali pada tahun 1992 dimana diyakini sebagai sebuah metode pembelajaran yang mengalami perkembangan yang sangat begitu pesat. Perkembangan SVM itu sendiri mejadi sebuah penggunaan teknik pembelajaran memakai beberapa hipotesis linier yang memiliki dimensi ruangan yang cukup tinggi (Ichwan, M., & Dewi, I. A. 2019). Tujuan SVM (*support vector machine*) dalam mengidentifikasi *hyperlane* secara terpisah diantara dua kelas berbeda sehingga termaksimalkan dan SVM (*support vector machine*) mengatur jarak data untuk memaksimalkan jarak antara data yang paling dekat dengan *hyperlane* lainnya (Irawan, 2021). SVM (*support vector machine*) juga memakai rumus fungsi matematika dan dikenal sebagai kernel guna mendapatkan sebuah data masukan baru (Firmansyah, 2019) . Adapun tipe kernel yang digunakan dalam penelitian ini ialah dengan menggunakan “*Radial*

$$K(x_i, x_j) = \frac{1}{(2\sigma^2)^{d/2}} \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \dots \dots (2.3)$$

Keterangan :

Di mana x_i dan x_j merupakan pasangan dua data training dan parameter $\sigma > 0$ adalah suatu konstanta. Untuk melakukan substitusi dot product di feature space pada fungsi kernel yang harus digunakan tergantung terhadap data, hal ini dikarenakan fungsi kernel akan menentukan new feature di mana hyperplane akan dicari.

Ketika ingin melakukan suatu substitusi pada dot product pada feature space, dimana memiliki fungsi kernel yang penggunaannya dipakai dengan ketergantungan sebuah data yang akan digunakan. Hal ini dilakukan agar mempermudah dalam mencari hyperlane yang diperlukan. Dalam penelitian ini, peneliti ingin menggunakan kernel *Gaussian Radial Basic Function* dan bagaimana penerapannya dalam mengklasifikasi data transfusi darah. kernel *Gaussian Radial Basic Function* sangat umum dan paling sering digunakan dalam kernel linear dibandingkan dengan kernel yang lainnya.

Konsep kernel *Gaussian Radial Basic Function* (RBF) digunakan untuk mengklasifikasikan data yang tidak dapat dipisahkan secara linear. Dengan penentuan parameter yang tepat, kernel RBF dapat memberikan kinerja yang baik dan hasil pelatihan dengan tingkat kesalahan yang rendah. Ruang fitur kernel RBF memiliki jumlah dimensi yang tak terbatas, yang dapat diatur oleh parameter, sehingga menghasilkan solusi linier yang berpotongan secara unik dan efektif dalam mengklasifikasikan data yang bersifat diskontinu. Oleh karena itu, kernel RBF seringkali menjadi pilihan terbaik untuk tugas klasifikasi yang kompleks

(Pambudi, R. 2022).

Dalam melakukan pemilihan dari fungsi kernel bukanlah hal yang mudah, namun pemilihan kernel ini bisa memberikan pengaruh pada kinerja dari sebuah algoritma terutama dalam algoritma SVM. Pada pengenalan pola tertentu, tak ada langkah lain yang lebih mudah untuk dapat mengetahui sebuah kernel tersebut apakah bisa memberikan akurasi yang tinggi dan baik untuk digunakan dalam sebuah algoritma. Adapun cara yang dilakukan agar mengetahui apakah sebuah kernel tersebut memiliki akurasi yang tinggi yakni membuat sebuah uji coba ataupun eksperimen yang memberikan hasil positif terhadap kernel tersebut. Dalam menggunakan algoritma SVM dapat dilakukan pada tahapan awal, selanjutnya dengan melihat fungsi kernelnya apakah kernel tersebut lebih banyak menghasilkan sebuah akurasi yang tinggi dibandingkan dengan kernel-kernel lainnya.

Algoritma SVM (*support vector machine*) ini adalah metode klasifikasi data transfusi darah dimana dengan melihat dan menentukan data dari batas antara dua kelas dengan jarak maksimal ke jarak terdekat. Dalam penentuan sebuah batas diantara kelas, maka garis pemisah (*hyperlane*) melakukan input space dimana ini berfungsi sebagai penentuan dalam mengukur margin *hyperlane* serta bagaimana mengetahui titik maksimalnya. Margin *hyperlane* disini dimaksudkan sebagai sebuah penanda yakni jarak antara kelas terdekat sebagai algoritma SVM. Algoritma SVM disini bertujuan sebagai pengklasifikasian sebuah data sehingga nantinya dapat dipisahkan secara linear (*linearly separable*) maupun non-linear (*nonlinear separable*). Dengan kata lain, SVM menggunakan teknik *hyperplane* untuk memisahkan data ke dalam kelas-kelas yang sesuai (Rizal, 2019).

SVM memiliki beberapa kelebihan, yaitu konsepnya yang jelas dengan sedikit parameter yang harus diatur, sehingga memberikan model- model klasifikasi namun dengan data yang lebih sedikit, dan dapat diimplementasikan secara mudah karena formula SVM dapat dirumuskan dalam masalah Quadratic Programming. Dengan kata lain, SVM memiliki konsep yang mudah dipahami dan dapat menghasilkan model klasifikasi yang baik dengan sedikit data latih dan pengaturan parameter yang sederhana. Selain itu, formula SVM mudah diimplementasikan karena dapat dirumuskan dalam masalah Quadratic Programming (Hermanto, 2020).

Variabel x_i digunakan untuk mewakili data dalam suatu dataset, sementara variabel y_i digunakan untuk mewakili kelas dalam dataset tersebut. Metode SVM digunakan untuk membuat dua kelas dari pemisahan dataset. Kelas pertama dipisahkan oleh hyperplane dengan nilai 1, sedangkan kelas lainnya memiliki nilai -1. Dengan demikian, kita dapat mengekspresikan hal ini dalam bentuk persamaan berikut ini.

$$x_i \cdot w + b \geq 1 \text{ untuk } Y_i = 1 \dots (2.4)$$

$$x_i \cdot w + b \leq -1 \text{ untuk } Y_i = -1 \dots (2.5)$$

Keterangan:

X_i = data ke- i

W = nilai bobot *support vector* yang tegak lurus dengan

hyperlane b = nilai bias

Y_i = kelas data ke- i

Berdasarkan rumus di atas maka dapat diketahui bahwa dalam support vector bobot (w) yang tegak lurus dengan hyperlane merupakan vektor yang

mempunyai peran sebagai garis tegak lurus diantara kedua titik pusat yaitu titik koordinat dengan hyperplane. Sedangkan, pada nilai bias (b) dimana titik koordinat yang relatif dari garis tersebut terhadap titik koordinat lainnya. Di bawah ini merupakan persamaan yang digunakan untuk menghitung nilai b, sedangkan persamaan lainnya digunakan untuk mencari nilai w.

$$b = - (w \cdot x^+ + w \cdot x^-) \dots\dots (2.6)$$

$$w = \sum_{i=1}^{\infty} y_i x_i \dots\dots\dots (2.7)$$

Keterangan :

b : nilai bias

$w \cdot x^+$: nilai bobot untuk kelas data positif

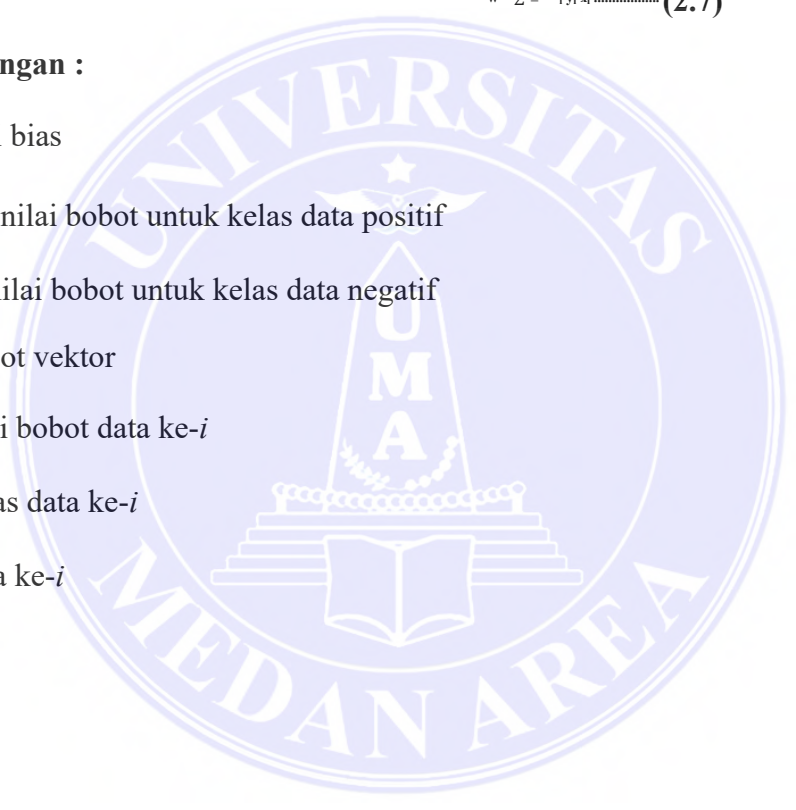
$w \cdot x^-$: nilai bobot untuk kelas data negatif

w : bobot vektor

a_i : nilai bobot data ke-*i*

y_i : kelas data ke-*i*

x_i : data ke-*i*



Berdasarkan rumus di atas maka dapat dilihat jika H_1 merupakan *hyperlane* pendukung dari kelas +1 dimana mempunyai fungsi $wx_1 + b = +1$

$$\text{margin} = |dH_1 - dH_2| = \dots \dots \dots (2.8)$$

Keterangan :

dH_1 = jarak *hyperlane*

pendukung kelas +1 dH_2 = jarak *hyperlane*

pendukung kelas -1

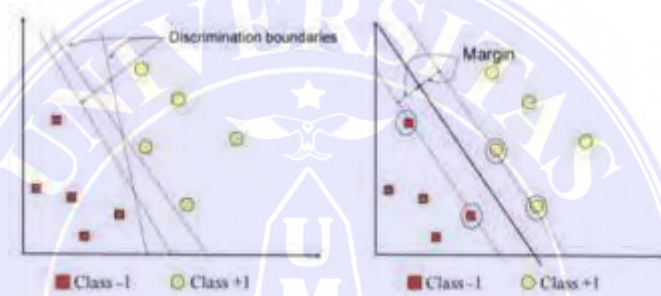
Berdasarkan penjelasan serta pengelolaan dalam algoritma SVM, kita dapat menjelaskannya secara sederhana sebagai upaya mencari garis pemisah terbaik di dalam ruang input. Garis tersebut berfungsi untuk memisahkan dua kelas yang berbeda, yaitu kelas positif yang ditandai dengan +1 dan kelas negatif yang ditandai dengan -1. Algoritma machine learning lainnya yang umum digunakan dan handal adalah Support Vector Machine (SVM). Support Vector Machine (SVM) merupakan salah satu algoritma pembelajaran mesin yang efektif untuk mengakomodasi banyak variabel dan banyak kelas. Ide Mesin vektor pendukung juga kadang-kadang disebut "pengklasifikasi margin besar" karena ketika mendefinisikan batas keputusan antara dua kelas, ia mencoba untuk memaksimalkan margin antara setiap kelas dan batas.

Ide dari Support Vector Machine adalah membuat hyperplane. Tujuannya untuk memisahkan satu set pelatihan sampel l dengan vektor data x_i dan label kelas yang sesuai y_i (x_i, y_i), \dots (x_l, y_l) $\mathbb{R}^n \times \{-1, 1\}$ dengan mencari vektor bobot $w \in \mathbb{R}^n$ dan offset $b \in \mathbb{R}$ dari sedemikian hingga memenuhi persamaan.

$$: \mathbb{R} \rightarrow \{-1, 1\}$$

$$\rightarrow (\cdot +)$$

Berdasarkan rumus diatas dapat dilihat beberapa pola yang merupakan anggota dari kedua kelas tersebut. Pola yang termasuk dalam kelas negatif ditandai dengan kotak, sedangkan dalam masalah klasifikasi, kita mencoba menemukan garis (garis hiper) yang dapat memisahkan kedua kelompok tersebut. Gambar 2.1 di bawah ini juga menunjukkan berbagai alternatif garis pemisah (batas diskriminasi) yang mungkin ada. (Monika & Furqon, 2023)



Gambar 2. 1 Hyperlane Class Negatif dan Positif

Pada Gambar 2.1 (a) di atas menyatakan bahwasaya alteratif pada garis pemisah (*discrimination boundaries*) memperlihatkan bahwa *pattern* yang menjadi satu dalam class-1 dilambangkan pada warna merah (kotak), sedangkan untuk *pattern* dalam class +1 ditandai dengan warna kuning. Sedangkan pada Gambar 2.1 (b) memperlihatkan adanya garis *hyperlane* yang tepat berada diantara kedua kelas. Dalam analisis ini prinsip dasar yang dipakai ialah dimana *hyperlane* yang paling baik dimana dapat meminimalisir kesalahan-kesalahan pada saat klasifikasi data serta mampu memaksimalkan margin geometrisnya seperti tampak pada Gambar 2.1 di atas.

Adapun proses dari SVM itu sendiri ialah pada saat usaha mencari lokasi hyperlanenya. Hyperlanne ini biasanya digunakan sebagai pemisah dua class

dimana dengan mengukur margin hyperlane yang akan digunakan. Setelah itu, maka kita akan menemukan titik maksimal dengan penggunaan margin antara hyperlane dengan pattern terdekatnya masing-masing. Biasanya pattern yang paling dekat ditandai sebagai support vector.

2.4. Klasifikasi

Klasifikasi merupakan sebuah proses dalam menentukan dan memilih model beserta fungsinya dimana dapat menjelaskan serta membedakan konsep yang ada pada kelas data. Tujuan dari klasifikasi ini ialah agar nantinya mampu memperkirakan dan memperhitungkan kelas yang dipakai dalam objek yang belum diketahui (Wibawa, 2019).

Klasifikasi merupakan sebuah teknik dalam *data mining* yang membagikan data menjadi kelompok-kelompok maupun kelas data terbaik yang sudah ditentukan. Klasifikasi yaitu metode *supervised learning* dimana dibutuhkan data-data yang memiliki label ke dalam kelompok maupun kelas yang sudah ditentukan, Setiap teknik tersebut menggunakan algoritma pembelajaran yang bertujuan untuk teridentifikasinya model yang dapat menemukan korelasi yang cocok diantara himpunan atribut maupun label kelas dari data yang diinputkan (Setio, 2020). Adapun akurasi klasifikasi ialah ukuran ketepatan klasifikasi yang menunjukkan performansi teknik klasifikasi secara keseluruhan (Nugroho, 2023). Semakin tinggi akurasi klasifikasi berarti performansi teknik klasifikasi juga semakin baik. Umumnya cara mengukur kinerja klasifikasi menggunakan matriks konfusi. Matriks konfusi mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya.

2.6. Penelitian Terdahulu

Tabel 2. 1 Penelitian Terdahulu

No	Penulis & Tahun	Metode	Keterangan
1	Naufal, M. F. (2021)	<i>SupportVector Machine (SVM)</i>	Dalam penelitian ini peneliti mendapatkan nilai MAPE yaitu sebanyak 10%. Maka dari itu dapat ditarik kesimpulan bahwasanya model akurasi yang dipakai termasuk sangat baik, dimana data yang digunakan dalam menganalisis berapa banyak jumlah wisatawan mancanegara yang datang ke Indonesia dari 6 pintu masuk.
2	(Nugroho, 2023)	<i>Support Vector Machine (SVM)</i>	Membahas tentang klasifikasi pendonor darah menggunakan dataset RFMTC dari

			<p>UCI dataset. Setelah melakukan uji coba, maka diperoleh hasil yang terbaik dimana dapat menghasilkan serta memberikan akurasi tertinggi dalam pengujian SVM adalah $\lambda = 2$, $\gamma = 0.5$, $\varepsilon = 0.005$, dan $C = 20$.</p> <p>Dengan menggunakan parameter tersebut, akurasi terbaik yang diperoleh adalah 72.64%</p>
3	Colanus Ivo Rally Drajana, 2019	<i>Support Vector Machine dan Forward Selection</i>	<p>Pada penelitian ini hasil eksperimen yang diperoleh menunjukkan bahwa series pada pembayaran algoritma SVM serta Forward Selection dimana memberikan kinerja performa yang paling</p>

			baik dibandingkan dengan SVM.
4	(Ramadhan, 2019)	<i>Random Forest</i>	Dalam penelitian pada tahun 2018, dilakukan pemodelan klasifikasi multikelas dengan menggunakan metode random forest untuk mengklasifikasikan mutu pendidikan berdasarkan hasil Ujian Nasional Berbasis Komputer (UNBK) dan skor akreditasi per butir instrumen pada tingkat SMA. Hasil penelitian menunjukkan bahwa model ini mencapai tingkat akurasi keseluruhan sebesar 83,49%.
5	(Tineges, 2020)	<i>Support Vector Machine (SVM)</i>	Dalam penelitian ini, dilakukan analisis pada layanan Indihome. Metode

			<p>yang digunakan adalah <i>Support Vector Machine</i>. Evaluasi dilakukan menggunakan Confusion Matrix dan hasilnya menunjukkan hasilnya yakni sebanyak 87%. Ketepatan antara hasil prediksi dengan data sebenarnya (precision) mencapai 86%, sedangkan tingkat keberhasilan sistem dalam memprediksi sebuah data (recall) mencapai 95%. Adapaun pada tingkat kesalahan keseluruhan data yang di prediksi (error rate) sebanyak 13%. Selain itu, nilai perbandingan rata-rata precision dan recall (F1-score) sudah mencapai hampir 90%.</p>
--	--	--	---

BAB III

METODOLOGI PENELITIAN

3.1. Alat dan Bahan Penelitian

Pada penelitian ini, peneliti memilih menggunakan perangkat yang membantu serta mempermudah peneliti agar nantinya dapat melakukan penelitian dengan lancar. Adapun perangkat keras dan perangkat lunak yang digunakan untuk melakukan penelitian ini yaitu sebagai berikut:

3.1.1. Perangkat Keras

Dalam penelitian ini penulis menggunakan peralatan perangkat keras untuk membantu penelitian ini agar dapat berjalan dengan lancar.

Tabel 3. 1 Perangkat Keras yang digunakan

No	Perangkat keras	Keterangan
1	Device	HP elitebook 820
2	Processor	Intel(R) Core(TM) i7-4600U CPU @ 2.10GHz 2.70 GHz
3	SSD	319 GB
4	RAM	8,00 GB

3.1.2. Perangkat Lunak

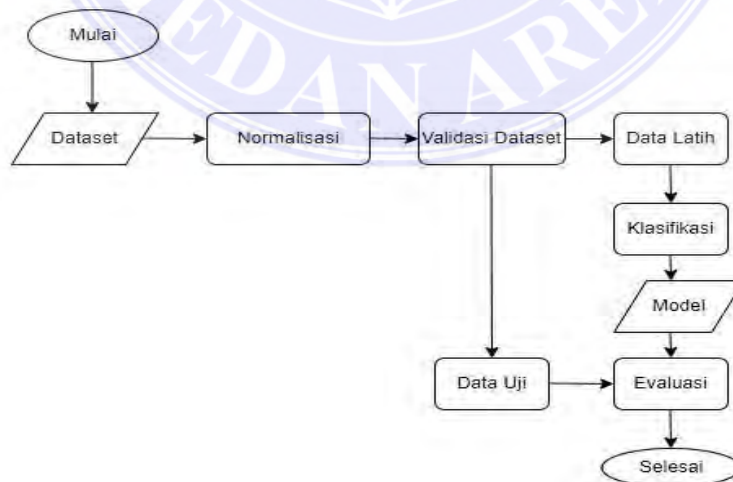
Dalam penelitian ini penulis menggunakan peralatan perangkat lunak untuk membantu penelitian ini agar dapat berjalan dengan lancar.

Tabel 3. 2 Perangkat Lunak yang digunakan

no	Perangkat lunak	Keterangan
1	Sitem Operasi	Windows 10 Pro
2	Google Coolab	RAM: 12 GB Disk : 107 GB
3	RAM	8,00 GB

3.2. Tahapan Penelitian

Studi ini melibatkan beberapa langkah Gambar 3.1. Tahap awal melibatkan pengumpulan dataset, yang diikuti oleh normalisasi data, validasi data, dan tahap klasifikasi. Dua algoritma klasifikasi yang digunakan dalam penelitian ini adalah *Support Vector Machine* dan *Random Forest*. Hasil dari tahap klasifikasi adalah sebuah model klasifikasi, dan evaluasi dilakukan dengan menggunakan akurasi berdasarkan confusion matrix.

**Gambar 3. 1 Kerangka Kerja Penelitian**

3.3. Dataset

Penelitian ini memanfaatkan kumpulan data transfusi darah yang diperoleh dari pusat layanan transfusi darah di kota Hsin-Chu, Taiwan. Data ini telah dipublikasikan di portal *UCI Machine Learning Repository*. Dataset yang digunakan terdiri dari lima atribut utama, yaitu *recency*, *frequency*, *monetary*, *time*, dan *class*. Penting untuk memastikan bahwa kelas dalam dataset seimbang, karena ketidakseimbangan kelas dapat menyebabkan kesalahan yang tidak diinginkan, terutama jika terdapat kesalahan *cost* yang sama secara berulang. Dalam kasus ketidakseimbangan, ukuran kinerja yang dapat digunakan adalah *error rate*, dihitung sebagai 1 minus akurasi. Namun, dalam konteks klasifikasi data dengan ketidakseimbangan kelas, nilai *error rate* tidak selalu mencerminkan performa model dengan baik.

Sebagai alternatif, ukuran performa yang lebih relevan dalam kasus ketidakseimbangan kelas adalah *recall*. *Recall* diukur dengan menggunakan persamaan tertentu dan memberikan informasi tentang kemampuan model untuk mengenali instance dari kelas minoritas. Hal ini penting karena fokus pada kelas minoritas dapat memberikan wawasan lebih baik tentang sejauh mana model dapat mengatasi ketidakseimbangan dan menghindari kesalahan yang berpotensi merugikan pada kelas minoritas. (Chawla, 2022).

Tabel 3. 3 Deskripsi Variabel Dataset

No	Variabel	Deskripsi
1	<i>Recency</i>	berapa bulan sejak terakhir kali mendonor

2	<i>Frequency</i>	berapa kali mendonasikan darah.
3	<i>Monetary</i>	jumlah darah yang didonasikan
4	<i>Time</i>	berapa bulan sejak pertama kali mendonor
5	Kelas	0 berarti dapat berdonasi, 1 tidak dapat berdonasi.

3.4. Data Analisis

Sampel yang digunakan sebagai model uji coba sebanyak 748 data sampel model evaluasi yang digunakan untuk penelitian ini dengan membagikan hasil pengelolaan data menjadi dua kelompok yaitu data training dan data testing yakni ada sebanyak 80% data training dan 20% data testing. Berikut ini merupakan sampel data tranfusi darah yang disajikan dalam tabel 3.4.

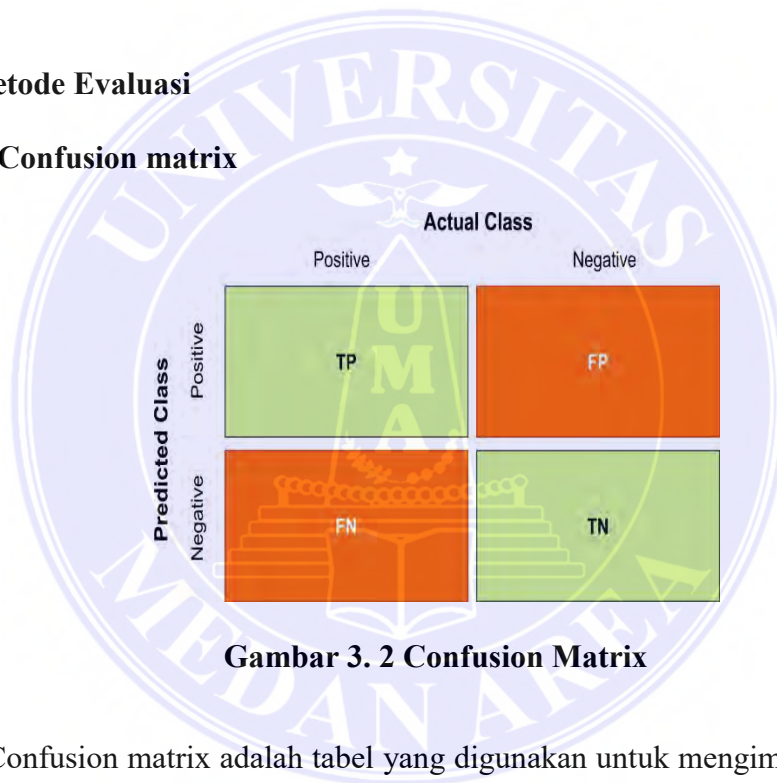
Tabel 3. 4 Sampel Data Transfusi Data

Recency	frequency	monetary	Time	class
2	50	12500	98	1
0	13	3250	28	1
1	16	4000	35	1
2	20	5000	45	1
1	24	6000	77	0
4	4	1000	4	0
2	7	1750	14	1
1	12	3000	35	0

2	9	2250	22	1
5	46	11500	98	1
4	23	5750	58	0
0	3	750	4	0
2	10	2500	28	1
1	13	3250	47	0

3.5. Metode Evaluasi

3.5.1. Confusion matrix



Gambar 3. 2 Confusion Matrix

Confusion matrix adalah tabel yang digunakan untuk mengimplementasikan serta mengklasifikasi pada kumpulan data uji yang nilainya diketahui matriks kebingungan itu sendiri bekerja sederhana dan mudah dimengerti, tetapi bisa membingungkan jika Anda tidak tahu terminologi terkait.

Berdasarkan pada Gambar 3.2 dimana Predicted Class mewakili jumlah instan data yang diketahui berada di grup label adalah benar. Penjelasan dari gambar di atas ialah, pertama True Positives (TP) merupakan jumlah benar dalam memprediksi pengklasifikasian dalam bentuk positif. Kedua, True Negatives

(TN) merupakan jumlah prediksi pengklasifikasian dalam bentuk negatif. Ketiga, False Positives (FP) ialah jumlah prediksi oleh pengklasifikasian menjadi positif dari yang awalnya negatif. Keempat, False Negatives (FN) merupakan jumlah prediksi oleh pengklasifikasian dari yang awalnya negatif menjadi positif.

3.5.2. Performa Measure

Dalam pengklasifikasian ada 4 nilai yang digunakan untuk mengukur kemampuan klasifikasi yang akan di olah yaitu akurasi, presicion, recal,dan fl score.

a. Akurasi

Akurasi Nilai akurasi diperoleh dari jumlah data positif prediksi positif dan data negatif prediksi negatif dibagi dengan jumlah total data dalam dataset (Nasution & Hayaty, 2019). Rumus akurasi adalah sebagai berikut.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad 3.1$$

b. Precision

Merupakan parameter penilaian yang menghitung nilai rata-rata presisi dari data hasil klasifikasi, yaitu jumlah data yang benar di antara nilai sebenarnya (Yudianto, 2020). Rumus Presicion adalah

$$Precision = \frac{TP}{TP + FP} \quad 3.2$$

c. Recall

Recall adalah ukuran kelengkapan model. mengingat persamaan perbandingan antara yang benar-benar positif dengan total sampel yang benar- benar positif (Prabowo, 2021) Rumus Recall adalah

$$Recal = \frac{TP}{TP + FN} \quad 3.3$$

d. F1 Score

Nilai F1-Score, juga dikenal sebagai F-Measure, diperoleh dari hasil Presisi dan Ingat antara kategori yang diprediksi dan kategori aktual (Chicco & Jurman, 2020). Rumus F1 Score adalah

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad 3.4$$

BAB V

KESIMPULAN DAN SARAN

5.1. Kesimpulan

Penelitian ini bertujuan untuk membandingkan efektivitas algoritma random forest dan SVM dalam mengklasifikasikan data donor darah menjadi kategori pendonor reguler dan tidak reguler. Analisis terhadap distribusi data transfusi darah mengungkapkan pola yang signifikan dalam variabel *recency*, *frequency*, *monetary*, dan *time*, yang memiliki implikasi langsung terhadap kinerja klasifikasi. Hasil evaluasi kinerja menunjukkan bahwa random forest secara keseluruhan mendapat akurasi lebih tinggi dibandingkan dengan SVM, dengan random forest menunjukkan keunggulan dalam menyeimbangkan klasifikasi antara pendonor reguler dan tidak reguler.

Algoritma Random Forest terbukti lebih baik dalam menangani kompleksitas dan variasi data transfusi darah, mencapai akurasi tertinggi dibandingkan algoritma SVM. Sebaliknya, SVM menunjukkan keunggulan dalam mengidentifikasi pendonor reguler tetapi mengalami kesulitan dalam klasifikasi pendonor tidak reguler dengan akurasi yang sama tinggi.

5.2. Saran

1. Eksplorasi Teknik Klasifikasi Lain, penggunaan teknik Deep Learning untuk membandingkan atau meningkatkan hasil klasifikasi.
2. Memperkaya Dataset, mendapatkan dataset yang lebih besar dan lebih beragam dari berbagai sumber akan sangat membantu dalam meningkatkan akurasi dan kegeneralan model.

DAFTAR PUSTAKA

- Adi M & Hasan. (2021). Perbandingan akurasi algoritma Support Vector Machine dengan Mobile Neural Network. *Jurnal Teknologi dan Informasi*, 4 (9).
- Anton K & Wijaya. (2022). Klasifikasi menggunakan data sekunder dengan algoritma Support Vector Machine. *Journal Informatic Engginering*, 3 (7).
- Bambang T & Susilo. (2021). Pengembangan sistem pengolahan data medis menggunakan algoritma Support Vector Machine (SVM). *Journal Sistem Informatic*, 5(14).
- Barhoom, A. M. (2019). Blood Donation Prediction Using Artificial Neural Network. *International Journal Of Academic Engineering Research*, 3(10).
- Budi P Siregar. (2020). Klasifikasi data medis menggunakan algoritma Random Forest. *Jurnal Teknik Informatika. Institut Teknologi Bandung*, 7(8).
- Chicco, D., & Jurman, G. (2020). The Advantages Of The Matthews Correlation Coefficient (MCC) Over F1 Score And Accuracy In Binary Classification Evaluation. *BMC Genomics*, 21(1).
- Chicco, D., Tötsch, N., & Jurman, G. (2021). The Matthews Correlation Coefficient (Mcc) Is More Reliable Than Balanced Accuracy, Bookmaker Informedness, And Markedness In Two-Class Confusion Matrix Evaluation. *Biodata Mining*, 14.
- Dwiputri, A., Supadmi, F. R. S., & Mumpuni, N. (2020). Produksi Darah Dan Karakteristik Pendorong Darah Berdasarkan Golongan Darah ABO Dan

Rhesus Di UTD PMI Kota Yogyakarta Tahun 2019. *Universitas Muhammadiyah Yogyakarta Undergraduate Conference, 1*.

Fauzi R & Ahmad. (2020). Analisis data menggunakan Algoritma Random Forest. *Jurnal Indonesia*. 9 (4).

Febriani, A., Rahmawati, T. T., & Sabna, E. (2018). Implementation Of Data Mining To Predict The Feasibility Of Blood Donors Using C4.5 Algorithm. *Indonesian Journal Of Artificial Intelligence And Data Mining, 1*(1).

Firmansyah, S., Gaol, J., Budi Susilo B A Mayor, S., Kelautan, T., Kelautan, M. T., Institut, P., & Bogor, P. (2019). *Journal Of Natural Resources And Environmental Management Perbandingan Klasifikasi SVM Dan Decision Tree Untuk Pemetaan Mangrove Berbasis Objek Menggunakan Citra Satelit Sentinel-2B Di Gili Sulat, Lombok Timur Comparison Of SVM And Decision Tree Classifier With Object Based Approach For Mangrove Mapping To Sentinel-2B Data On Gili Sulat, Lombok Timur*.

Hadi, I., Santoso, L. W., & Tjondrowiguno, A. N. (2020). Sistem Rekomendasi Film Menggunakan User-Based Collaborative Filtering Dan K-Modes Clustering. *Jurnal Infra*, 3(1).

Hendriyana, H., Karo, I. M. K., & Dewi, S. (2022). Analisis perbandingan Algoritma Support Vector Machine, Naive Bayes dan Regresi Logistik untuk Memprediksi Donor Darah. *Jurnal Teknologi Terpadu*, 8(2), 121-126.

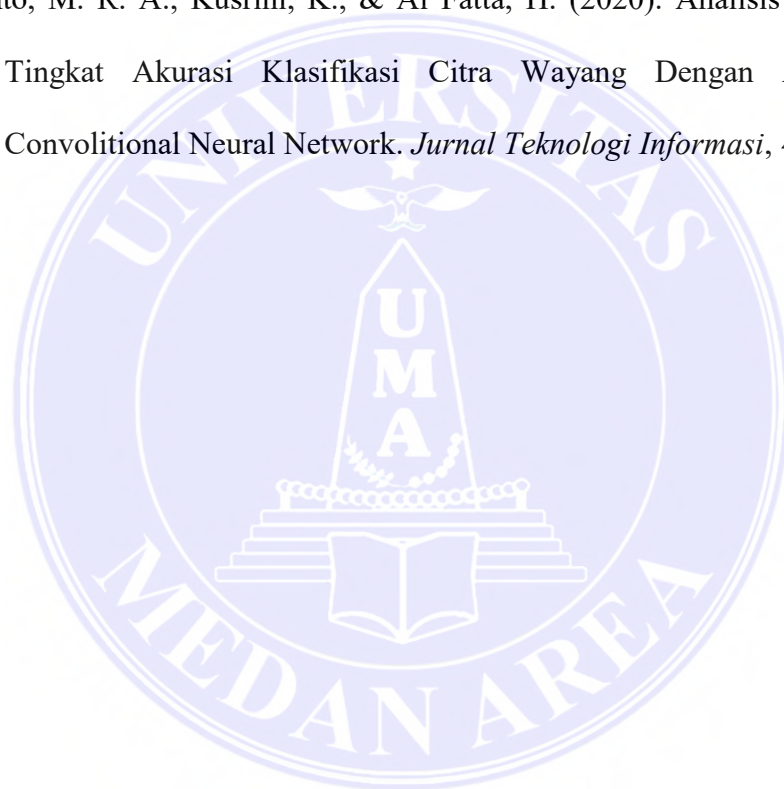
- Hermanto, H., Mustopa, A., & Kuntoro, A. Y. (2020). Algoritma Klasifikasi Naive Bayes Dan Support Vector Machine Dalam Layanan Komplain Mahasiswa. *JITK (Jurnal Ilmu Pengetahuan Dan Teknologi Komputer)*, 5(2).
- Indrayani, R. (2019). Analisa Perbandingan Algoritme Naive Bayes Dan Decision Tree Pada Klasifikasi Data Transfusi Darah. *Jurnal Ilmiah Teknologi Infomasi Terapan*, 5(1).
- Irawan, D., Perkasa, E. B., Yurindra, Y., Wahyuningsih, D., & Helmud, E. (2021). Perbandingan Klassifikasi SMS Berbasis Support Vector Machine, Naive Bayes Classifier, Random Forest Dan Bagging Classifier. *Jurnal Sisfokom (Sistem Informasi Dan Komputer)*, 10(3), 432–437.
- Monika, I. P., & Furqon, M. T. (2021). Penerapan Metode Support Vector Machine (SVM) Pada Klasifikasi Penyimpangan Tumbuh Kembang Anak. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 2(10).
- Nasution, M. R. A., & Hayaty, M. (2019). Perbandingan Akurasi Dan Waktu Proses Algoritma K-NN Dan SVM Dalam Analisis Sentimen Twitter. *Jurnal Informatika*, 6(2).
- Nugroho, E. B., Furqon, M. T., & Hidayat, N. (2023). Klasifikasi Pendoror Darah Menggunakan Metode Support Vector Machine Pada Dataset RFMTC. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIHK)*, 2(10).

- Pambudi, R. (2022). Deteksi Penggunaan Masker dengan Algoritma RBF Support Vector Machine. *The Journal on Machine Learning and Computational Intelligence (JMLCI)*, 1(2).
- Prabowo, T. T. (2021). Efektivitas Sistem Temu Kembali Informasi Perpustakaan Digital Institut Seni Indonesia (ISI) Yogyakarta Dalam Tinjauan Recall Dan Precision. *Media Pustakawan*, 28(1).
- Ramadhan, A., Susetyo, B., & Indahwati. (2019). Penerapan metode klasifikasi random forest dalam mengidentifikasi faktor Penting penilaian mutu pendidikan. *Jurnal Pendidikan Dan Kebudayaan*, 4(2).
- Rizal, R. A., Girsang, I. S., & Prasetyo, S. A. (2019). Klasifikasi Wajah Menggunakan Support Vector Machine (SVM). *REMIK (Riset Dan E-Jurnal Manajemen Informatika Komputer)*, 3(2).
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3.
- Siti A & Putri. (2019). Klasifikasi transfusi darah. *Jurnal Tenaga medis Indonesia*, 2(8).
- Situmorang, P. R., Sihotang, W. Y., & Novitarum, L. (2020). Identifikasi Faktor- Faktor Yang Mempengaruhi Kelayakan Donor Darah Di Stikes Santa Elisabeth Medan Tahun 2019. *Jurnal Analis Medika Biosains (JAMBS)*, 7(2).
- Tineges, R., Triayudi, A., & Sholihati, I. D. (2020). Analisis Sentimen

Terhadap Layanan Indihome Berdasarkan Twitter Dengan Metode Klasifikasi Support Vector Machine (SVM). *Jurnal Media Informatika Budidarma*, 4(3).

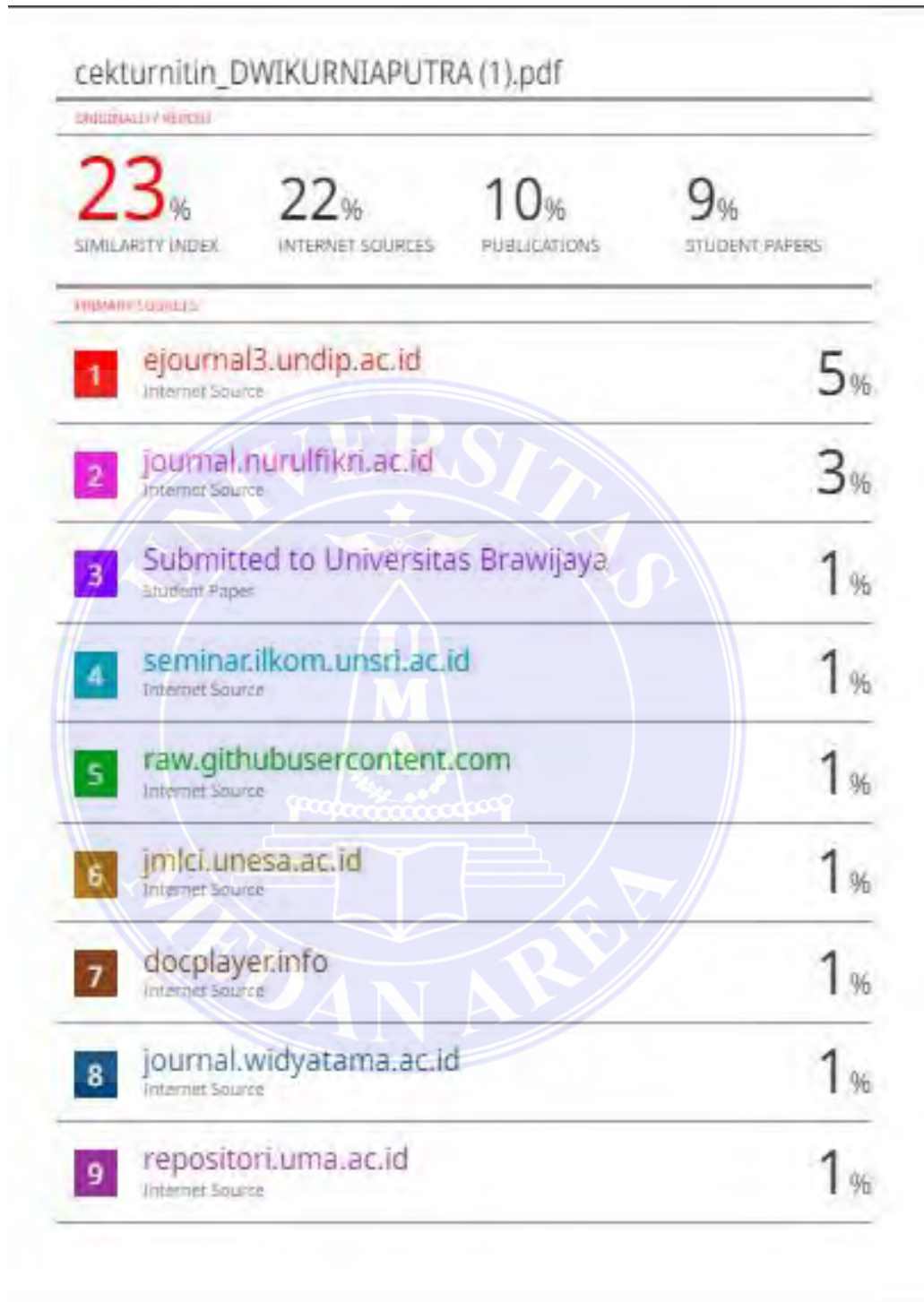
Wibawa, A. P., Guntur, M., Purnama, A., Fathony Akbar, M., & Dwiyanto, F. A. (2022). Metode-Metode Klasifikasi. *Prosiding Seminar Ilmu Komputer Dan Teknologi Informasi*, 3(1).

Yudianto, M. R. A., Kusriani, K., & Al Fatta, H. (2020). Analisis Pengaruh Tingkat Akurasi Klasifikasi Citra Wayang Dengan Algoritma Convolutional Neural Network. *Jurnal Teknologi Informasi*, 4(2).

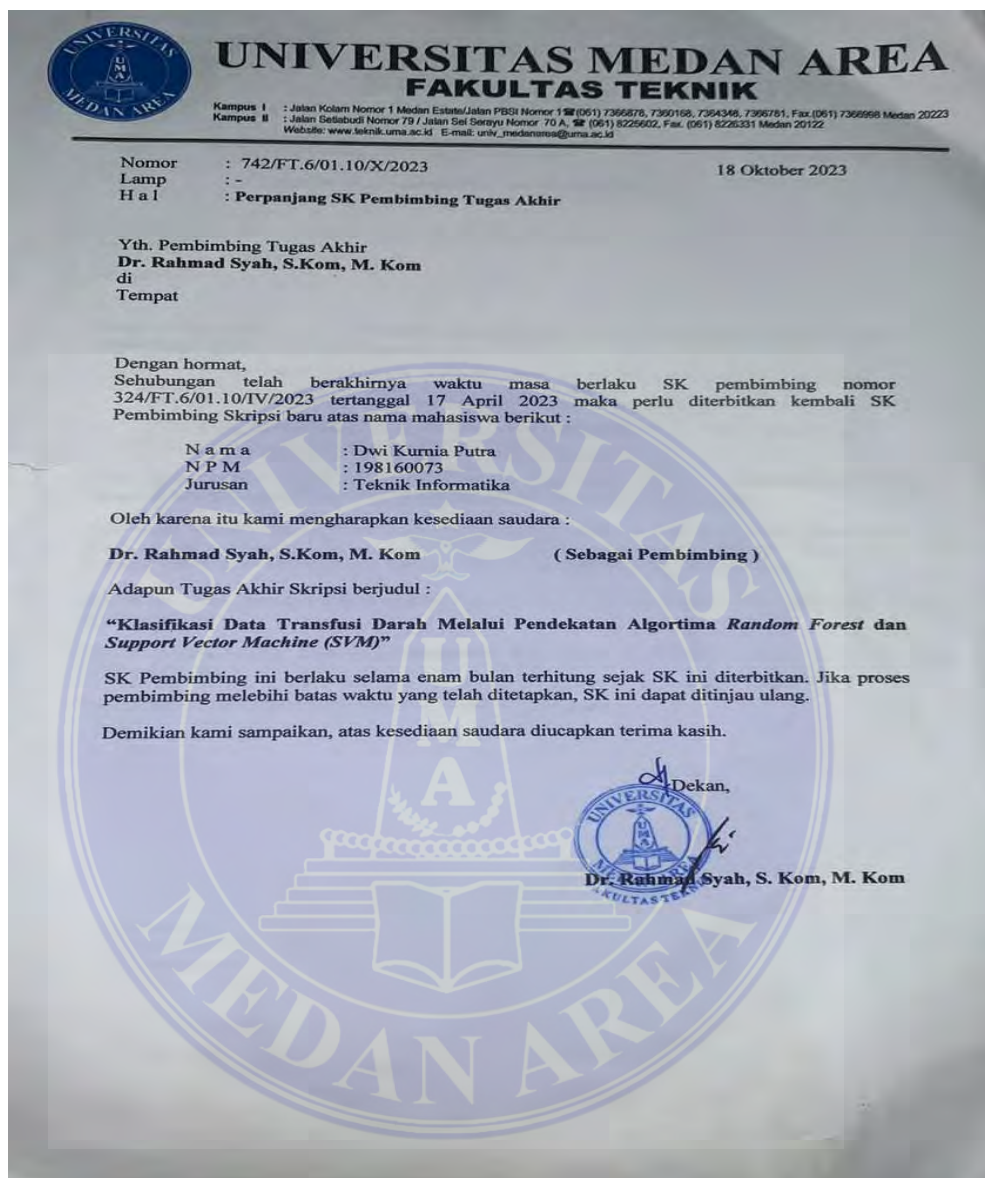


LAMPIRAN

Lampiran hasil plagiasi



Lampiran SK pembimbing tugas akhir



UNIVERSITAS MEDAN AREA
FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estabul/Jalan PBSI Nomor 1 ☎ (061) 7366876, 7360188, 7364348, 7366761, Fax (061) 7366968 Medan 20223
Kampus II : Jalan Setiabudi Nomor 79 / Jalan Sei Sanyin Nomor 70 A, ☎ (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanarea@uma.ac.id

Nomor : 742/FT.6/01.10/X/2023
Lamp : -
Hal : Perpanjang SK Pembimbing Tugas Akhir

18 Oktober 2023

Yth. Pembimbing Tugas Akhir
Dr. Rahmad Syah, S.Kom, M. Kom
di
Tempat

Dengan hormat,
Sehubungan telah berakhirnya waktu masa berlaku SK pembimbing nomor 324/FT.6/01.10/IV/2023 tertanggal 17 April 2023 maka perlu diterbitkan kembali SK Pembimbing Skripsi baru atas nama mahasiswa berikut :

N a m a : Dwi Kurnia Putra
N P M : 198160073
Jurusan : Teknik Informatika

Oleh karena itu kami mengharapkan kesediaan saudara :

Dr. Rahmad Syah, S.Kom, M. Kom (Sebagai Pembimbing)

Adapun Tugas Akhir Skripsi berjudul :


"Klasifikasi Data Transfusi Darah Melalui Pendekatan Algoritma Random Forest dan Support Vector Machine (SVM)"

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. Rahmad Syah, S. Kom, M. Kom

Lampiran surat penelitian dan pengambilan data tugas akhir

**UNIVERSITAS MEDAN AREA**
FAKULTAS TEKNIK

Kampus I : Jalan Kolam Nomor 1 Medan Estate/Jalan PPSI Nomor 1 Medan (061) 7366878, 7360158, 7364348, 7366781, Fax.(061) 7366999 Medan 20223
Kampus II : Jalan Selsebukil Nomor 79 / Jalan Sei Serayu Nomor 70 A, Kf. (061) 8225602, Fax. (061) 8226331 Medan 20122
Website: www.teknik.uma.ac.id E-mail: univ_medanama@uma.ac.id

Nomor : 725 /FT.6/01.10/X/2023 16 Oktober 2023
Lamp : -
Hal : Penelitian Dan Pengambilan Data Tugas Akhir

Yth. Wakil Rektor Bid. Pengembangan SDM & Adm. Keuangan
Jln. Kolam No.1
Di
Medan

Dengan hormat, kami mohon kesediaan ibu kiranya berkenan untuk memberikan izin dan kesempatan kepada mahasiswa kami tersebut dibawah ini :

NO	N A M A	N P M	PRODI
1.	Dwi Kurnia Putra	198160073	Teknik Informatika

Untuk melaksanakan Penelitian dan Pengambilan Data Tugas Akhir di **Laboratorium Komputer Program Studi Teknik Informatika Fakultas Teknik Universitas Medan Area**.

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 :

Klasifikasi Data Transfusi Darah melalui Pendekatan Algoritma Random Forest dan Support Vector Machine (SVM).

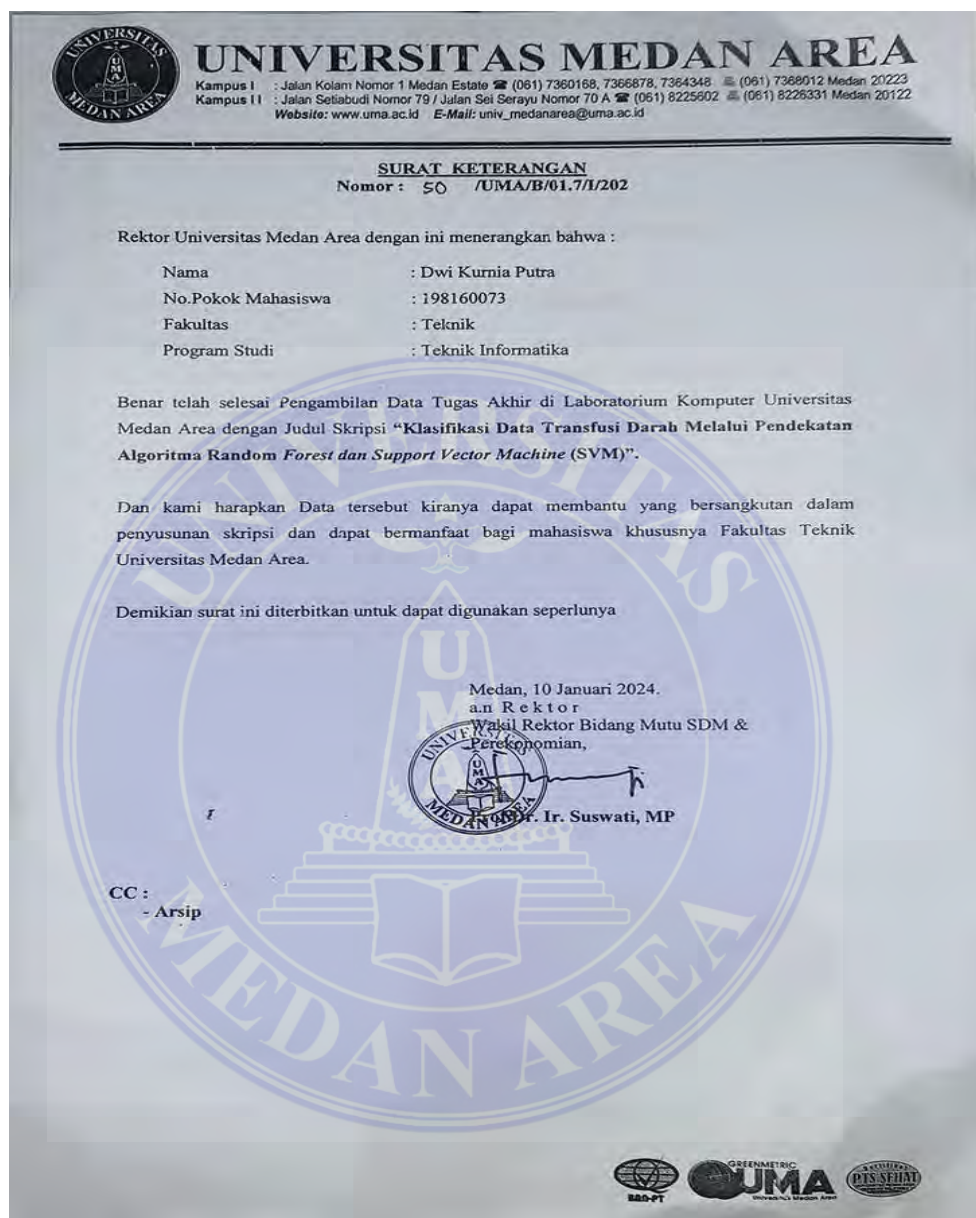
Mohon kiranya tanggal Surat Izin Pengambilan Data Tugas Akhir agar disesuaikan dengan tanggal Terbitnya SK ini.

Atas perhatian dan kerja sama yang baik diucapkan terima kasih.


Dr. Rahmad Syah, S.Kom, M. Kom

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

Lampiran surat keterangan selesai penelitian



Lampiran Source Code Google Colaboratory

Import library yang dibutuhkan

```
import numpy as np
```

```
import pandas as pd
```

```
from google.colab import files
```

Mengunggah file dari perangkat lokal Anda ke Google Colab

```
uploaded = files.upload()
```

Mengambil nama file yang diunggah (asumsi satu file diunggah)

```
file_name = list(uploaded.keys())[0]
```

Membaca file CSV ke dalam DataFrame

```
df = pd.read_csv(file_name)
```

Menampilkan beberapa baris pertama dari DataFrame

```
print(df.head())
```

variasi dan karakteristik data di setiap kolom

```
for i in df.columns:
```

```
    dis = len(df[i].unique())
```

```
    print(f"{i} - {dis}")
```

#Mengimpor pustaka dan membuat data sampel untuk dua distribusi

```
import matplotlib.pyplot as plt
```

```
import seaborn as sns
```

Memplot histogram dari kedua distribusi pada subplot

```
fig, axes = plt.subplots(nrows=2, ncols=2)
```

```
axes = axes.flatten()
```

```
fig.set_size_inches(20, 20)

for ax, col in zip(axes, df.columns):

    sns.distplot(df[col], ax=ax)

plt.show()

# Membuat scatter plot

sns.scatterplot(x='Recency', y='Frequency', hue='class', data=df,palette='deep')

# Membuat matriks korelasi

correlation_matrix = df.corr()

# Menentukan ukuran gambar heatmap

plt.figure(figsize=(15, 10))

# Membuat heatmap dari matriks korelasi

sns.heatmap(correlation_matrix, annot=True, cmap='Blues', linewidths=0.5,

fmt='.2f')

# Menambahkan judul

plt.title('Correlation Matrix Heatmap')

# Menampilkan gambar

plt.show()

# Klasifikasi Menggunakan Algoritma Random Forest

from sklearn.model_selection import train_test_split

from sklearn.ensemble import RandomForestClassifier

from sklearn.metrics import confusion_matrix, classification_report

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt
```

```
# Replace 'path_to_your_file.csv' with the path to your CSV
file data = pd.read_csv('Blood Transfusion Service data.csv')

# Correcting DataFrame name and column name

X = data[['Recency', 'Frequency', 'Monetary', 'Time']]
y = data['Class(Target)']

# Splitting the data into training and testing sets (80/20)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=42)

# Creating the Random Forest Classifier

rf_classifier = RandomForestClassifier(random_state=42)

# Training the model

rf_classifier.fit(X_train, y_train)

# Predicting the test set results

y_pred = rf_classifier.predict(X_test)

# Computing the confusion matrix

conf_matrix = confusion_matrix(y_test, y_pred)

conf_matrix_report = classification_report(y_test, y_pred)

# Displaying the confusion matrix as a heatmap with custom labels, size, and
colormap

plt.figure(figsize=(8, 6))

sns.heatmap(conf_matrix, annot=True, fmt='g', cmap='coolwarm',
            xticklabels=['Pendonor Reguler', 'Pendonor Tidak Reguler'],
            yticklabels=['Pendonor Reguler', 'Pendonor Tidak Reguler'])

plt.xlabel('Predicted')
```

```
plt.ylabel('Actual')

plt.title('Confusion Matrix')

plt.show()

# Printing the classification report

print("Classification Report:")

print(conf_matrix_report)

# Klasifikasi Menggunakan Algoritma SVM

from sklearn.model_selection import train_test_split

from sklearn.svm import SVC

from sklearn.metrics import confusion_matrix, classification_report

import pandas as pd

import seaborn as sns

import matplotlib.pyplot as plt

# Replace 'path_to_your_file.csv' with the path to your CSV

file_data = pd.read_csv('Blood Transfusion Service data.csv')

# Correcting DataFrame name and column name

X = data[['Recency', 'Frequency', 'Monetary', 'Time']]

y = data['Class(Target)']

# Splitting the data into training and testing sets (80/20)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2,

random_state=42)

# Creating the Support Vector Machine Classifier (SVM)

svm_classifier = SVC(random_state=42)
```

```
# Training the model

svm_classifier.fit(X_train, y_train)

# Predicting the test set results

y_pred = svm_classifier.predict(X_test)

# Computing the confusion matrix

conf_matrix = confusion_matrix(y_test, y_pred)

conf_matrix_report = classification_report(y_test, y_pred)

# Displaying the confusion matrix as a heatmap with custom labels, size, and
colormap

plt.figure(figsize=(8, 6))

sns.heatmap(conf_matrix, annot=True, fmt='g', cmap='coolwarm',
            xticklabels=['Pendonor Reguler', 'Pendonor Tidak Reguler'],
            yticklabels=['Pendonor Reguler', 'Pendonor Tidak Reguler'])

plt.xlabel('Predicted')

plt.ylabel('Actual')

plt.title('Confusion Matrix')

plt.show()

# Printing the classification report

print("Classification Report:")

print(conf_matrix_report)
```