

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Objek penelitian dalam studi ini difokuskan pada pengembangan model prediksi stok permintaan darah berbasis algoritma *Random Forest*, *Gradient Boosting*, dan *hybrid Random Forest-Gradient Boosting* pada Palang Merah Indonesia (PMI) Kabupaten Tangerang. Penelitian ini bertujuan untuk memberikan pemahaman yang mendalam terhadap pola permintaan darah dengan mempertimbangkan berbagai aspek spesifik. Data yang digunakan dalam penelitian ini mencakup periode dari tahun 2021 hingga 2024 sehingga memungkinkan pengamatan terhadap perubahan pola permintaan stok darah dari waktu ke waktu. Untuk memberikan gambaran awal mengenai karakteristik data penelitian, berikut disajikan Gambar 3.1 mengenai visualisasi tren permintaan darah secara tahunan yang menunjukkan pola dan fluktuasi permintaan darah di PMI Kabupaten Tangerang selama periode penelitian.



Gambar 3.1 Visualisasi Tren Permintaan Darah Tahunan (2021-2024)

Berdasarkan visualisasi di atas, terlihat bahwa terdapat tren peningkatan permintaan darah dari tahun 2021 hingga 2024, dengan puncak permintaan terjadi pada tahun 2024. Pola ini mengindikasikan adanya faktor-faktor tertentu yang mempengaruhi dinamika permintaan darah, sehingga diperlukan model prediksi yang mampu mengakomodasi variabilitas tersebut. Untuk mengembangkan model prediksi yang akurat terhadap pola permintaan darah seperti yang terlihat pada

visualisasi di atas, diperlukan pemahaman terhadap berbagai variabel yang mempengaruhi permintaan tersebut. Penelitian ini menggunakan beberapa atribut kunci yang mencakup jenis darah, dimensi waktu, dan volume permintaan. Adapun *field* yang dipakai dalam penelitian tertera pada Tabel 3.1 berikut.

Tabel 3.1 Keterangan Atribut Penelitian

Field	Tipe Data	Pengertian
Golongan Darah	String	Klasifikasi darah berdasarkan antigen atau antibodi (A, B, O, AB)
Jenis Darah	String	Jenis dari komponen-komponen dalam darah
• AHF (Anti-Hemophilic Factor)	String	Komponen darah yang mengandung faktor pembekuan.
• FFP (Fresh Frozen Plasma)	String	Plasma darah yang dibekukan dalam waktu tertentu setelah donor.
• Leucodepleted	String	Komponen darah yang telah melalui proses pengurangan leukosit.
• Leucodepleted-Pediatrik	String	Komponen darah yang sudah melalui proses pengurangan leukosit dan dikemas dalam volume lebih kecil untuk pasien anak-anak.
• PRC (Packed Red Cells)	String	Sel darah merah pekat yang telah dipisahkan plasma nya.
• PRC-Pediatrik	String	Sel darah merah pekat yang dikemas dalam volume lebih kecil untuk transfusi pada bayi dan anak-anak.
• TC (Thrombocyte Concentrate)	String	Konsentrat trombosit yang digunakan untuk pasien dengan gangguan pembekuan darah.
• TC - Apheresis	String	Trombosit yang diperoleh melalui prosedur analisis, di mana hanya komponen trombosit yang diambil

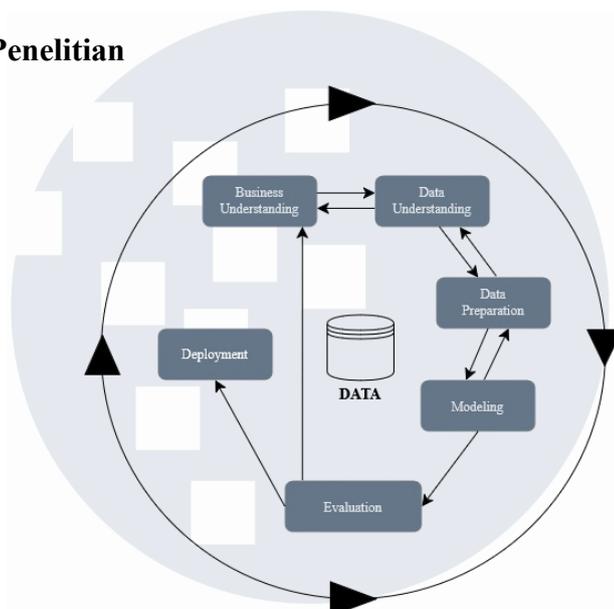
Field	Tipe Data	Pengertian
		dari donor, memiliki kualitas lebih baik dibandingkan TC biasa.
• WB (Whole Blood)	String	Darah utuh yang mengandung semua komponen (sel darah merah, sel darah putih, trombosit, dan plasma), digunakan dalam situasi darurat atau kehilangan darah yang signifikan.
• Plasma-Konvalesen	String	Plasma darah dari penyintas penyakit tertentu yang mengandung antibodi.
• PRC-Pediatrik-50cc, PRC-Pediatrik-75cc, PRC-Pediatrik-100cc	String	Sel darah merah pekat yang dikemas dalam ukuran tertentu (50cc, 75cc, atau 100cc) untuk disesuaikan dengan kebutuhan transfusi pasien anak-anak atau bayi.
• FFP-NAT (Fresh Frozen Plasma - Nucleic Acid Testing)	String	Plasma beku segar yang telah melalui uji NAT untuk mendeteksi virus seperti HIV, Hepatitis B, dan Hepatitis C
• PRC-NAT (Packed Red Cells - Nucleic Acid Testing)	String	Sel darah merah pekat yang telah diuji menggunakan metode NAT untuk memastikan keamanan dari infeksi virus yang ditularkan melalui darah.
Bulan	String	Bulan permintaan darah berlangsung
Tahun	Int	Tahun permintaan darah berlangsung
Jumlah	Int	Total permintaan darah (kantong)

Data tersebut akan melalui tahapan pengolahan dan analisis untuk menghasilkan model prediksi yang akurat. Model yang dihasilkan diharapkan mampu mengintegrasikan keunggulan dari masing-masing algoritma untuk meningkatkan akurasi prediksi. Selain itu, hasil prediksi akan divisualisasikan

dalam bentuk web *dashboard* menggunakan Streamlit. Visualisasi ini dirancang untuk memudahkan pengguna dalam memahami pola dan tren permintaan darah serta mendukung pengambilan keputusan yang lebih baik di tingkat operasional PMI.

3.2 Metode Penelitian

3.2.1 Alur Penelitian



Gambar 3.2 Diagram Alur Penelitian

Pada alur penelitian ini, framework yang digunakan adalah CRISP-DM (*Cross Industry Standard Process for Data Mining*). Gambar 3.1 merupakan alur yang diimplementasikan pada penelitian ini.

3.2.2 Metode Data Mining

Metode *data mining* dalam penelitian ini menggunakan CRISP-DM. Adapun pilihan *framework* selain CRISP-DM ialah KDD (*Knowledge Discovery in Databases*) dan SEMMA (*Sample, Explore, Modify, Model, Assess*). Untuk perbandingan *framework* CRISP-DM, KDD, dan SEMMA dapat dilihat pada Tabel 3.2 berikut.

Tabel 3.2 Perbandingan Framework CRISP-DM, KDD, dan SEMMA

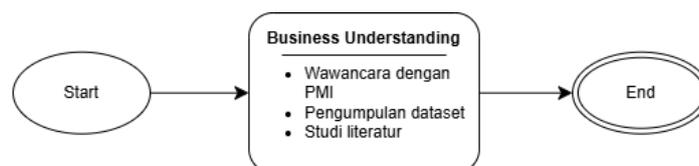
Aspek	CRISP-DM	KDD	SEMMA
Fokus Utama	Proses data mining yang iteratif dan aplikatif dalam berbagai industri.	Menemukan pola tersembunyi dalam data besar.	Fokus pada pengolahan statistik dan model eksploratif.
Tahapan	<ol style="list-style-type: none"> 1. <i>Business Understanding</i> 2. <i>Data Understanding</i> 3. <i>Data Preparation</i> 4. <i>Modeling</i> 5. <i>Evaluation</i> 6. <i>Deployment</i> 	<ol style="list-style-type: none"> 1. <i>Data Selection</i> 2. <i>Data Preprocessing</i> 3. <i>Data Transformation</i> 4. <i>Data Mining</i> 5. <i>Interpretation/Evaluation</i> 	<ol style="list-style-type: none"> 1. <i>Sample</i> 2. <i>Explore</i> 3. <i>Modify</i> 4. <i>Model</i> 5. <i>Assess</i>
Keunggulan	<ul style="list-style-type: none"> • Fleksibel dan iteratif, cocok untuk berbagai industri. • Menekankan pemahaman bisnis sebelum analisis data. • Memiliki tahap <i>deployment</i> yang memungkinkan implementasi model secara 	<ul style="list-style-type: none"> • Berfokus pada penemuan pola dalam data skala besar. • Cocok untuk eksplorasi data berbasis <i>Machine Learning</i>. 	<ul style="list-style-type: none"> • Efektif dalam eksperimen model statistik dan eksplorasi data. • Mudah digunakan untuk analisis eksploratif dalam lingkungan SAS.

Aspek	CRISP-DM	KDD	SEMMA
	nyata.		
Kelemahan	<ul style="list-style-type: none"> • Bisa lebih kompleks dibandingkan pendekatan lain. • Membutuhkan lebih banyak iterasi untuk mendapatkan model optimal. 	<ul style="list-style-type: none"> • Tidak memiliki struktur standar untuk deployment model. • Cenderung berfokus pada eksplorasi data tanpa tahap bisnis. 	<ul style="list-style-type: none"> • Tidak mencakup pemahaman bisnis dan tahap <i>deployment</i>. • Lebih cocok untuk eksplorasi statistik daripada implementasi model di industri.
Waktu Penggunaan	<ul style="list-style-type: none"> • Jika ingin menerapkan <i>data mining</i> untuk pengambilan keputusan bisnis. • Jika memerlukan pendekatan iteratif dengan evaluasi 	<ul style="list-style-type: none"> • Jika tujuan utama adalah menemukan pola tersembunyi dalam data besar. • Jika tidak memerlukan implementasi bisnis langsung. 	<ul style="list-style-type: none"> • Jika ingin eksplorasi data statistik sebelum membangun model prediktif. • Jika menggunakan SAS sebagai platform utama.

Aspek	CRISP-DM	KDD	SEMMA
	berulang.		

CRISP-DM dipilih sebagai *framework* utama dalam penelitian ini karena menyediakan kerangka kerja sampai tahap *deployment* dan iteratif, yang tidak hanya fokus pada aspek teknis pemodelan data tetapi juga pada pemahaman konteks bisnis secara mendalam [36]–[40]. Berbeda dengan KDD yang langsung fokus pada data selection tanpa tahap *business understanding*, CRISP-DM memulai proses dengan pemahaman kebutuhan *stakeholder* PMI, *workflow* operasional, dan dampak keputusan prediksi terhadap keselamatan pasien yang krusial dalam domain kesehatan. Selain itu, CRISP-DM memiliki tahap *deployment* yang terstruktur untuk implementasi model dalam lingkungan operasional nyata, sedangkan KDD dan SEMMA tidak menyediakan panduan *deployment* yang komprehensif padahal implementasi *dashboard* interaktif merupakan tujuan akhir penelitian ini. Sifat iteratif CRISP-DM juga memungkinkan evaluasi berulang dan perbaikan model berdasarkan *feedback* operasional, sehingga model dapat beradaptasi dengan perubahan pola permintaan darah akibat faktor musiman, epidemiologi, atau demografi di Kabupaten Tangerang. Dengan orientasi praktis untuk implementasi industri dan fleksibilitas lintas sektor, CRISP-DM merupakan pilihan metodologi yang tepat untuk mengembangkan solusi prediksi yang andal, *implementable*, dan *sustainable* dalam mendukung pengambilan keputusan manajemen stok darah di PMI. Berikut adalah 5 tahapan CRISP-DM yang akan diterapkan pada penelitian ini:

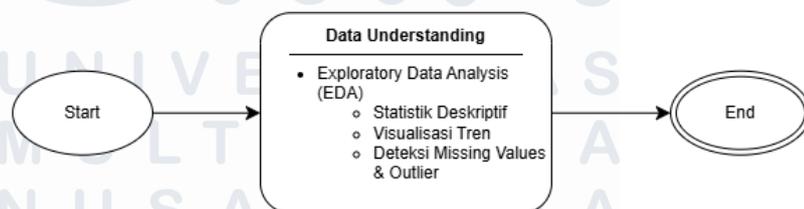
1. *Business Understanding*



Gambar 3.3 Diagram Business Understanding

Gambar 3.3 menunjukkan skema *Business Understanding* dalam metodologi CRISP-DM bertujuan untuk memahami secara mendalam konteks bisnis, kebutuhan, serta tujuan analisis sebelum masuk ke tahap eksplorasi dan pemodelan data. Dalam penelitian ini, langkah pertama yang dilakukan adalah wawancara dengan *stakeholder*, yaitu pihak UTD PMI Kabupaten Tangerang, untuk mengetahui bagaimana pencatatan dan pengelolaan stok darah dilakukan serta kendala yang dihadapi, terutama dalam perencanaan pemenuhan permintaan darah. Selanjutnya, dilakukan studi literatur untuk menganalisis penelitian terdahulu mengenai metode prediksi permintaan darah dan sistem manajemen stok darah guna menemukan pendekatan yang paling sesuai. Selain itu, penelitian ini juga melakukan analisis data historis dengan meninjau data permintaan dan persediaan darah yang telah dicatat dalam aplikasi Sidoni, termasuk pola tren berdasarkan golongan darah, jenis darah, serta faktor musiman yang berpengaruh. Data yang dianalisis merupakan data permintaan darah di PMI Kabupaten Tangerang yang diperoleh melalui observasi langsung, mencakup periode 2021-2024 dengan atribut utama seperti Jenis Darah, Golongan Darah, dan Jumlah Permintaan Darah berdasarkan Bulan dan Tahun.

2. *Data Understanding*

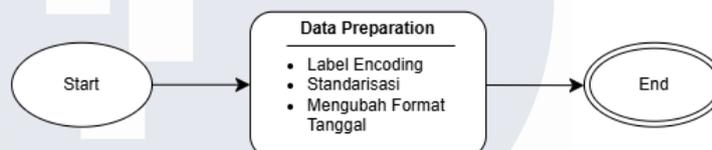


Gambar 3.4 Diagram *Data Understanding*

Tahap *Data Understanding* dalam metodologi CRISP-DM berfokus pada pemahaman mendalam terhadap data yang digunakan dalam penelitian ini. *Flow* tahapan ini secara sederhana dapat dilihat pada Gambar 3.4. Proses ini mencakup beberapa langkah penting, salah

satunya adalah *Exploratory Data Analysis* (EDA). EDA dilakukan untuk memahami distribusi data, mengidentifikasi pola atau tren, serta mendeteksi anomali atau *missing values* yang dapat memengaruhi kualitas analisis. Teknik yang digunakan dalam EDA meliputi statistik deskriptif, seperti *mean*, *median*, dan standar deviasi untuk melihat kecenderungan data, serta visualisasi data guna memahami pola permintaan darah dari waktu ke waktu. Selain itu, tahap ini juga mencakup deteksi dan analisis *missing values* serta *outliers*.

3. *Data Preparation*

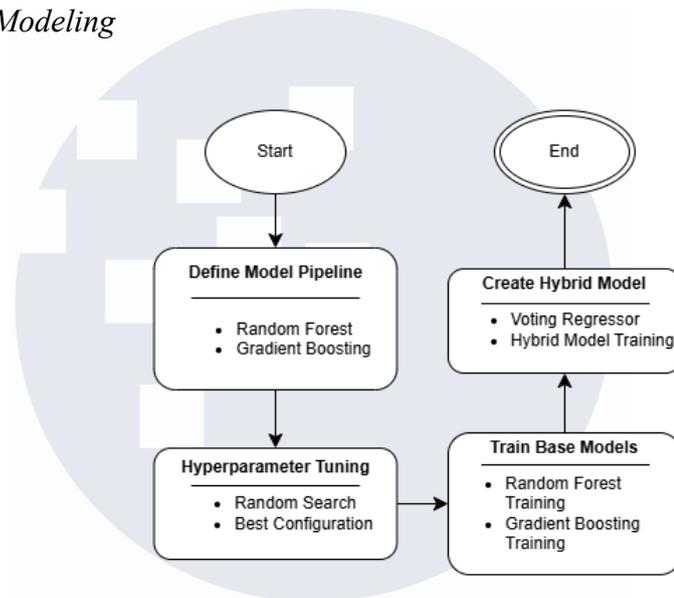


Gambar 3.5 Diagram Data Preparation

Gambar 3.5 mencakup skema singkat mengenai proses pengolahan dan pembersihan data yang telah dikumpulkan. Untuk memastikan model dapat memahami variabel kategori dengan baik, dilakukan proses *label encoding*, yaitu mengonversi data kategori seperti bulan ke dalam bentuk numerik. Selanjutnya, dilakukan normalisasi atau standarisasi data guna memastikan skala data seragam, terutama jika model yang digunakan sensitif terhadap perbedaan skala. Terakhir, mengubah format data menjadi bentuk yang sesuai dengan kebutuhan analisis, misalnya mengonversi kolom tanggal ke dalam format *datetime* agar dapat digunakan dalam analisis deret waktu. Pada tahap ini, *outliers* tidak dihapus dari dataset karena dianggap sebagai bagian dari variasi alami dalam populasi. Penelitian sebelumnya menunjukkan bahwa *outliers* tidak selalu merupakan kesalahan data, melainkan bisa menjadi indikasi adanya fenomena unik yang perlu dipertimbangkan [41]. Penghapusan *outliers* pada

penelitian ini dapat menghilangkan informasi berharga dan merusak pemahaman terhadap variasi alami dalam populasi yang sedang diteliti serta menurunkan performa model. Oleh karena itu, dalam analisis ini, *outliers* dipertahankan untuk menghindari distorsi hasil yang bisa terjadi dengan mengabaikan variasi yang sah.

4. Modeling



Gambar 3.6 Diagram Modeling

Dalam tahap *modeling*, model *Random Forest*, *Gradient Boosting*, dan *Hybrid* diimplementasikan untuk membangun model prediksi permintaan darah. Alur pemodelan dapat dilihat pada Gambar 3.6. Untuk memberikan gambaran komprehensif mengenai perbedaan mendasar antara algoritma yang umum digunakan dalam pemodelan prediktif, berikut disajikan Tabel 3.3 yang menyajikan informasi perbandingan dari *Decision Tree*, *Random Forest*, dan *Gradient Boosting*. Tabel ini akan memudahkan pemahaman mengapa pendekatan *Random Forest* dan *Gradient Boosting* dipilih dalam penelitian ini.

Tabel 3.3 Perbandingan Decision Tree, Random Forest, dan Gradient Boosting

Algoritma	Jenis/Metode	Keunggulan	Kelemahan
Decision Tree	Model pohon keputusan tunggal	<ul style="list-style-type: none"> • Mudah dipahami dan diinterpretasikan • Cepat dalam proses pelatihan • Tidak memerlukan pra-pemrosesan data yang rumit 	<ul style="list-style-type: none"> • Rentan terhadap <i>overfitting</i> • Sensitif terhadap <i>noise</i> • Stabilitas rendah (perubahan data dapat mengubah hasil)
Random Forest	<i>Ensemble learning (bagging)</i> dengan banyak pohon	<ul style="list-style-type: none"> • Mengurangi <i>overfitting</i> dengan rata-rata prediksi dari banyak pohon • Lebih stabil dan <i>robust</i> • Menangani data berdimensi tinggi dengan baik 	<ul style="list-style-type: none"> • Model lebih kompleks dan kurang interpretatif • Memerlukan sumber daya komputasi lebih tinggi
Gradient Boosting	<i>Ensemble learning (boosting)</i> yang	<ul style="list-style-type: none"> • Sangat akurat dengan menangkap pola non-linier yang 	<ul style="list-style-type: none"> • Proses pelatihan lebih lambat • Rentan

Algoritma	Jenis/Metode	Keunggulan	Kelemahan
	membangun pohon secara bertahap	kompleks • Mampu memperbaiki kesalahan model secara iteratif	terhadap <i>overfitting</i> jika tidak dikontrol dengan baik • Parameter <i>tuning</i> yang kompleks

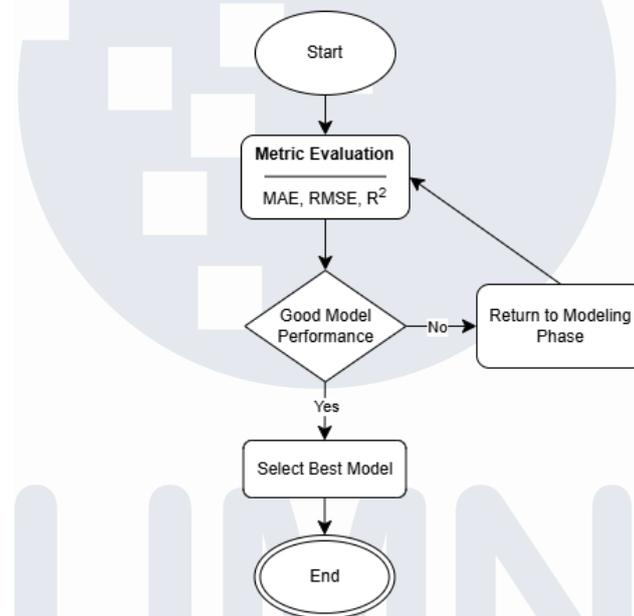
Pemilihan algoritma Random Forest dan Gradient Boosting dalam penelitian ini didasarkan pada keunggulan masing-masing metode dalam menangani kompleksitas data prediksi permintaan darah. Random Forest, sebagai metode *ensemble* berbasis *bagging*, memiliki keunggulan dalam mengurangi *overfitting* dengan menggabungkan prediksi dari banyak pohon keputusan. Selain itu, algoritma ini lebih stabil dalam menangani data berdimensi tinggi serta mampu mengidentifikasi fitur yang paling berpengaruh dalam prediksi. Namun, karena pendekatannya berbasis rata-rata prediksi dari banyak model, Random Forest cenderung kurang fleksibel dalam menangkap pola non-linear yang kompleks.

Di sisi lain, Gradient Boosting menggunakan pendekatan *boosting* yang secara iteratif memperbaiki kesalahan model sebelumnya, menjadikannya lebih akurat dalam menangkap pola non-linear yang sulit dikenali oleh model lain. Metode ini sangat efektif dalam meningkatkan akurasi prediksi, terutama pada data yang memiliki hubungan kompleks antar variabel. Namun, Gradient Boosting memiliki proses pelatihan yang lebih lambat dibandingkan Random

Forest dan rentan terhadap *overfitting* jika tidak dikonfigurasi dengan baik.

Pemilihan model hybrid yang menggabungkan Random Forest dan Gradient Boosting juga dilakukan untuk memanfaatkan keunggulan masing-masing algoritma. Pendekatan gabungan ini memungkinkan pemanfaatan kekuatan agregasi dari Random Forest dan kemampuan optimasi iteratif dari Gradient Boosting.

5. Evaluation

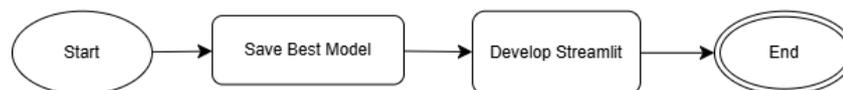


Gambar 3.7 Diagram Evaluation

Evaluasi model merupakan tahap kritis dalam proses ini untuk memahami sejauh mana model yang dihasilkan memenuhi tujuan awal yang ditetapkan. Dalam penelitian ini, model prediksi yang dikembangkan diuji dengan menggunakan metrik evaluasi yang sesuai, seperti *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, dan *R-squared*, untuk menentukan akurasi dan kinerja prediksi. Adapun gambaran proses evaluasi dapat dilihat pada Gambar 3.7. Hasil evaluasi dengan nilai MAE terkecil dan R^2

tertinggi akan dipilih sebagai *best model* yang akan digunakan sebagai model prediksi.

6. *Deployment*



Gambar 3.8 Diagram *Deployment*

Gambar 3.8 mengilustrasikan tahap *deployment* yang melibatkan implementasi model prediksi ke dalam bentuk website *dashboard*. Setelah melewati tahap evaluasi, model yang berhasil dievaluasi akan disimpan dan diintegrasikan ke *web dashboard* interaktif menggunakan Streamlit. Visualisasi ini bertujuan untuk memberikan pandangan yang jelas dan informatif mengenai tren permintaan darah, sehingga pihak manajemen dapat mengambil keputusan yang lebih baik berdasarkan data yang ada. Dengan memanfaatkan hasil prediksi dalam pengambilan keputusan, PMI Kabupaten Tangerang dapat meningkatkan efektivitas kelola stok darahnya.

3.3. Teknik Pengumpulan Data

3.3.1 Wawancara

Pada penelitian ini dilakukan wawancara secara langsung dengan Ibu Tuti Hamsyah selaku Kepala Bidang Pelayanan PMI Kabupaten Tangerang. Wawancara dilakukan pada tanggal 24 November 2024 di kantor PMI Kabupaten Tangerang yang beralamatkan di Kadu Jaya, Kec. Curug, Tangerang. Wawancara ini bertujuan untuk mendapatkan informasi mendalam tentang sistem pengelolaan darah yang saat ini digunakan serta mengidentifikasi kebutuhan dan tantangan yang dihadapi dalam memprediksi permintaan darah.

3.3.2 Studi Pustaka

Tinjauan pustaka dilakukan untuk membangun landasan teoritis yang kuat dalam pengembangan model prediksi permintaan darah. Tinjauan ini mencakup penelitian tentang teknik-teknik *machine learning* untuk prediksi deret waktu, khususnya yang berkaitan dengan manajemen persediaan di bidang kesehatan untuk mendapatkan gambaran permasalahan yang semakin kuat.

3.3.3 Sampling Penelitian

Dalam penelitian ini, populasi yang menjadi fokus adalah data permintaan darah yang dikelola oleh Palang Merah Indonesia (PMI) Kabupaten Tangerang. Sampel data yang digunakan mencakup informasi yang diambil langsung dari PMI. Data tersebut terdiri atas beberapa kolom penting, yaitu Jenis Darah, Golongan Darah, serta Jumlah Permintaan Darah yang terperinci berdasarkan bulan dan tahun. Teknik sampling yang digunakan dalam penelitian ini adalah *purposive sampling*, yaitu teknik pengambilan sampel berdasarkan kriteria atau pertimbangan tertentu. Dalam hal ini, data yang digunakan dipilih karena memenuhi kriteria waktu (2021–2024), relevansi dengan penelitian, dan kelengkapan informasi yang diperlukan. Secara total, terdapat 2880 data permintaan darah yang diperoleh pada PMI Kabupaten Tangerang.

3.4. Teknik Analisis Data

Analisis data pada penelitian ini dilakukan menggunakan bahasa pemrograman Python dengan *tools* Jupyter Notebook. Jupyter Notebook dipilih karena kecepatan komputasi yang tinggi, memastikan eksekusi perintah dan perhitungan efisien dalam mengolah dataset permintaan darah pada PMI Kabupaten Tangerang, termasuk dalam proses pemodelan *machine learning* untuk prediksi permintaan darah. Selain itu, keunggulan visualisasi interaktifnya mempermudah pemahaman pola dan distribusi data dengan tampilan yang dinamis. Fleksibilitas dalam integrasi bahasa pemrograman, terutama Python, membuat Jupyter Notebook menjadi alat yang sangat adaptif untuk pengolahan dan analisis data yang efektif.

Sebagai langkah awal, data yang diperoleh dari PMI Kabupaten Tangerang akan melalui tahapan *data understanding*. Pada proses ini dilakukan eksplorasi lebih lanjut mengenai informasi data. Proses eksplorasi data (EDA) bertujuan untuk memahami karakteristik dataset sebelum dilakukan pemrosesan lebih lanjut. Tahapan ini mencakup pengecekan jenis data, visualisasi distribusi data serta hubungan antar variabel guna mengidentifikasi pola atau tren yang mungkin relevan. Eksplorasi ini juga mencakup pendeteksian *missing values*, yang dapat mempengaruhi kualitas data dan perlu ditangani sebelum model prediksi diterapkan. Berikutnya dilakukan tahapan *data preparation* yang mencakup transformasi data, standarisasi, dan perubahan format data sesuai dengan kebutuhan tahap berikutnya, yaitu *modeling*. Tahap persiapan data ini sangat penting untuk memastikan dataset yang digunakan dalam *modeling* sudah bersih dan dapat diandalkan.

Tahap *modeling* dimulai dengan penerapan algoritma Random Forest, Gradient Boosting, dan Hybrid. Ketiga algoritma ini dipilih karena kemampuan masing-masing dalam menangani data yang besar dan kompleks. Model-model ini dilatih menggunakan dataset yang sudah dipersiapkan sebelumnya, yang mencakup variabel-variabel seperti jenis dan golongan darah serta fluktuasi permintaan darah berdasarkan bulan dan tahun. Hasil prediksi dari ketiga algoritma ini kemudian dibandingkan untuk melihat model mana yang memiliki performa paling baik. Setelah tahap *modeling* selesai, dilakukan evaluasi terhadap performa model dengan menggunakan metrik evaluasi yang sesuai. Evaluasi ini bertujuan untuk mengukur akurasi model dalam memprediksi permintaan darah dan untuk memastikan bahwa model yang dikembangkan memenuhi tujuan analisis data. Selain metrik evaluasi standar, juga dilakukan perhitungan *Confidence Interval* (CI) untuk memberikan konfirmasi mengenai tingkat kepercayaan dari prediksi yang dihasilkan. CI berfungsi untuk menunjukkan rentang nilai prediksi yang sebenarnya kemungkinan besar akan berada, sehingga dapat memberi gambaran ketidakpastian model dan membantu dalam pengambilan keputusan yang lebih akurat [42]–[44].

Langkah terakhir yakni *Deployment*, dilakukan dengan mengembangkan website menggunakan Streamlit. Bentuk visualisasi prediksi permintaan darah yang telah dibuat pada Jupyter Notebook akan diintegrasikan ke dalam web ini sehingga dapat digunakan oleh pengguna secara interaktif. Tahap ini bertujuan untuk menyajikan hasil prediksi dalam format yang mudah dipahami dan dapat diakses oleh pihak yang berkepentingan, seperti pengelola UTD PMI.

3.4.1 Variabel Penelitian

Pada penelitian ini, terdapat dua buah variabel penelitian, yakni variabel dependen dan variabel independen.

1. Variabel Dependen

Variabel dependen merupakan variabel yang mempunyai keterikatan atau dipengaruhi oleh variabel independen. Variabel dependen dalam penelitian ini menggambarkan jumlah permintaan darah yang akan diprediksi berdasarkan berbagai faktor lainnya. Di mana model prediksi akan digunakan untuk mengestimasi nilai dari jumlah permintaan darah pada PMI Kabupaten Tangerang.

2. Variabel Independen

Variabel independen adalah variabel yang dianggap sebagai penyebab atau faktor yang memengaruhi variabel dependen. Dalam penelitian ini, variabel independen terdiri atas:

a. Jenis Darah

Jenis darah yang dimaksud mencakup berbagai jenis darah yang ada, seperti *Whole Blood (WB)*, *Packed Red Cells (PRC)*, dan lain-lain. Jenis darah ini dapat memengaruhi pola permintaan darah yang berbeda-beda.

b. Golongan Darah

Variabel ini mengacu pada golongan darah, seperti A, B, O, dan AB. Golongan darah juga dapat mempengaruhi kebutuhan permintaan darah pada PMI.

c. Bulan

Waktu dalam satuan bulan yang mencakup pola fluktuasi permintaan darah berdasarkan waktu tertentu dalam tahun.

d. Tahun

Variabel ini merujuk pada tahun yang dapat memengaruhi tren permintaan darah karena mungkin terdapat perbedaan permintaan antar tahun.

Penjelasan lebih lanjut mengenai variabel-variabel tersebut dapat dilihat pada Tabel 3.4 berikut:

Tabel 3.4 Definisi Operasional Variabel Penelitian

Variabel	Jenis	Rentang/Kategori	Keterangan
Golongan Darah	Input	A, B, AB, O	Golongan darah berdasarkan antigen atau antibodi.
Jenis Darah	Input	PRC, PRC-Pediatrik, PRC-NAT, TC, TC-Apheresis, WB, FFP, FFP-NAT, AHF, Plasma-Konvalesen, Leucodepleted, Leucodepleted-Pediatrik, PRC-Pediatrik-50cc, PRC-Pediatrik-75cc, PRC-Pediatrik-100cc	Jenis komponen darah yang diminta berdasarkan klasifikasi laboratorium darah PMI
Bulan	Input	Januari–Desember	Menunjukkan bulan terjadinya permintaan darah.
Tahun	Input	2021–2024	Menunjukkan tahun permintaan darah berlangsung.
Jumlah Permintaan	Target	0–1882 kantong/bulan	Jumlah total permintaan darah yang menjadi

Variabel	Jenis	Rentang/Kategori	Keterangan
			variabel target dalam pemodelan.

3.5. Teknik Pengujian

Pengujian hasil akhir pada penelitian ini menggunakan metrik evaluasi. Metrik Evaluasi digunakan untuk menilai performa model secara kuantitatif dan kualitatif. Metrik ini berperan penting dalam memvalidasi hipotesis penelitian dan membandingkan performa model yang digunakan. Metrik evaluasi cukup beragam, tergantung pada tujuan pemodelan, apakah segmentasi, klasifikasi, prediksi, dan lain-lain. Pada penelitian ini digunakan *MAE (Mean Absolute Error)*, *RMSE (Root Mean Squared Error)*, dan *R² (R-squared)*.

1) MAE (*Mean Absolute Error*)

Mean Absolute Error menghitung selisih rata-rata antara nilai yang dihitung dan nilai aktual. MAE juga dikenal sebagai akurasi yang bergantung pada skala karena menghitung kesalahan dalam pengamatan yang diambil pada skala sama yang digunakan untuk memprediksi akurasi model *machine learning* [31]. Interpretasinya yang jelas terhadap prediksi rata-rata, ketahanan nya terhadap *outlier*, dan sifatnya yang praktis untuk diimplementasikan menjadikan MAE sebagai model evaluasi yang unggul. Semakin rendah atau semakin nilai MAE mendekati 0 maka model prediksi bersifat semakin bagus.

$$\frac{\sum_{i=1}^n |Y_i - X_i|}{n}$$

Rumus 2.5 Rumus Menghitung MAE

2) RMSE (*Root Mean Square Error*)

RMSE adalah nilai akar kuadrat dari rata-rata kuadrat semua error. RMSE dianggap sebagai metrik evaluasi yang sangat baik untuk melakukan prediksi numerik [31]. RMSE cocok untuk membandingkan kesalahan prediksi antara

berbagai model atau konfigurasi untuk suatu variabel tertentu, bukan antar variabel, karena bergantung pada skala. RMSE memiliki satuan yang sama dengan variabel target. RMSE yang lebih rendah menunjukkan kinerja model yang lebih baik, tetapi nilainya harus dibandingkan dengan skala variabel target agar masuk akal.

$$\sqrt{\sum_{i=1}^n \frac{(\hat{Y}_i - Y_i)^2}{n}}$$

Rumus 2.6 Rumus Menghitung RMSE

3) R² (*Coefficient of Determination*)

Coefficient of Determination (R²) adalah suatu ukuran yang digunakan untuk mengukur seberapa baik model prediksi dapat menjelaskan variasi data yang ada tergantung pada rasio deviasi total hasil yang dijelaskan oleh model [31]. R² memiliki rentang nilai antara 0 dan 1, nilai 1 menunjukkan bahwa model prediksi dapat menjelaskan 100% variasi data yang ada, sedangkan 0 menunjukkan bahwa model prediksi tidak dapat menjelaskan variasi data sama sekali.

$$R^2 = 1 - \frac{\sum(Y_i - \hat{Y})^2}{\sum(Y_i - \bar{Y})^2}$$

Rumus 2.7 Rumus Menghitung R²