

BAB II

LANDASAN TEORI

2.1 Tinjauan Pustaka

2.1.1 Rujukan Rumah Sakit

Proses rujukan rumah sakit melibatkan penyampaian informasi medis yang relevan dari dokter atau tenaga medis yang merujuk kepada rumah sakit yang dituju, sehingga tim medis di sana dapat mempersiapkan diri untuk menyambut pasien dan memberikan perawatan yang sesuai. Rujukan ini bisa dilakukan dalam berbagai kasus, mulai dari diagnosis yang kompleks, perawatan spesialis tertentu, hingga intervensi medis yang mendesak.

Biasanya, rujukan rumah sakit dilakukan setelah evaluasi dan pertimbangan yang cermat oleh dokter atau tenaga medis yang merawat pasien, dengan tujuan memberikan perawatan yang terbaik sesuai dengan kebutuhan medis pasien tersebut. Proses ini juga melibatkan koordinasi antara berbagai pihak, termasuk dokter primer, spesialis, dan fasilitas medis, untuk memastikan bahwa pasien mendapatkan perawatan yang tepat dan tepat waktu.

2.1.1.1 Faktor Penyebab Rujukan Rumah Sakit

Rujukan rumah sakit adalah proses di mana seorang dokter atau tenaga medis merujuk pasien ke rumah sakit tertentu untuk pemeriksaan lebih lanjut, perawatan, atau penanganan spesifik yang tidak dapat dilakukan di fasilitas medis tempat pasien tersebut awalnya berkonsultasi. Rujukan ini biasanya terjadi ketika dokter primer atau tenaga medis yang merawat pasien menganggap bahwa pasien

membutuhkan perhatian dan perawatan yang lebih intensif atau spesialis, yang mungkin tidak tersedia di fasilitas medis lokal mereka.

2.1.2 Machine Learning

Pembelajaran Mesin atau Machine learning adalah cabang kecerdasan buatan yang memungkinkan komputer untuk belajar dari data yang ada dan menganali pola secara otomatis. Ada beberapa jenis utama *Machine Learning*:

1. **Supervised Learning:** Model belajar dari data yang telah diberi label. Tujuan utama adalah untuk memprediksi label yang tepat untuk data baru. Contoh algoritma supervised learning termasuk *Regresi Linier*, *Naive Bayes*, *K-Nearest Neighbors (KNN)*, *Decision Trees*, *Random Forest*, dan *Support Vector Machines (SVM)*.
2. **Unsupervised Learning:** Model belajar dari data yang tidak memiliki label. Tujuannya adalah untuk menemukan struktur dalam data, seperti pola atau kelompok, tanpa panduan eksternal. Contoh algoritma *unsupervised learning* termasuk *K-Means Clustering*, *Hierarchical Clustering*, dan *Principal Component Analysis (PCA)*.

2.1.3 Klasifikasi

Menurut Towa P. Hmakotrda dan J.N.B. Tairas (1995), klasifikasi adalah pengelompokan sistematis dari objek, gagasan, buku, atau benda lain ke dalam golongan tertentu berdasarkan ciri-ciri yang sama. Dalam kehidupan sehari-hari, klasifikasi sudah sering dilakukan oleh manusia. Contohnya di supermarket, pasar, atau toko buku, di mana pedagang mengelompokkan barang dagangan yang sejenis

dalam satu kelompok yang sama untuk memudahkan pembeli dalam memilih kebutuhan yang diperlukan.

1. Regresi Logistik (*Logistic Regression*): Memprediksi probabilitas dari sebuah kejadian berdasarkan variabel independen.
2. *K-Nearest Neighbors* (KNN): Mengklasifikasikan data berdasarkan mayoritas dari kategori-kategori tetangga terdekatnya.
3. *Support Vector Machines* (SVM): Mencari hyperplane terbaik untuk memisahkan dua kelas data dengan margin maksimal.
4. *Decision Trees*: Menggunakan serangkaian keputusan untuk mencapai klasifikasi pada data berdasarkan fitur-fitur yang ada.
5. *Random Forest*: Ensemble dari pohon keputusan yang digunakan untuk menggabungkan hasil prediksi dari beberapa model.
6. *Naive Bayes*: Mengaplikasikan teorema Bayes dengan asumsi bahwa fitur-fitur yang digunakan saling independen.
7. *Neural Networks*: Jaringan dari neuron buatan yang digunakan untuk mempelajari pola-pola kompleks dalam data.

2.1.3.1 Random Forest

Random Forest adalah sekumpulan beberapa pohon keputusan (*tree*) yang masing-masing pohon bergantung terhadap nilai piksel yang diambil secara acak dan independen untuk setiap vektor (Breiman, 2001). Metode ini cenderung tidak mengalami overfitting dan dapat melakukan pemrosesan dengan cepat, sehingga memungkinkan pengguna untuk memproses sebanyak mungkin pohon yang diinginkan (Breiman and Clutler, 2005).

2.1.3.2 Support Vector Machine (SVM)

Support Vector Machine adalah klasifikasi dalam pembelajaran yang disupervisi. Algoritma SVM bekerja dengan menggunakan pemetaan non-linear untuk mengubah data pelatihan asli ke dimensi yang lebih tinggi (Alven Ritonga, 2018).

2.1.3.3 Evaluasi *Matrix Confusion*

Hasil dari klasifikasi akan dievaluasi untuk mengukur tingkat keberhasilan metode yang digunakan dalam pengujian. Salah satu metode evaluasi dalam analisis sentimen adalah menggunakan *Confusion Matrix*. *Confusion Matrix* adalah metode untuk mengevaluasi hasil dari implementasi algoritma klasifikasi melalui sebuah tabel, seperti yang terlihat pada gambar 2.1 di bawah.

Tabel 2. 1 Evaluasi dari *Confusion Matrix*

		Nilai Aktual	
		Positive	Negative
Nilai Prediksi	Positive	TP	FP
	Negative	FN	TN

TP (*True Positive*) adalah jumlah dokumen positif yang diklasifikasikan oleh sistem sebagai kelas positif, sedangkan FP (*False Positive*) adalah jumlah dokumen negatif yang diklasifikasikan sebagai kelas positif. FN (*False Negative*) adalah jumlah dokumen positif yang diklasifikasikan oleh sistem sebagai kelas negatif,

sedangkan TN (*True Negative*) adalah jumlah dokumen negatif yang diklasifikasikan sebagai kelas negatif oleh sistem.

Nilai-nilai ini akan digunakan untuk menghitung accuracy, precision, recall, dan F-measure. Accuracy adalah perbandingan antara jumlah prediksi yang benar dengan total jumlah prediksi. Precision adalah perbandingan antara jumlah prediksi benar untuk data positif dengan total prediksi untuk data positif. Recall adalah perbandingan antara jumlah prediksi benar untuk data positif dengan total data positif sebenarnya. F-measure adalah rata-rata harmonis dari precision dan recall.

2.1.4 Metode CRISP-DM

CRISP-DM adalah metode yang menggunakan model proses pengembangan data yang sering digunakan oleh para ahli untuk memecah masalah. Metodologi ini terdiri dari enam tahap, yaitu Business Understanding, Data Understanding, Data Preparation, Modelling, Evaluation, dan Deployment. Berikut penjelasan masing-masing tahapan dalam metodologi ini.

1. *Business Understanding* (Pemahaman Bisnis)

Pada tahap ini, fokusnya adalah memahami kebutuhan dan tujuan dari perspektif bisnis. Ini melibatkan pendefinisian masalah dalam data mining dan menentukan rencana serta strategi untuk mencapai tujuan tersebut.

2. *Data Understanding* (Pemahaman Data)

Tahap ini dimulai dengan mengumpulkan data, mendeskripsikan data, dan mengevaluasi kualitas data yang diperoleh.

3. *Data Preparation* (Persiapan Data)

Pada tahap ini, dataset akhir dibentuk dari data mentah. Proses ini mencakup pembersihan data (*Data Cleaning*), pemilihan data (*Data Selection*), serta

transformasi data (Data Transformation) untuk dijadikan input dalam tahap pemodelan.

4. *Modelling* (Permodelan)

Tahap ini melibatkan penggunaan metode statistik dan Machine Learning untuk menentukan teknik data mining, alat bantu, dan algoritma yang akan diterapkan. Selanjutnya, teknik dan algoritma tersebut diterapkan pada data dengan bantuan alat bantu. Jika diperlukan penyesuaian data untuk teknik tertentu, proses dapat kembali ke tahap persiapan data.

5. *Evaluation* (Pengujian)

Hasil dari data mining yang dihasilkan pada tahap pemodelan diinterpretasikan dan dievaluasi. Tujuan evaluasi ini adalah memastikan model yang diterapkan sesuai dengan tujuan yang ditetapkan pada tahap pertama.

6. *Deployment* (Penyebaran)

Tahap *Deployment* adalah tahap yang paling berharga dari proses CRISP-DM. Perencanaan untuk deployment dimulai selama Business Understanding dan harus mencakup cara menghasilkan nilai model, mengonversi skor menjadi keputusan, dan mengintegrasikan keputusan tersebut dalam sistem operasional

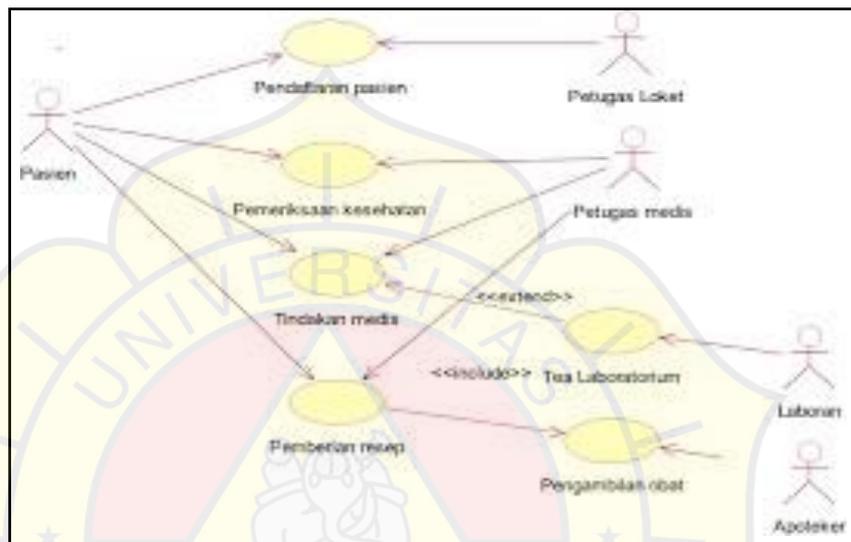
Pada akhirnya, rencana deployment harus mengakui bahwa model tidak statis. Model dibangun dari data yang mewakili kondisi pada waktu tertentu, sehingga perubahan waktu dapat mengubah karakteristik data. Oleh karena itu, model harus dipantau dan mungkin diganti dengan model yang telah diperbarui.

2.1.5 UML

Menurut Prihandoyo, M. T.(2018:127) “*Unified Modeling Language*” adalah salah satu metode pemodelan visual yang digunakan dalam perancangan dan

pembuatan perangkat lunak yang berorientasi objek. UML merupakan standar penulisan atau semacam blueprint yang mencakup proses bisnis dan penulisan kelas-kelas dalam bahasa yang spesifik. Terdapat beberapa jenis diagram dalam UML sebagai berikut:

2.1.5.1 Use Case Diagram



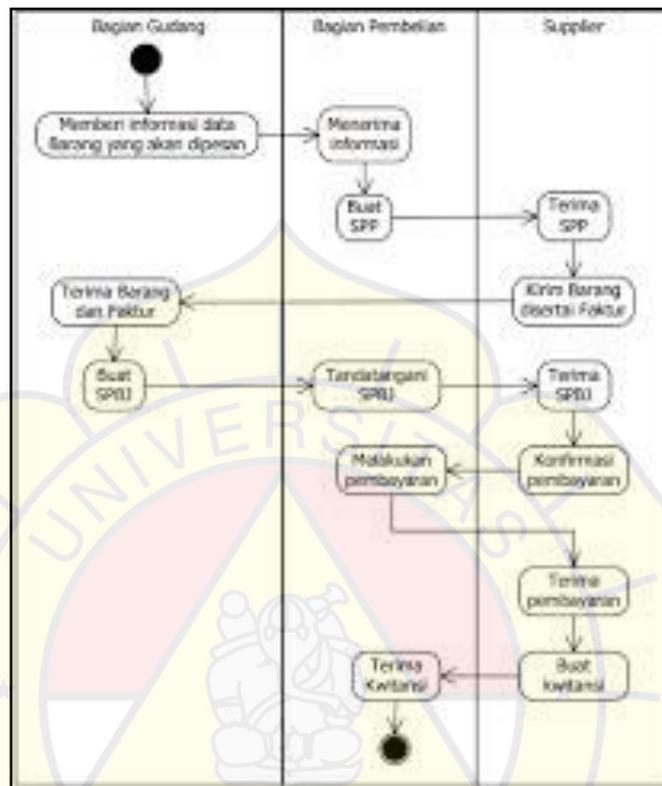
Gambar 2. 1 Usecase Diagram

Menurut Prihandoyo, M. T.(2018:127) “*Use Case* merupakan gambaran dari fungsionalitas yang di harapkan dari sebuah sistem dan merepresentasikan sebuah interaksi antara aktor dan sistem. Didalam *use case* terdapat aktor yang merupakan pekerjaan di sistem”. Diagram *Use Case* bersifat statis dan sangat penting untuk mengorganisasi dan memodelkan perilaku sistem yang dibutuhkan serta diharapkan pengguna, dapat di lihat gambar 2.1 di atas.

2.1.5.2 Activity Diagram

Menurut Prihandoyo, M. T.(2018:127) “*Activity Diagram* Merupakan gambaran air dari aktivitas-aktivitas didalam sistem yang berjalan”. *Activity*

diagram ini bersifat dinamis dalam pemodelan fungsi-fungsi suatu sistem dan memberi tekanan pada aliran kendali antar objek, dapat di lihat gambar 2.2 di bawah.



Gambar 2. 2 Activity Diagram

2.1.6 Pemrograman Sistem

2.1.6.1 Python

Python adalah bahasa pemrograman tingkat tinggi yang populer, dikenal karena sintaksis yang mudah dipahami dan dipelajari. Ia sering digunakan dalam berbagai bidang seperti pengembangan web, ilmu data, kecerdasan buatan, dan lainnya. Python memiliki berbagai keunggulan, termasuk kemudahan penggunaan, dukungan untuk berbagai paradigma pemrograman, serta ekosistem modul yang kaya.

2.1.6.2 HTML

HTML (Hypertext Markup Language) adalah markah bahasa standar yang dipergunakan untuk membuat dan merancang halaman situs. HTML memungkinkan pembuat web untuk menentukan struktur dan konten sebuah halaman web dengan menggunakan elemen dan tag yang ditetapkan. Setiap elemen HTML menggambarkan berbagai bagian dari halaman, seperti judul, paragraf teks, gambar, tautan, dan lainnya. Dengan menggabungkan berbagai elemen ini, pembuat web dapat membuat halaman web yang interaktif dan informatif. HTML berfungsi sebagai dasar bagi pengembangan web, dan sering digunakan bersama dengan CSS (Cascading Style Sheets) dan JavaScript untuk menghasilkan pengalaman web yang kaya dan dinamis.

2.1.6.3 Flask

Flask adalah kerangka kerja web ringan dan ramah pemula untuk Python. Ia dirancang untuk membuat pengembangan web yang cepat dan mudah dengan penekanan pada kesederhanaan dan modularitas. Flask memberikan alat yang diperlukan untuk membangun aplikasi web dari yang sederhana hingga yang kompleks, dengan dukungan untuk pengembangan API, templating, penanganan formulir, dan masih banyak lagi, menjadikannya pilihan yang populer di kalangan pengembang Python.

2.1.6.4 MySQL

Menurut Kustiyahningsih (2011:145), “MySQL adalah sebuah basis data yang mengandung satu atau jumlah tabel”. Tabel terdiri atas sejumlah baris, dan setiap baris mengandung satu atau lebih kolom.

Dengan berbagai keunggulan yang dimilikinya, *software* database ini banyak digunakan oleh praktisi untuk membangun berbagai proyek. Fasilitas API (*Application Programming Interface*) yang dimiliki oleh MySQL memungkinkan berbagai aplikasi komputer yang ditulis dengan berbagai bahasa pemrograman untuk mengakses basis data MySQL.

MySQL termasuk jenis RDBMS (*Relational Database Management System*). RDBMS sendiri mengenal istilah seperti tabel, baris, dan kolom yang digunakan dalam perintah-perintah di MySQL.

2.2 Kajian Penelitian Terdahulu

Penelitian yang dilakukan saat ini memiliki beberapa referensi dan bersifat untuk mengembangkan dari beberapa penelitian terdahulu.

2.2.1 Paper 1

Penerapan *Data Mining* dalam Analisis Prediksi Kanker Paru Menggunakan Algoritma *Random Forest*. Sari, L., Romadloni, A. & Listyaningrum, R. 2023. Vol.14, No.01, Januari 2023. SINTA 3.

- **Tujuan Penelitian :** Penelitian dari *Paper 1* membahas tentang melakukan klasifikasi terhadap kasus penyakit kanker secara dini guna untuk mengetahui penyakit penderita lebih awal untuk mendapatkan penanganan yang tepat. Penelitian ini memiliki hasil akhir dalam bentuk *akurasi*, *recall* dan *precision* terhadap algoritma yang digunakan.
- **Metodologi Penelitian :** Penelitian Paper 1 menggunakan metodologi data mining pada kasus klasifikasi penyakit kanker. Data mining yang digunakan menggunakan penerapan KDD sebagai Teknik penyelesaian dari kasus yang

dilakukan. Hal ini dimaksudkan karena pada penelitian ini tidak memiliki sebuah deployment dari hasil akhir yang telah didapatkan.

- **Temuan Utama :** Penelitian Paper 1 menggunakan data dari Kaggle.com yang berisi informasi tentang beberapa data pengidap penyakit kanker. Peneliti menggunakan algoritma *Random Forest* untuk melakukan penelitian terbarunya dan melakukan pengukuran *akurasi*, *recall* serta *precision* dari algoritma tersebut. Peneliti menggunakan beberapa metode pendukung guna untuk menyelesaikan penelitian yang dilakukan. Penelitian menggunakan pemodelan pada penyocokan data training dan data testing dengan menggunakan K-Fold Cross Validation dengan nilai $K=10$. Setelah menemukan hasil dari pemrosesan algoritma *Random Forest*, peneliti menggunakan diagram AUC untuk menentukan Tingkat keefektifan yang dihasilkan oleh algoritma *Random Forest*.
- **Kesimpulan :** Peneliti menggunakan Algoritma *Naïve Bayes* sebagai parameter pembandingan dengan Algoritma *Random Forest*. Pada penelitian yang sama dengan menggunakan Algoritma *Naïve Bayes*, hasil dari proses klasifikasi memiliki nilai akurasi sebesar 95,1% sedangkan jika dibandingkan dengan proses klasifikasi dari Algoritma *Random Forest* yang memiliki akurasi sebesar 98,3% tentu memiliki gap yang cukup besar untuk sebuah penelitian. Penelitian memiliki kesimpulan untuk kasus klasifikasi penyakit kanker, Algoritma *Random Forest* memiliki ketepatan proses yang lebih tinggi dibanding dengan Algoritma *Naïve Bayes*.

2.2.2 Paper 2

Klasifikasi Penyakit Stunting Dengan Menggunakan Algoritma Support Vector Machine Dan Random Forest. Mahmin ,B., Dinda Betaria, H., Oloan, S. 2003. Vol.6, No.02, December 2023. SINTA 4.

- **Tujuan Penelitian :** Penelitian *paper 2* mengembangkan model yang dapat memprediksi keberadaan atau risiko stunting pada anak-anak berdasarkan faktor-faktor yang relevan, seperti pola makan, status gizi, lingkungan, dan faktor-faktor lainnya. Penelitian ini bertujuan untuk membantu dalam deteksi penyakit stunting, yang dapat mengarah pada *intervensi* yang lebih efektif dan tepat waktu untuk mencegah dampak jangka panjangnya pada pertumbuhan dan perkembangan anak.
- **Metodologi Penelitian :** Penelitian *paper 2* menjelaskan Teknologi kesehatan ini, khususnya dalam machine learning, mempengaruhi klasifikasi penyakit dengan cepat. SVM dan *Random Forest* digunakan untuk mengklasifikasi faktor penyebab stunting. SVM adalah teknik untuk meramalkan dengan model machine learning, sementara *Random Forest* menggunakan pemisahan *biner rekursif* dalam struktur pohon untuk prediksi. Kedua algoritma ini memainkan peran penting dalam memprediksi klasifikasi model *machine learning*.
- **Temuan Utama :** Penelitian *paper 2* menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest* untuk mengklasifikasi faktor-faktor penyebab stunting pada anak-anak, menyoroti kemampuan *machine learning* dalam mengatasi masalah kesehatan yang kompleks. Penelitian berhasil mengidentifikasi faktor-faktor risiko yang berkontribusi pada kondisi stunting pada anak-anak, memperkuat pemahaman tentang penyakit tersebut. SVM dan

Random Forest terbukti efektif dalam memprediksi klasifikasi penyakit stunting, memberikan kontribusi berharga dalam diagnosis dan pengelolaan kondisi tersebut.

- **Kesimpulan** : Penelitian menggunakan algoritma SVM (*Support Vector Machine*) memiliki akurasi sebesar 65.6% untuk *data testing* dan 62.7% untuk *data training*. Di sisi lain, Algoritma *Random Forest* menunjukkan akurasi yang lebih tinggi, dengan hasil 88.2% untuk *data testing* dan 98.8% untuk *data training*. Perbedaan ini menandakan performa yang lebih baik dari *Random Forest* dibandingkan SVM. Setelah melalui proses *hypertuning*, akurasi SVM meningkat menjadi 81%, mendekati tingkat akurasi *Random Forest*. Hal ini menunjukkan bahwa proses *hypertuning* membantu meningkatkan kinerja model SVM, meskipun *Random Forest* tetap memiliki performa yang lebih baik dalam mengklasifikasikan penyakit stunting dalam penelitian ini.

2.2.3 Paper 3

Perbandingan *Support Vector Machine* dan *Modified Balanced Random Forest* dalam Deteksi Pasien Penyakit Diabetes. Mahendra Dwifabri, P., Muhammad Irvan, T., Adnan Imam, H., Adiwijaya, A. 2021. Vol.5, No.02, April 2021. SINTA 2.

- **Tujuan Penelitian** : membandingkan kinerja dua metode klasifikasi, yaitu *Support Vector Machine* (SVM) dan *Modified Balanced Random Forest*, dalam mendeteksi pasien dengan penyakit diabetes. tujuannya untuk mengevaluasi efektivitas relatif dari kedua algoritma tersebut dalam mengidentifikasi pasien yang mungkin menderita diabetes. Hal ini dapat membantu dalam menentukan metode yang lebih unggul atau efisien dalam mendiagnosis penyakit tersebut.

- **Metodologi Penelitian :** *Paper 3* Penelitian ini akan menggunakan data pasien untuk membandingkan dua cara untuk menemukan orang-orang yang mungkin menderita diabetes. Setelah data dipersiapkan, Kemudian kita akan menggunakan dua metode untuk mengklasifikasi penyakit diabetes, algoritma yang di pakai adalah *Support Vector Machine* dan *Modified Balanced Random Forest*. Kedua metode ini akan kita uji dengan memberikan mereka beberapa data dan melihat seberapa akurat dari kedua algoritma tersebut.
- **Temuan Utama :** *Modified Balanced Random Forest* menunjukkan akurasi yang lebih baik dalam mendeteksi pasien penyakit diabetes dibandingkan dengan *Support Vector Machine* (SVM). Meskipun SVM memberikan akurasi yang layak, *Modified Balanced Random Forest* secara signifikan meningkatkan kemampuan dalam memprediksi keberadaan penyakit diabetes.
- **Kesimpulan :** *Modified Balanced Random Forest* (MBRF) lebih efektif daripada *Support Vector Machine* (SVM) dalam mengklasifikasikan data pasien diabetes dari dataset GulaKarya Medika. MBRF menghasilkan akurasi hingga 97,8%, sementara SVM hanya mencapai 91,48%. Studi ini juga menunjukkan bahwa MBRF atas *Random Forest* biasa dan hasil yang lebih baik dibandingkan dengan penelitian sebelumnya yang menggunakan SVM.