

## **BAB. II** **LANDASAN TEORI**

### **2.1. Penyakit Malaria**

Pengendalian malaria memerlukan strategi terpadu yang mencakup berbagai langkah pencegahan dan pengobatan. Beberapa metode pencegahan yang umum dilakukan antara lain penggunaan kelambu berinsektisida, penyemprotan insektisida di dalam ruangan, serta pengelolaan lingkungan untuk mengurangi habitat nyamuk. Pengobatan yang cepat dan tepat dengan obat antimalaria sangat penting guna mencegah komplikasi. Selain itu, pengembangan vaksin juga menjadi salah satu upaya penting, seperti vaksin RTS,S/AS01 yang telah diimplementasikan di beberapa negara Afrika dan terbukti mampu menekan angka kejadian malaria.

Keterkaitan malaria dengan penelitian sangat kuat karena pengendalian penyakit ini memerlukan data yang akurat dan metode analisis yang sesuai. Studi epidemiologi membantu memetakan wilayah dengan risiko penularan tinggi, sehingga langkah intervensi dapat difokuskan pada area prioritas. Penelitian di laboratorium juga berperan dalam mengidentifikasi jenis Plasmodium yang dominan di suatu daerah, sehingga terapi dapat disesuaikan dengan spesies parasit yang ditemukan.

Kemajuan penelitian di bidang teknologi kesehatan turut berkontribusi pada pengembangan sistem deteksi dini dan pemantauan penyebaran malaria. Pendekatan berbasis data modern, seperti pemodelan spasial dan prediksi menggunakan kecerdasan buatan, memungkinkan identifikasi potensi penyebaran

lebih cepat. Inovasi ini mendukung upaya pencegahan secara proaktif sebelum penyakit meluas.

Dengan memadukan strategi pengendalian yang efektif dan dukungan penelitian berkelanjutan, diharapkan angka kesakitan dan kematian akibat malaria dapat ditekan secara signifikan. Penerapan teknologi dan metode ilmiah dalam memprediksi, mendeteksi, serta mengendalikan malaria menjadi kunci keberhasilan untuk mencapai target eliminasi penyakit ini di berbagai negara.

## **2.2. Machine Learning**

Machine Learning adalah cabang dari kecerdasan buatan( Artificial Intelligence) yang berfokus pada pengembangan algoritma dan model yang memungkinkan sistem komputer untuk belajar dari data dan meningkatkan kinerjanya secara otomatis tanpa harus diprogram secara eksplisit. Konsep utamanya adalah membangun model matematis yang dapat mengenali pola dan hubungan di dalam data, sehingga mampu membuat prediksi atau mengambil keputusan berdasarkan informasi baru. Dengan memanfaatkan data dalam jumlah besar, Machine Learning dapat memberikan solusi yang lebih adaptif dan akurat dibandingkan metode pemrograman konvensional.

Dalam penerapannya, Machine Learning dibagi menjadi beberapa jenis utama, seperti supervised learning, unsupervised learning, dan reinforcement learning. Supervised learning melibatkan proses pembelajaran dari data yang sudah memiliki label atau kategori tertentu, sedangkan unsupervised learning digunakan untuk menemukan pola atau struktur tersembunyi pada data yang tidak memiliki label. Sementara itu, reinforcement learning mengandalkan prinsip trial and error,

di mana sistem belajar berdasarkan umpan balik dari tindakan yang dilakukan. Ketiga jenis pembelajaran ini dapat disesuaikan dengan kebutuhan dan tujuan analisis, tergantung pada ketersediaan data dan permasalahan yang ingin diselesaikan.

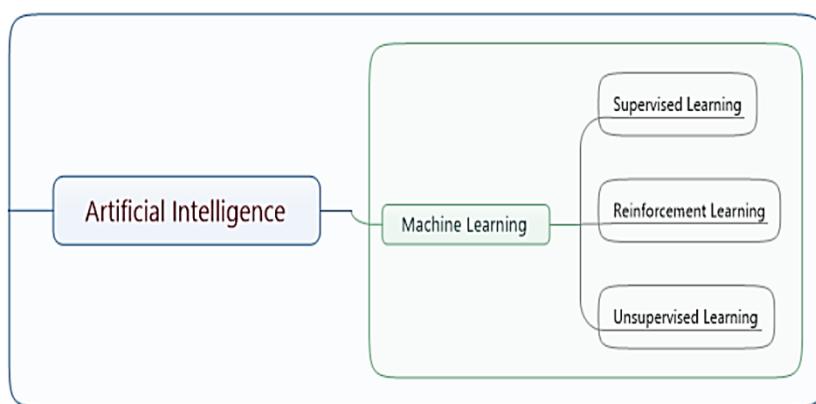
Keunggulan Machine Learning terletak pada kemampuannya dalam mengolah data dalam jumlah besar dengan kecepatan tinggi, sekaligus menyesuaikan model secara dinamis ketika menerima data baru. Hal ini memungkinkan sistem untuk memberikan hasil yang lebih relevan dan presisi, bahkan ketika pola data bersifat kompleks atau terus berubah. Dengan dukungan metode yang beragam, seperti algoritma klasifikasi, regresi, clustering, hingga deep learning, Machine Learning mampu digunakan di berbagai bidang mulai dari pengenalan wajah, analisis sentimen, deteksi penipuan, hingga sistem rekomendasi.

Penerapan Machine Learning juga memberikan peluang besar untuk meningkatkan efisiensi dan efektivitas pengolahan data dalam skala besar. Melalui pemanfaatan model yang tepat, organisasi dapat memperoleh wawasan yang lebih mendalam dari data yang dimilikinya, sehingga membantu dalam proses pengambilan keputusan strategis. Selain itu, kemampuannya untuk melakukan pembelajaran berkelanjutan menjadikan Machine Learning relevan dalam menghadapi tantangan di era big data dan transformasi digital yang menuntut analisis cepat, akurat, dan adaptif.

### **2.2.1. Teknik Machine Learning**

Hasil kajian terbaru menunjukkan bahwa pendekatan dalam machine literacy dapat dikelompokkan menjadi tiga tipe utama, yakni Supervised Learning,

Unsupervised literacy, dan underpinning literacy( 11). Masing- masing tipe memiliki mekanisme pembelajaran dan tujuan yang berbeda, mulai dari pemanfaatan data berlabel, eksplorasi pola tersembunyi, hingga pembelajaran berbasis umpan balik lingkungan. Hubungan hierarkis antara artificial intelligence dan machine literacy, beserta posisi ketiga tipe pembelajaran tersebut, dapat divisualisasikan secara konseptual pada Gambar 2.1.



**Gambar 2. 1. Skema Artificial Intelligence dan Machine Learning [10]**

Supervised Learning merupakan pendekatan pembelajaran mesin di mana data pelatihan telah dilengkapi dengan marker atau target affair yang jelas. Tujuannya adalah untuk mengajarkan model agar dapat memetakan input ke affair yang sesuai berdasarkan contoh- contoh yang diberikan. Teknik ini sangat umum digunakan dalam klasifikasi, yakni mengelompokkan data ke dalam kategori yang telah ditentukan, maupun dalam regresi, yaitu memprediksi nilai kontinu seperti harga, suhu, atau volume penjualan. Dengan memahami hubungan antara variabel input dan affair, model yang dihasilkan mampu melakukan prediksi pada data baru dengan tingkat akurasi tertentu. Contoh algoritma yang kerap digunakan meliputi Decision Tree, Support Vector Machine( SVM), Logistic Retrogression, k- Nearest Neighbor( k- NN), Random Forest, Naïve Bayes, hingga berbagai arsitektur Neural

Network ultramodern. Berbeda dengan supervised literacy, Unsupervised literacy beroperasi tanpa adanya marker atau target affair pada data. Sistem hanya menerima data mentah dan berusaha menemukan pola, keteraturan, atau struktur tersembunyi di dalamnya. Salah satu bentuk umum dari teknik ini adalah clustering, yang berfokus pada pengelompokan data berdasarkan kesamaan karakteristik, seperti pengelompokan pelanggan berdasarkan perilaku belanja. Ada pula metode association rule learning yang bertujuan menemukan hubungan atau aturan yang sering muncul dalam kumpulan data, misalnya kebiasaan konsumen membeli produk tertentu secara bersamaan.

Pendekatan ini banyak dimanfaatkan untuk eksplorasi data dalam tahap awal analisis, di mana pola atau hubungan antarvariabel belum diketahui secara pasti. underpinning literacy menempati posisi unik karena sifatnya yang adaptif terhadap lingkungan dinamis, menggabungkan prinsip trial- and- error dengan umpan balik berupa price atau penalty. Sistem ini tidak memerlukan data berlabel seperti supervised literacy, namun juga tidak sekadar menemukan pola seperti unsupervised literacy. Sebaliknya, agen pembelajaran membuat serangkaian keputusan berdasarkan interaksi langsung dengan lingkungan, menyesuaikan strategi untuk memaksimalkan hasil jangka panjang.

Pendekatan ini sering dimodelkan menggunakan Markov Decision Process (MDP) yang memformulasikan masalah ke dalam status, aksi, transisi, dan imbalan. Beberapa algoritma populer yang digunakan meliputi Q- literacy, SARSA, dan Temporal Difference (TD). Dalam penerapannya, underpinning learning terbagi menjadi dua pendekatan utama. Metode berbasis model terlebih dahulu

mempelajari gambaran lingkungan, lalu menentukan strategi optimal dari informasi tersebut, seperti pada algoritma SARSA yang mempertimbangkan hubungan antara aksi dan keadaan berikutnya. Sementara metode tanpa model (model-free styles), seperti Q- Learning dan algoritma TD, langsung memperbarui kebijakan berdasarkan pengalaman tanpa perlu membangun representasi eksplisit dari lingkungan.

Keunggulan underpinning learning terletak pada kemampuannya belajar secara mandiri dari pengalaman, yang menjadikannya cocok untuk aplikasi seperti robotika, permainan strategi, hingga sistem rekomendasi adaptif, meskipun proses pelatihannya sering kali memerlukan waktu dan sumber daya yang signifikan.

### **2.3. *Naive Bayes***

Naive Bayes adalah algoritma klasifikasi yang berlandaskan pada teorema Bayes dengan asumsi bahwa setiap variabel prediktor bersifat independen satu sama lain. Proses kerjanya dilakukan dengan menghitung probabilitas suatu kelas berdasarkan data yang tersedia, lalu menentukan kelas dengan probabilitas tertinggi sebagai hasil prediksi. Meskipun memiliki konsep yang sederhana, metode ini mampu menghasilkan prediksi yang cukup akurat pada berbagai kasus klasifikasi, khususnya jika data pelatihan berjumlah besar dan distribusinya sesuai dengan asumsi independensi.

Tahapan kerja Naive Bayes dimulai dengan menghitung probabilitas prior untuk setiap kelas, kemudian mengombinasikannya dengan probabilitas likelihood dari masing-masing atribut terhadap kelas tersebut. Asumsi independensi yang digunakan mengartikan bahwa antar atribut tidak saling memengaruhi, sehingga

proses perhitungan menjadi lebih efisien dan cepat. Walaupun dalam praktiknya asumsi ini jarang terpenuhi sepenuhnya pada data nyata, algoritma ini tetap menunjukkan performa yang memuaskan dalam banyak penelitian.

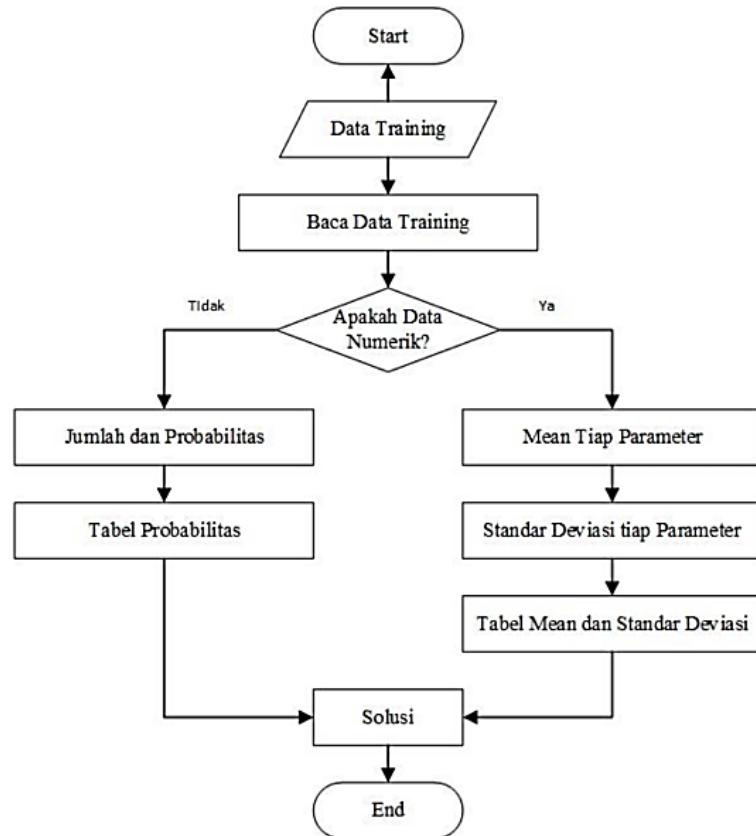
Keunggulan utama Naive Bayes terletak pada kemampuannya menangani dataset berdimensi tinggi, di mana jumlah atribut yang dianalisis cukup banyak. Selain itu, metode ini tidak membutuhkan proses pelatihan yang rumit, sehingga waktu komputasi yang dibutuhkan relatif singkat. Keistimewaan lainnya adalah sifatnya yang tahan terhadap noise pada data, serta fleksibilitasnya dalam mengolah data baik yang bersifat kategorikal maupun numerik melalui teknik estimasi probabilitas yang tepat.

Dalam penerapan analisis data, Naive Bayes sering digunakan untuk mengklasifikasikan data ke dalam kategori tertentu, seperti deteksi spam pada email, analisis sentimen, maupun pengelompokan dokumen berdasarkan topik. Mekanisme probabilistiknya membuat hasil klasifikasi dapat dijelaskan secara matematis, sehingga memudahkan interpretasi. Selain itu, algoritma ini cocok digunakan pada data dengan jumlah sampel besar karena skalabilitasnya yang tinggi.

Penggunaan Naive Bayes juga efektif pada penelitian yang melibatkan variabel-variabel independen yang dapat dihitung probabilitasnya. Dengan proses perhitungan yang relatif sederhana, peneliti dapat memperoleh model prediksi yang dapat digunakan untuk memetakan data baru secara cepat. Algoritma ini mampu memberikan gambaran yang jelas mengenai pengaruh setiap atribut terhadap

kategori yang diprediksi, sehingga membantu dalam proses pengambilan keputusan berbasis data.

Kombinasi antara kemudahan implementasi, efisiensi waktu komputasi, serta tingkat akurasi yang memadai membuat Naive Bayes menjadi salah satu pilihan populer dalam pemodelan klasifikasi. Algoritma ini mampu diaplikasikan pada berbagai domain ilmu, baik di bidang bisnis, kesehatan, pendidikan, maupun teknologi, sehingga menjadi alat analisis yang serbaguna. Dengan memahami konsep dasar dan cara kerja Naive Bayes, peneliti dapat mengoptimalkan penggunaannya untuk menghasilkan model klasifikasi yang andal dan relevan dengan tujuan analisis.



Gambar 2. 2. Alur Kerja Algoritma Naïve Bayes [11]

Dari gambar 2.2 dapat dijelaskan alur kerja dari algoritma Naïve Bayes dijelaskan sebagai berikut [11]:

1. Baca data *training*.
2. Hitung jumlah atau probabilitas, namun apabila data numerik maka:
3. Cari nilai mean dan standar deviasi dari masing-masing parameter yang merupakan data numerik.
4. Cari nilai probabilitas dengan cara menghitung jumlah data yang sesuai dari dari kategori yang sama dibagi dengan jumlah data pada kategori tersebut.
5. Mendapatkan nilai dalam table mean, standar deviasi dan probabilitas.

Kelebihan algoritma naïve bayes adalah data latih yang digunakan sedikit, sehingga perhitungan dapat dilakukan lebih cepat dan efisien. Kelemahan naïve bayes adalah jika terjadi kesalahan dalam pemilihan fitur, akurasinya menurun dan waktu perhitungannya bertambah [12].

#### **2.4. Confusion Matrix**

Confusion Matrix adalah sebuah tabel yang digunakan untuk mengevaluasi kinerja model klasifikasi dengan membandingkan antara hasil prediksi model dan label sebenarnya. Tabel ini memuat informasi tentang jumlah prediksi yang benar maupun salah dalam setiap kategori kelas. Dengan adanya Confusion Matrix, kita dapat mengetahui secara rinci seberapa baik model dalam mengklasifikasikan data ke dalam kelas yang tepat. Evaluasi ini menjadi penting karena tidak hanya menilai akurasi secara keseluruhan, tetapi juga memberikan gambaran mengenai distribusi kesalahan prediksi yang dilakukan oleh model.

Struktur Confusion Matrix umumnya berbentuk persegi dengan jumlah baris dan kolom yang sama, di mana baris mewakili kelas sebenarnya (actual class) dan kolom mewakili kelas hasil prediksi (predicted class). Dalam kasus klasifikasi biner, terdapat empat komponen utama, yaitu True Positive (TP), False Positive (FP), True Negative (TN), dan False Negative (FN). True Positive menunjukkan jumlah data positif yang diprediksi positif secara benar, sedangkan True Negative menunjukkan jumlah data negatif yang diprediksi negatif dengan tepat. Sebaliknya, False Positive adalah data negatif yang salah diprediksi sebagai positif, dan False Negative adalah data positif yang diprediksi negatif secara keliru.

Dari Confusion Matrix, berbagai metrik evaluasi dapat dihitung, seperti akurasi, presisi, recall, dan F1-score. Akurasi mengukur proporsi prediksi yang benar dari seluruh data yang diuji. Presisi mengukur tingkat ketepatan prediksi positif yang dihasilkan model, sedangkan recall mengukur kemampuan model dalam menemukan seluruh data yang benar-benar positif. F1-score adalah rata-rata harmonis antara presisi dan recall, yang berguna ketika keseimbangan antara keduanya diperlukan. Metrik-metrik ini memberikan penilaian yang lebih komprehensif dibandingkan hanya mengandalkan akurasi saja, terutama pada data yang tidak seimbang. Ilustrasi tabel confusion matrix dapat dilihat pada gambar 2.3 berikut.

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

Gambar 2.3. Confusion Matrix [13]

Gambar 2.3 menampilkan bentuk tabel confusion matrix, yaitu representasi yang digunakan untuk mengukur kinerja suatu model klasifikasi. Dalam tabel ini, True Positive (TP) merepresentasikan jumlah data yang secara aktual termasuk dalam kelas positif dan diprediksi sebagai positif oleh model. False Positive (FP) menunjukkan jumlah data yang secara aktual berada pada kelas negatif namun diprediksi sebagai positif. False Negative (FN) adalah jumlah data yang sebenarnya termasuk kelas positif namun diprediksi sebagai negatif. Sementara itu, True Negative (TN) menggambarkan jumlah data yang secara aktual berada pada kelas negatif dan diprediksi negatif oleh model.

Penggunaan Confusion Matrix menjadi sangat bermanfaat ketika ingin memahami kelemahan spesifik dari suatu model. Misalnya, sebuah model mungkin memiliki akurasi yang tinggi, tetapi ternyata banyak melakukan kesalahan pada kelas tertentu. Dengan melihat sebaran TP, FP, TN, dan FN, peneliti dapat mengidentifikasi pola kesalahan tersebut dan menentukan strategi perbaikan, seperti menyeimbangkan data, mengubah parameter model, atau memilih algoritma lain yang lebih sesuai dengan karakteristik data.

Selain itu, Confusion Matrix juga membantu dalam mengevaluasi dampak dari kesalahan prediksi. Dalam beberapa kasus, kesalahan memprediksi kelas positif sebagai negatif atau sebaliknya memiliki konsekuensi yang berbeda. Misalnya, pada sistem deteksi penyakit, kesalahan False Negative bisa lebih berbahaya daripada False Positive. Oleh karena itu, analisis Confusion Matrix tidak hanya fokus pada angka akurasi, tetapi juga mempertimbangkan konteks dan risiko dari kesalahan klasifikasi yang terjadi.

Melalui analisis yang dihasilkan dari Confusion Matrix, peneliti dapat memberikan interpretasi yang lebih mendalam terkait performa model yang dibangun. Hasil ini dapat menjadi dasar untuk membandingkan efektivitas beberapa algoritma klasifikasi, mengidentifikasi kebutuhan optimasi, dan merumuskan rekomendasi perbaikan model di masa mendatang. Dengan demikian, Confusion Matrix tidak hanya berfungsi sebagai alat evaluasi, tetapi juga sebagai panduan strategis dalam pengembangan model yang lebih akurat dan andal.

Confusion matrix berfungsi sebagai alat evaluasi yang mampu memperlihatkan serta membandingkan antara nilai aktual (actual value) dan hasil prediksi model (predicted value). Melalui matriks ini, dapat dihitung sejumlah metrik evaluasi seperti accuracy, precision, recall, dan F1-score [21].

1. Akurasi (Accuracy) mengukur tingkat ketepatan model dalam melakukan klasifikasi, yaitu perbandingan antara jumlah prediksi benar (positif dan negatif) dengan keseluruhan data yang diuji. Rumusnya dituliskan sebagai:

2. Presisi (Precision) merupakan probabilitas bahwa data yang diprediksi sebagai positif benar-benar termasuk dalam kelas positif. Metrik ini mengindikasikan sejauh mana prediksi positif yang dihasilkan model sesuai dengan kondisi sebenarnya. Rumusnya adalah:

3. Daya Ingat (Recall) atau *Sensitivity* adalah probabilitas bahwa data yang sebenarnya positif berhasil diidentifikasi dengan benar sebagai positif oleh model. Metrik ini mengukur kemampuan model dalam menemukan semua data positif yang ada. Rumusnya dituliskan sebagai:

4. F1-Score atau *F-Measure* adalah ukuran harmonisasi antara nilai *precision* dan *recall*, sehingga mampu memberikan gambaran kinerja model yang seimbang dalam menghindari kesalahan tipe I (*false positive*) dan tipe II (*false negative*). Rumusnya sebagai berikut:

## 2.5. Alat Bantu Analisis

### 2.5.1. RapidMiner

RapidMiner merupakan sebuah platform perangkat lunak di bidang *data science* yang dikembangkan oleh perusahaan dengan nama yang sama. Perangkat ini menyediakan lingkungan kerja yang terintegrasi untuk berbagai keperluan, seperti persiapan data, penerapan *machine learning*, *deep learning*, *text mining*, hingga analisis prediktif. Pemanfaatannya cukup luas, mulai dari sektor bisnis dan

komersial, hingga penelitian, pendidikan, pelatihan, *rapid prototyping*, serta pengembangan aplikasi. RapidMiner mendukung seluruh tahapan proses pembelajaran mesin, termasuk tahap pra-pemrosesan data, validasi dan optimasi model, hingga penyajian hasil dalam bentuk visual [23]. Sebagai salah satu perangkat lunak yang banyak digunakan dalam pengolahan *data mining*, RapidMiner juga memiliki modul khusus untuk *text mining*. Modul ini berfungsi untuk menganalisis teks, menemukan pola-pola tersembunyi pada himpunan data yang besar, serta mengombinasikan hasilnya dengan metode statistik, kecerdasan buatan, dan basis data [24].

Awalnya, RapidMiner dikenal dengan nama YALE (*Yet Another Learning Environment*), yang dikembangkan oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Unit Kecerdasan Buatan Universitas Teknik Dortmund pada tahun 2001. Kemudian, pada tahun 2006, pengembangannya dilanjutkan oleh Rapid-I, sebuah perusahaan yang didirikan oleh Ingo Mierswa dan Ralf Klinkenberg. Setahun kemudian, tepatnya pada 2007, nama YALE secara resmi diubah menjadi RapidMiner. Selanjutnya, pada 2013, Rapid-I melakukan *rebranding* menjadi RapidMiner. Saat ini, RapidMiner tidak hanya digunakan di dunia bisnis, tetapi juga di bidang pendidikan, penelitian, pelatihan, dan pembuatan *prototype*. Fungsinya meliputi seluruh proses *machine learning* seperti pra-pemrosesan data, visualisasi, pembuatan model, hingga optimasi [25].



**Gambar 2. 4. Tampilan Aplikasi RapidMiner [14]**

RapidMiner bersifat *open source* dan dirancang sebagai solusi lengkap untuk analisis *data mining*, *text mining*, dan analisis prediktif. Perangkat ini dapat digunakan secara mandiri atau diintegrasikan ke dalam produk lain sebagai *data mining engine*. Dibangun dengan bahasa pemrograman Java, RapidMiner dapat dijalankan di berbagai sistem operasi. Perangkat ini memanfaatkan beragam teknik analisis deskriptif maupun prediktif untuk membantu pengguna memperoleh wawasan yang mendukung pengambilan keputusan terbaik. Secara keseluruhan, RapidMiner menyediakan sekitar 500 *operator* yang mencakup fungsi *input-output*, *data preprocessing*, hingga visualisasi [27].

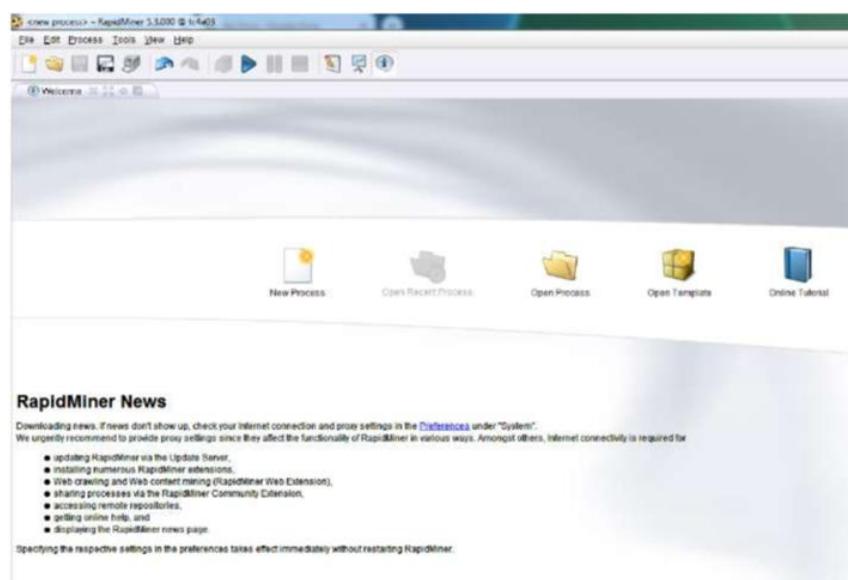
### **2.5.2. Bagian RapidMiner**

RapidMiner memiliki antarmuka pengguna (UI) yang dirancang secara intuitif, sehingga memudahkan pengoperasian bagi penggunanya. Dalam aplikasi ini, antarmuka tersebut dikenal dengan istilah *Perspective*. Secara umum, terdapat

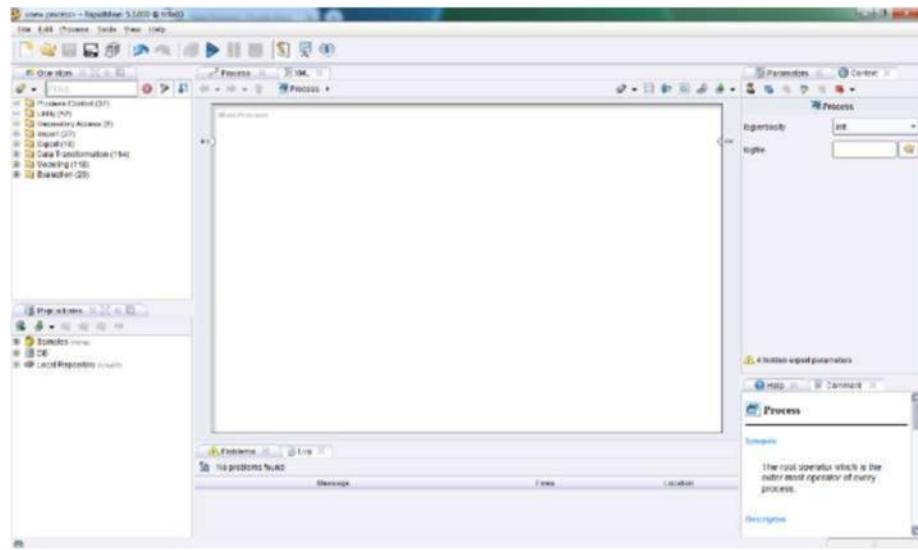
tiga jenis *Perspective* yang tersedia, yaitu Welcome Perspective, Design Perspective, dan Result Perspective [28].

Ketika pertama kali membuka RapidMiner, pengguna akan disambut dengan Welcome Perspective, seperti terlihat pada Gambar 2.5. Selanjutnya, Design Perspective menjadi area utama tempat pengguna melakukan aktivitas analisis di RapidMiner. Pada perspective ini, pengguna dapat merancang sekaligus mengatur alur proses analisis data. Seperti yang diperlihatkan pada Gambar 2.6, *Design Perspective* memuat beberapa *view* yang memiliki fungsi masing-masing untuk mendukung proses analisis.

Sementara itu, Result Perspective merupakan tampilan yang digunakan untuk menyajikan hasil dari proses analisis yang telah dilakukan. Bentuk hasil yang ditampilkan dapat beragam, mulai dari grafik, teks, pohon keputusan (*tree*), tabel, dan format lainnya. Variasi tampilan ini menyesuaikan dengan karakteristik data serta metode analisis yang digunakan.

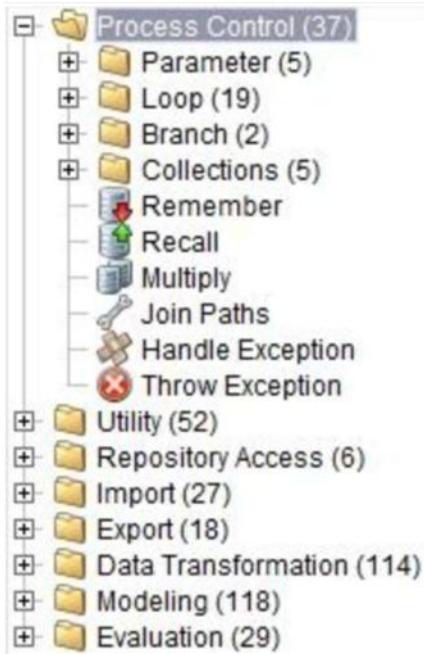


Gambar 2. 5. Welcome Perspective RapidMiner [15]



**Gambar 2. 6. Design Perspective RapidMiner [15]**

Sebagai tampilan untuk lingkungan kerja, Design Perspective memiliki beberapa view. Operator View adalah view yang paling utama dan penting. Semua operator atau langkah-langkah kerja dalam RapidMiner ditampilkan dalam operator view dalam bentuk grup hierarkis, sehingga operator ini dapat digunakan dalam proses analisis, seperti yang ditunjukkan pada Gambar 2.7.



**Gambar 2.7. Operator View RapidMiner [15]**

Berdasarkan Gambar 2.7, operator view terbagi ke dalam beberapa kelompok dengan fungsi yang berbeda-beda. Pertama, terdapat Process Control, yaitu kelompok operator yang berisi logika percabangan dan perulangan (looping) untuk mengatur jalannya alur proses analisis data. Kedua, Utility, yang berperan sebagai kumpulan operator pendukung, misalnya untuk macros, login, subproses, dan fungsi tambahan lainnya. Ketiga, Repository Access, yang digunakan untuk membaca maupun menulis data pada repository. Keempat, Import, yang memiliki fungsi untuk mengambil atau memuat data dan objek dari berbagai sumber, seperti file, basis data, dan media lainnya. Kelima, Export, yang berfungsi kebalikan dari Import, yakni menulis atau menyimpan data ke dalam format tertentu. Selanjutnya adalah Data Transformation, yaitu operator yang digunakan untuk melakukan transformasi baik pada data maupun metadata. Berikutnya, Modelling, yang

memuat berbagai metode dan teknik data mining untuk mengolah data sesuai kebutuhan analisis. Terakhir, Evaluation, yang digunakan untuk mengukur dan menilai kualitas hasil keluaran (output) dari proses yang dijalankan.

Selain operator view, terdapat pula repository view yang memegang peranan penting dalam Design Perspective. Pada bagian ini, pengguna dapat mengelola serta mengatur seluruh rangkaian proses analisis data menjadi suatu proyek yang terstruktur. Secara bersamaan, repository view juga berfungsi sebagai sumber penyimpanan data beserta elemen-elemen terkait metadata. Sementara itu, process view memiliki peran untuk menampilkan tahapan-tahapan tertentu dalam proses analisis sekaligus menyediakan tautan menuju langkah-langkah berikutnya, sehingga memudahkan navigasi dan pengendalian alur kerja analisis data.

## 2.6. Penelitian Terdahulu

Penelitian terdahulu memberikan dasar teoritis yang kuat untuk penelitian baru. Dengan memahami hasil dan temuan dari studi sebelumnya, peneliti dapat membangun kerangka kerja yang lebih baik. Tabel 2.1 di bawah ini menampilkan lima penelitian sebelumnya.

**Tabel 2. 1. Penelitian Terdahulu**

No	Judul	Penulis	Metode dan Data	Hasil
1	Classification of Malaria Complication Using CART (Classification and	(Irmanita et al., 2021)	1. Penelitian ini menggunakan dua metode, yaitu CART ( <i>Classification and</i>	Hasil penelitian menunjukkan bahwa model CART menghasilkan

	and Regression Tree) and Naïve Bayes		<p><i>Regression Tree)</i> dan Naïve Bayes</p> <p>2. Dataset yang digunakan terdiri dari 337 pasien malaria di Nigeria, dengan 17 atribut yang mencakup gejala seperti demam, kedinginan, sakit kepala, diare, dan lainnya.</p>	<p>akurasi tertinggi sebesar 81.2%</p> <p>dalam memprediksi komplikasi malaria, khususnya untuk komplikasi hipoglikemia.</p> <p>Evaluasi dilakukan menggunakan metrik seperti F- score, presisi, dan recall. Hasil evaluasi menunjukkan F- score sebesar 0.529, presisi 0.4, dan recall 0.717.</p> <p>Penelitian ini juga menemukan bahwa penggunaan dataset gejala memberikan akurasi yang lebih tinggi dibandingkan dengan dataset asli yang mencakup semua atribut, karena atribut yang tidak relevan dapat menurunkan akurasi prediksi.</p>
--	---	--	---	---

2	<p>Malaria Disease Prediction and Grading System: A Performance Model of Multinomial Naïve Bayes (MNB) Machine Learning in Nigerian Hospitals</p>	(Atoyebi et al., 2023)	<ol style="list-style-type: none"> <li>Penelitian ini menggunakan Multinomial Naïve Bayes (MNB) sebagai metode utama untuk memprediksi dan mengklasifikasikan penyakit malaria.</li> <li>Data yang digunakan dalam penelitian ini diambil dari dua rumah sakit publik di Abuja, Nigeria, dengan total 2121 instance dan 15 atribut, di mana 13 atribut adalah numerik dan 2 atribut adalah output.</li> </ol>	<p>Hasil dari penelitian menunjukkan bahwa model MNB memiliki akurasi sebesar 97% dalam memprediksi penyakit malaria. MNB lebih efisien dalam mengklasifikasikan data dengan noise dan outlier. Penelitian ini menyimpulkan bahwa model MNB dapat digunakan sebagai alat diagnostik yang efektif untuk mendukung praktisi medis dalam pengambilan keputusan terkait penyakit malaria.</p>
3	<p>Improved Malaria Outbreak Predictive Model Using Naïve Bayes</p>	(Aliyu dan Bada, 2023)	<ol style="list-style-type: none"> <li>Penelitian ini menggunakan dua metode klasifikasi, yaitu Naïve Bayes dan Artificial</li> </ol>	<p>Hasil dari penelitian menunjukkan bahwa model Naïve Bayes memiliki akurasi sebesar</p>

	and Artificial Neural Network	<p>Neural Network (ANN).</p> <p>2. Data yang digunakan dalam penelitian ini dikumpulkan dari Badan Meteorologi Kebbi dan Pusat Medis Federal selama dua musim hujan pada tahun 2014-2015. Data tersebut mencakup variabel meteorologis yang berpengaruh terhadap kejadian malaria.</p>	<p>90%, dengan sensitivitas 92.6% dan spesifisitas 77.7%. Model ini mampu memprediksi kemungkinan terjadinya wabah malaria dengan probabilitas 0.62 untuk adanya wabah. Sementara itu, model ANN menunjukkan hasil yang lebih baik dengan akurasi 98%, sensitivitas 96.4%, dan spesifisitas 100%. Penelitian ini menyimpulkan bahwa kedua model tersebut efektif dalam memprediksi wabah malaria, namun ANN memberikan hasil yang lebih akurat dibandingkan Naïve Bayes.</p>
--	-------------------------------	--	--

4	Classification Of Malaria Types Using Naïve Bayes Classification	(Ariandi et al., 2023)	<ol style="list-style-type: none"> <li>Penelitian ini menggunakan algoritma Naïve Bayes untuk mengklasifikasikan jenis-jenis malaria berdasarkan data yang dikumpulkan.</li> <li>Data yang digunakan dalam penelitian ini mencakup variabel-variabel seperti jenis penemuan, bulan/tahun, jenis kelamin, alamat dusun, jenis parasit, dan berbagai gejala yang dialami pasien. Dataset yang digunakan terdiri dari 166 data poin yang mencakup 10 variabel.</li> </ol>	<p>Hasil dari penelitian menunjukkan bahwa model Naïve Bayes memiliki akurasi yang sangat tinggi dalam mengklasifikasikan jenis malaria. Dalam pengujian model, akurasi tertinggi yang dicapai adalah 99.84% dengan pembagian data 50% untuk pelatihan dan 50% untuk pengujian. Model ini menunjukkan kinerja yang baik, terutama dalam mengidentifikasi kelas Quartana, Tertiana, dan Tropica, dengan nilai precision, recall, dan F1-score yang tinggi. Hasil ini menunjukkan bahwa algoritma</p>
---	--	------------------------	--	---

				Naïve Bayes efektif dalam memprediksi dan mengklasifikasikan jenis-jenis malaria berdasarkan data yang tersedia.
5	Artificial Intelligence System for Malaria Diagnosis	(Barracloog et al., 2024)	<ol style="list-style-type: none"> <li>Penelitian ini menggunakan beberapa metode machine learning, yaitu Artificial Neural Network (ANN), Naïve Bayes, Random Forest, dan Ensemble Methods.</li> <li>Data yang digunakan dalam penelitian ini terdiri dari 4000 responden yang mengisi kuesioner mengenai gejala malaria dan riwayat kesehatan. Data ini mencakup informasi demografis, faktor penghalang akses</li> </ol>	<p>Hasil penelitian menunjukkan bahwa model yang dikembangkan menggunakan algoritma Naïve Bayes mencapai akurasi 100% dalam mengklasifikasikan jenis-jenis malaria, termasuk Malignant, Tertian, Quartan, dan Suspected malaria. Model ini mampu memberikan diagnosis yang cepat dan akurat, dengan waktu pemrosesan yang sangat singkat (sekitar 0.01 detik untuk pengujian).</p>

			<p>layanan kesehatan, serta gejala dan tanda-tanda malaria.</p>	<p>Penelitian ini juga menyoroti pentingnya fitur gejala dan faktor non-gejala dalam meningkatkan akurasi diagnosis malaria.</p>
--	--	--	---	--