BAB II
LANDASAN TEORI

2.1 Tinjauan Pustaka

Tinjauan pustaka adalah kajian terhadap penelitian sebelumnya yang merupakan langkah penting dalam penelitian ini. Proses ini melibatkan peninjauan dan penelusuran hasil penelitian yang telah ada dengan mempertimbangkan keterkaitan terhadap topik yang sedang diteliti. Pada penelitian ini tinjauan pustaka dibagi menjadi 4 tabel dengan bahasan yang berbeda, yaitu membahas prediksi RF dalam mendeteksi penyakit kardiovaskular, perbandingan RF dengan algoritma lain dalam mendeteksi penyakit, perkembangan penelitian RF dalam mendeteksi penyakit, dan penelitian terkait *Shapley Additive Explanations*.

Tabel 2.1 Prediksi RF dalam mendeteksi Penyakit Kardiovaskular

Prediksi Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
1	(Anshori dkk., 2022)	Prediksi Pasien Dengan Penyakit Kardiovaskular Menggunakan Random Forest	Random Forest (RF)	1.Dataset: Data dari sumber public (Mendeley). Atribut penelitian mencakup variabel usia, tekanan darah, kadar kolesterol, dan detak jantung maksimum. 2.Pre-processing Data: Normalisasi min-max 3.Teknik Validasi: Menggunakan cross-validation dengan k-fold = 10 4.Model pembelajaran mesin: mem-bandingkan RF dengan Logistic Regression dan Ensemble Method	Model *Random Forest* memberikan akurasi yang lebih tinggi (98%) dibandingkan dengan model *ensemble method* (96,75%) dan *logistic regression* (91,61%) Menghasilkan nilai *AUC* (Area Under Curve) sebesar 0,9969,

Prediksi Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
					yang menunjukkan bahwa model sangat baik dalam melakukan klasifikasi
2	(Sumwiza dkk, 2023)	Enhanced Cardiovascular Disease Prediction Model Using Random Forest Algorithm	Ensemble learning berbasis Random Forest (RF)	1.Dataset: data dari kaggle 2.*Pre-processing* Data: missing values, penghapusan oulier, seleksi atribut 3.Teknik Validasi: Model dievaluasi menggunakan confusion matrix, yang mengukur (lanjutan) akurasi, presisi, recall, dan F1-score 4.Model Pembelajaran Mesin: Membanding-kan model RF dengan beberapa algoritma	*Random Forest* mencapai akurasi tertinggi sebesar (96%) dan setelah diseleksi atribut jadi (99%). Hasil ini lebih lebih tinggi dibandingkan dengan model lainnya, *KNN (*95%), *Logistic Regression* (87%), dan *SVM* (85%).
3	(Yang dkk, 2020)	Study Of Cardiovascular Disease Prediction Model Based On Random Forest In Eastern China	Random Forest	1.Dataset: 29.930 subjek risiko tinggi PKV dengan beberapa faktor. 2. Identifikasi faktor risiko dan Normalisasi serta pembersihan data 3.Teknik Validasi: menggunakan Area Under the Curve (AUC) 4.Model Pembelajaran Mesin: Membanding-kan beberapa algoritma	*Random Forest* memiliki kinerja terbaik (0,787) lebih tinggi dibandingkan model lain seperti *Naïve Bayes* (0,7074), *Bagged Trees* (0,7448), dan *AdaBoost* (0,7862)

Pada tabel 2.1 terdapat 3 jurnal yang membahas mengenai prediksi RF dalam mendeteksi PKV. Penelitian (Anshori dkk, 2022) mengembangkan model prediksi PKV dengan membandingkan algoritma RF dengan *Logistic Regression* dan *Ensemble Method*. Penelitian ini menggunakan dataset public yang berisi 1000 data pasien dan 14 atribut klinis, seperti usia, tekanan darah, kolesterol, dan detak jantung maksimum. Proses pengolahan data dilakukan dengan normalisasi min-max dan validasi silang 10-fold untuk meningkatkan generalisasi model. Hasil penelitian menunjukkan bahwa model RF berhasil mendapatkan akurasi 98% dan terbukti lebih baik dibanding *Logistic Regression* dan *Ensemble Method*.

Penelitian selanjutnya, (Sumwiza dkk, 2023) mengombinasikan RF dan teknik *feature selection* berbasis korelasi serta data mining, menggunakan dataset dari Kaggle yang diproses menjadi 769 data bersih. Penelitian ini berfokus untuk mendeteksi dini PKV sebagai upaya menurunkan angka kematian di negara berkembang. Model yang dikembangkan diuji menggunakan *confusion matrix*, dan hasilnya menunjukkan bahwa RF menghasilkan akurasi tertinggi sebesar 99%, dibandingkan KNN (95%), SVM (85%), dan *Logistic Regression* (87%).

Sementara itu, (Yang dkk, 2020) melakukan penelitian berskala besar di wilayah timur China menggunakan data dari 29.930 subjek berisiko tinggi PKV. Penelitian ini mengidentifikasi lebih dari 30 faktor risiko signifikan, termasuk usia lanjut, obesitas, dan kebiasaan merokok. Beberapa pendekatan pembelajaran mesin dikembangkan dan dibandingkan, dengan RF menunjukkan performa terbaik, mencapai nilai AUC sebesar 0,787, yang lebih tinggi dibandingkan regresi multivariat (AUC 0,7143), CART, dan *Naïve Bayes.*

Secara keseluruhan, ketiga studi tersebut menunjukkan konsistensi algoritma RF dalam memprediksi risiko PKV dengan akurasi dan performa yang lebih tinggi dibandingkan algoritma lain. Temuan ini menegaskan bahwa RF merupakan pendekatan yang efektif dan dapat diandalkan dalam sistem pendukung keputusan medis untuk deteksi dini PKV.

Tabel 2.2 Perbandingan RF dengan Algoritma ML lainnya

Komparasi Algoritma Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
1	(Pandey dkk, 2025)	A Systematic Review On Cardiovascular Disease	Systematic Literature Review (SLR) terhadap 50	Parameter yang digunakan dalam analisis mencakup berbagai metrik	Random Forest (RF) terbukti lebih efektif dibandingkan

Komparasi Algoritma Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
		Detection And Classification	Jurnal tahun 2021 -2024. dengan PRISMA (Preferred Reporting Items for Systematic Reviews and Met Analyses)	evaluasi kinerja model klasifikasi berdasarkan Akurasi, Presisi, Recall, dan F1-Score Selain itu, parameter lainnya mencakup jumlah sampel, tipe dataset, jenis sinyal, dan algoritma yang digunakan.	algoritma lain karena memiliki kemampuan yang kuat dalam menangani data berdimensi tinggi dan atribut yang saling berinteraksi, serta resisten terhadap *overfitting*.
2	(Azmi dkk, 2022)	A Systematic Review On Machine Learning Approaches For Cardiovascular Disease Prediction Using Medical Big Data	SLR terhadap 41 jurnal tentag algoritma ML untuk prediksi PKV. seperti DT, RF, SVM, KNN, NB, LR,dan Artificial Neural Network untuk Menganalisis big data medis besar dari berbagai sumber.	Parameter dalam penelitian ini meliputi *accuracy*, *precision*, *recall*, dan *F1-score*. Selain itu, beberapa studi juga menggunakan validasi silang (10-fold cross-validation) untuk memastikan konsistensi performa model. Proses *pre-processing*, *feature selection* dan pembagian data menjadi data pelatihandan pengujian (misalnya 80:20 atau 70:30) juga menjadi bagian penting dari metode evaluasi.	RF secara konsisten memiliki performa unggul dibanding algoritma lainnya dalam sebagian besar studi yang dianalisis. RF memiliki tingkat akurasi tinggi hingga 100% pada data pelatihan dan di atas 90% pada data pengujian.
3	(Zobair dkk, 2023)	Systematic Review Of Internet Of Medical Things For Cardiovascular Disease Prevention Among	Systematic Literature Review (SLR) dengan terhadap 32 jurnal mengguna-kan metode PRISMA	1.Parameter biometrik seperti sinyal elektro-kardiogram (ECG), detak jantung, laju pernapasan, tekanan darah	Hasil menunjukkan bahwa wearable IoMT ECG sensors yang dikombinasikan dengan algoritma AI mampu memberikan

Komparasi Algoritma Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
		Australian First Nations	(Preferred Reporting Items for Systematic Reviews and Meta-Analyses).	(BP), kadar oksigen (SpO2). 2.Parameter gaya hidup dan kondisi fisiologis lainnya.	prediksi PKV dengan akurasi tinggi, bahkan mencapai 99.56% dalam beberapa model. RF dinilai lebih unggul dalam beberapa penelitian yang direview karena kemampuannya dalam menangani data besar, mengenali pola kompleks, serta memberikan hasil klasifikasi yang stabil dan akurat dibandingkan algoritma lain seperti ANN, KNN, dan SVM.
4	(Haganta Depari dkk, 2022)	Perbandingan Model Decision Tree, Naive Bayes dan Random Forest untuk Prediksi Klasifikasi Penyakit Jantung	Decision Tree (DT), Naïve Bayes (NB), dan Random Forest (RF)	1.Dataset: data penyakit jantung dengan berbagai parameter medis (usia, tekanan darah, kadar kolesterol, detak jantung) 2.Pre-processing Data: *missing values,* normalisasi data, pembagian dataset 3.Teknik Validasi: untuk menguji akurasi model, dataset dibagi 2, data training dan data uji. 4.Model Pembelajaran Mesin: DT dan NB	*Random Forest* memiliki kinerja terbaik (0,787) lebih tinggi dibandingkan model lain seperti *NB* (0,7074), *DT* (0,7448), dan *NB* (0,7862)

Komparasi Algoritma Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
5	(Azizan dkk, 2023)	Analisis Perbandingan Metode Klasifikasi Menggunakan Regresi Logistik Biner, Random Forest, dan Support Vector Machine pada CVD Dataset	Regresi Logistik Biner, Random Forest, Support Vector Machine.	1.Dataset: Berasal dari aggle yang berisi 70.000 data pasien yang terdiri dari 13 variabel. 2.Pre-processing Data: pengecekan *missing values,* deteksi *outlier,* standard-risasi data, spil dataset dan *feature selection* 3.Teknik Validasi: *tuning hyper-parameter* (Grid Search CV)	RF memiliki performa terbaik dengan akurasi 73,2%, sensitivitas 66,7%, dan AUC 80,3%.

Pada tabel 2.2 terdapat 5 jurnal yang membahas mengenai perbandingan RF dengan algoritma lainnya. Penelitian oleh (Pandey dkk, 2025) melakukan SLR terhadap 50 jurnal dari tahun 2021–2024 terkait deteksi dan klasifikasi PKV menggunakan berbagai pendekatan kecerdasan buatan, termasuk *machine learning*, *deep learning*, dan metode hibrida. RF diidentifikasi sebagai algoritma yang unggul karena kemampuannya menangani dataset dengan banyak atribut, toleransi terhadap *overfitting*, serta kinerjanya yang konsisten tinggi dalam metrik evaluasi seperti akurasi, *precision, recall*, dan F1-*score*.

Demikian pula penelitian yang dilakukan (Azmi dkk, 2022) menganalisis 41 artikel penelitian berbasis Big Data medis, membandingkan algoritma seperti RF, *Decision Tree, SVM, KNN, Naïve Bayes, Logistic Regression*, dan Artificial Neural Network. Hasil menunjukkan bahwa RF memiliki akurasi pelatihan hingga 100% dan akurasi pengujian 97,29%, menjadikannya metode dengan performa tertinggi dalam sebagian besar studi.

Lebih lanjut, penelitian (Zobair dkk, 2023) dalam SLR terhadap 32 jurnal juga menemukan bahwa RF secara konsisten unggul dibandingkan algoritma lain seperti *Decision Tree, KNN, dan Naïve Bayes*. Penelitian ini menekankan efektivitas RF dalam menangani data tidak seimbang dan kompleks, serta kemampuannya mengidentifikasi atribut prediktif penting yang berkaitan dengan gaya hidup dan variabel fisiologis seperti usia, tekanan darah, kolesterol, dan diabetes.

Penelitian eksperimental oleh (Haganta dkk, 2022) juga memperkuat keunggulan RF melalui perbandingan langsung dengan *Decision Tree* dan Naïve Bayes dalam klasifikasi

pasien jantung. Setelah dilakukan *preprocessing* data dan evaluasi model, hasil menunjukkan bahwa RF memiliki akurasi tertinggi, menegaskan bahwa algoritma ini lebih efektif dan andal dalam klasifikasi kondisi jantung.

Terakhir, penelitian (Azizan dkk, 2023) yang membandingkan RF dengan Logistic Regression dan SVM menggunakan dataset besar dari Kaggle berisi 70.000 data pasien. Dengan menerapkan preprocessing menyeluruh dan feature selection berbasis ANOVA serta Chi-Square, RF tetap menunjukkan performansi terbaik dengan akurasi 73,2%, sensitivitas 66,7%, dan AUC 80,3%. Ini menegaskan bahwa RF memiliki keseimbangan yang baik antara deteksi positif dan kemampuan klasifikasi keseluruhan.

Berdasarkan kelima studi tersebut, dapat disimpulkan bahwa *Random Forest* secara konsisten menunjukkan performa terbaik dibandingkan algoritma ML lainnya dalam konteks prediksi PKV. Keunggulan RF dalam menangani data besar, kompleks, dan tidak seimbang menjadikannya metode yang sangat direkomendasikan dalam pengembangan sistem deteksi dini berbasis data medis.

Tabel 2.3 Perkembangan Penelitian RF Dalam Mendeteksi Penyakit

Perkembangan Penelitian Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
1	(Dadiyala dkk, 2024)	Progressive Heart Disease Prediction Model Using Machine Learning: A Comprehensive Staging Approach	Menerapkan 3 algoritma ML, yaitu Decision Tree , KNN , dan Random Forest untuk memprediksi penyakit jantung, setelah itu akurasi setiap model dievaluasi untuk menentukan algoritma yang paling efektif.	Model dikembangkan menggunakan 11 atribut, yaitu usia, jenis kelamin, tipe nyeri dada, tekanan darah saat istirahat, kadar kolesterol, gula darah puasa, hasil ECG saat istirahat, denyut jantung maksimum, angina yang dipicu oleh olahraga, ST depression (oldpeak), dan slope segmen ST.	Hasil menunjukkan bahwa Random Forest adalah algoritma dengan performa terbaik untuk prediksi penyakit jantung, dengan akurasi sebesar 87.12%. Untuk klasifikasi tahap keparahan penyakit (staging), Decision Tree menunjukkan akurasi tertinggi yaitu 98%.
2	(Alqahtani dkk, 2022)	Cardiovascular Disease	Ensemble learning	1.Dataset: data dari kaggle,	Model RF memiliki akurasi

Perkembangan Penelitian Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
		Detection using Ensemble Learning	berbasis kecerdasan buatan (AI) yang menggabungkan beberapa algoritma machine learning dan deep learning untuk klasifikasi.	yang berisi 70.000 data pasien. 2.Preprocessing Data: missing values, penghapusan oulier, seleksi atribut 3.Teknik Validasi: menggunakan akurasi dan AUC-ROC sebagai metrik utama untuk menilai performa model klasifikasi. 4.Model Pembelajaran Mesin: Algoritma yang digunakan RF, KNN, DT, XGB, DNN, KDNN.	terbaik di antara model pembelajaran mesin (88,65%). Model ensembel gabungan pembelajaran mesin (88,70%) dengan skor ROC-AUC (93%) yang menunjukkan hasil prediksi paling baik.
3	(Reddy dkk, 2023)	Machine Learning based Mobile App for Heart Disease Prediction	Membangun model prediksi penyakit jantung yang diimplemen-tasikan dalam bentuk aplikasi mobile. Tiga algoritma diterapkan dan dibandingkan, yaitu *LR, Artificial Neural Network, Multilayer Perceptron,* dan *RF.*	Prediksi dilakukan berdasarkan beberapa parameter kesehatan utama yang berperan dalam risiko penyakit jantung. Di antaranya adalah usia, jenis kelamin, tekanan darah, cerebral palsy, kadar gula darah puasa, dan indikator lainnya yang umum digunakan	Dari ketiga algoritma yang diuji, RF menghasilkan performa terbaik dalam hal akurasi prediksi, sehingga dipilih sebagai algoritma utama dalam aplikasi. Model ini berhasil meng-identifikasi apakah seseorang memiliki penyakit jantung atau tidak. Aplikasi yang dibuat bertujuan untuk meminimalkan biaya diagnosis serta mengurangi

Perkembangan Penelitian Random Forest					
No	Penulis	Judul	Metode	Parameter	Hasil
				dalam diagnosis kardiovaskular.	subjektivitas manusia dalam penilaian medis awal.

Pada Tabel 2.3 terdapat 3 jurnal mengenai perkembangan penelitian RF dalam mendeteksi PKV. Seperti yang telah di jelaskan pada ke tabel 2.1 dan tabel 2.2, algoritma RF telah diteliti secara luas dalam mendeteksi PKV, hal ini menunjukkan potensi signifikan dalam meningkatkan diagnosis dini dan penilaian risiko. Dari segi performa dan akurasi, pada penelitian (Dadiyala dkk, 2024) menunjukkan bahwa RF memiliki performa klasifikasi tertinggi dibandingkan dengan algoritma lain seperti *Decision Tree dan K-Nearest Neighbor (KNN),* dengan akurasi mencapai 87,12% dalam mendeteksi risiko PKV berdasarkan 11 parameter medis. Keunggulan RF terletak pada kemampuannya menangani data multivariabel yang kompleks serta mengurangi *overfitting* melalui pendekatan *ensemble learning*. Hasil yang stabil dan akurat membuat algoritma ini cocok untuk implementasi di dunia nyata dalam sistem deteksi dini penyakit jantung.

Dari segi model optimasi dan *ensemble methods*, penelitian (Alqahtani dkk, 2022) menggunakan pendekatan *ensemble learning* dengan kombinasi beberapa algoritma *machine learning* dan *deep learning* pada dataset besar berisi 70.000 data pasien dari Kaggle. RF berperan ganda sebagai model klasifikasi dan alat seleksi atribut, menunjukkan performa tertinggi di antara algoritma lain seperti KNN, *Decision Tree, Extreme Gradient Boosting (XGB), dan Deep Neural Networks (DNN).* RF berhasil mencapai akurasi 88,65% dan AUC-ROC sebesar 92%, memperkuat posisinya sebagai algoritma paling andal dalam prediksi PKV berbasis data medis berskala besar.

Dari segi aplikasi dan implementasi, ditunjukkan oleh (Reddy dkk, 2023) yang mengembangkan aplikasi mobile berbasis machine learning untuk deteksi dini penyakit jantung. Aplikasi ini dibangun menggunakan MIT App *Inventor* dan terintegrasi dengan *Firebase*, memanfaatkan RF untuk menganalisis data kesehatan seperti usia, tekanan darah, dan gula darah puasa. Aplikasi ini tidak hanya memberikan diagnosis instan, tetapi juga memungkinkan pengguna menyimpan dan membagikan data medis, menjadikannya alat yang mudah diakses, hemat biaya, dan efektif, terutama di wilayah dengan keterbatasan fasilitas kesehatan.

Berdasarkan penelitian terdahulu yang sudah dijabarkan pada penjelasan diatas, banyak penelitian yang membahas mengenai sistem prediksi penyakit kardiovaskular

menggunakan RF karena RF merupakan model *machine learning* memberikan akurasi yang cukup baik jika dibandingkan dengan metode sederhana lainnya. Namun demikian, masih belum ada penelitian prediksi risiko PKV dengan keterbatasan pada data dan memanfaatkan integrasi data klinis multisumber. Oleh karena itu, penelitian ini diarahkan untuk mengembangkan model prediksi PKV berbasis Random Forest yang tidak hanya akurat, tetapi juga interpretatif melalui analisis SHAP, sehingga mampu memberikan pemahaman yang lebih mendalam terhadap kontribusi setiap variabel dalam pengambilan keputusan dan mendukung penerapan model secara lebih transparan dalam konteks klinis

Tabel 2.4 Penelitian Terkait Shapley Additive Explanations (SHAP)

No	Penulis	Judul	Hasil SHAP
1	(Miranda dkk, 2023)	Understanding Arteriosclerotic Heart Disease Patients Using Electronic Health Records: A Machine Learning and Shapley Additive Explanations Approach	Hasil penelitian menunjukkan hemoglobin merupakan atribut paling berpengaruh dalam mendeteksi pasien dengan penyakit jantung aterosklerotik. Hasil penelitian ini mampu menjembatani kesenjangan antara performa.
2	(Vimbi dkk, 2024)	Interpreting Artificial Intelligence Models: A Systematic Review On The Application Of LIME And SHAP In Alzheimer's Disease Detection	Hasil penelitian menunjukkan menunjukkan bahwa SHAP lebih unggul dalam menunjukkan kontribusi setiap atribut secara konsisten di seluruh dataset dan dapat disimpulkan bahwa SHAP dianggap lebih informatif dibandingkan LIME.
3	(Tusongtuoheti dkk, 2024)	Predicting the risk of subclinical atherosclerosis based on interpretable machine models in a Chinese T2DM population	Hasil penelitian menunjukkan bahwa usia, albumin, total protein, kolesterol total, dan kreatinin serum merupakan lima variabel yang paling berpengaruh terhadap prediksi risiko SCAS.
4	(Choi dkk, 2024)	Explainable Model Using Shapley Additive Explanations Approach on Wound Infection after Wide Soft Tissue Sarcoma Resection: "Big Data" Analysis Based on Health Insurance Review and Assessment Service Hub	Hasil penelitian menunjukkan bahwa transfusi darah, usia lanjut, dan status sosial ekonomi memiliki kontribusi paling besar terhadap risiko infeksi, yang divisualisasikan melalui *dependence plot* menggunakan nilai SHAP

Pada tabel 2.4 terdapat 4 jurnal yang membahas mengenai penelitia terkait SHAP. Penelitian (Miranda dkk, 2023) menggunakan SHAP untuk menganalisis kontribusi

beberapa atribut hematologi seperti hemoglobin, hematokrit, dan leukosit terhadap hasil prediksi model penyakit jantung aterosklerotik menggunakan data rekam medis elektronik di Indonesia. Hasil penelitian menunjukkan bahwa hemoglobin merupakan atribut paling berpengaruh dalam mendeteksi pasien dengan penyakit jantung aterosklerotik. Hasil penelitian ini mampu menjembatani kesenjangan antara performa model yang tinggi dan kebutuhan akan interpretasi yang jelas di lingkungan klinis.

Penelitian selanjutnya, (Vimbi dkk, 2024) melakukan penelitian menggunakan dua metode XAI, yaitu LIME dan SHAP untuk menjelaskan hasil prediksi penyakit Alzheimer. Hasilnya adalah LIME mampu menyoroti atribut pasien seperti *cognitive scores*, citra MRI lokal, dan biomarker tertentu, namun hasilnya cenderung bervariasi dan kurang stabil antar percobaan. Sementara itu, SHAP mengidentifikasi atribut penting seperti *volume hippocampus, atrofi kortikal*, atau nilai *kognitif* tertentu yang konsisten memengaruhi prediksi Alzheimer dan memberikan penjelasan yang lebih komprehensif. Berdasarkan kedua hasil tersebut, SHAP lebih unggul dalam menunjukkan kontribusi setiap atribut secara konsisten di seluruh dataset dan dapat disimpulkan bahwa SHAP dianggap lebih informatif dibandingkan LIME. Peneliti (Tusongtuoheti dkk, 2024) melakukan penelitian mengenai model prediksi berbasis *machine learning* pada populasi pasien diabetes tipe 2 di Tiongkok. Melalui hasil SHAP, peneliti dapat mengetahui bahwa usia, albumin, total protein, kolesterol total, dan kreatinin serum merupakan lima variabel yang paling berpengaruh terhadap prediksi risiko SCAS.

Penelitian yang terakhir (Choi dkk, 2024) menggunakan SHAP untuk memahami hubungan kompleks antar variabel dalam model pembelajaran mesin. SHAP berfungsi untuk menghitung pengaruh masing-masing atribut terhadap kemungkinan infeksi luka pascaoperasi. Hasil penelitian menunjukkan bahwa transfusi darah, usia lanjut, dan status sosial ekonomi memiliki kontribusi paling besar terhadap risiko infeksi, yang divisualisasikan melalui *dependence plot* menggunakan nilai SHAP.

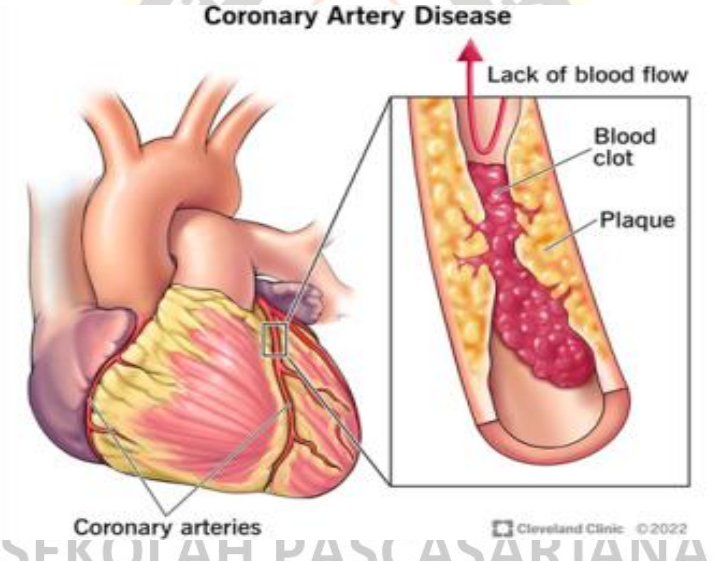
Secara keseluruhan, SHAP bertujuan untuk memberikan penjelasan yang konsisten dan transparan mengenai kontribusi suatu atribut (Alkhanbouli dkk, 2025). Melalui SHAP, peneliti dan tenaga medis dapat memahami faktor mana yang paling berpengaruh terhadap hasil prediksi, misalnya usia, tekanan darah, kadar kolesterol, atau kebiasaan merokok dalam kasus PKV. Dengan demikian, SHAP membantu menjelaskan mengapa suatu pasien diprediksi memiliki risiko penyakit tertentu.

2.2 Dasar Teori

2.2.1 Penyakit Kardiovaskular (*Cardiovascular Disease*)

Penyakit kardiovaskular (PKV) adalah penyakit yang menyerang jantung dan pembuluh darah. Berdasarkan data dari *World Health Organization* (WHO) diperkirakan dalam setahun sebanyak 17,9 juta orang meninggal akibat PKV. Dari jumlah tersebut, 85% disebabkan oleh serangan jantung dan stroke (WHO, 2021) yang terjadi akibat penyumbatan pembuluh darah yang menghambat aliran darah ke organ vital seperti jantung dan otak (Teo dan Rafiq, 2021).

Salah satu contoh PKV adalah penyakit jantung koroner, yaitu kondisi dimana pembuluh darah koroner yang berfungsi menyuplai oksigen dan nutrisi ke otot jantung mengalami penyempitan atau penyumbatan akibat penumpukan plak lemak (*aterosklerosis*) (Bill dan Lakshmi, 2024). Visualisasi penyakit jantung koroner dapat dilihat pada Gambar 2.1



Gambar 2.1 Penyakit Jantung Koroner

Pada Gambar 2.1 *aterosklerosis* menyebabkan aliran darah ke jantung terganggu sehingga dapat menurunkan fungsi jantung dan meningkatkan risiko terjadinya nyeri dada berlebihan yang akan menyebabkan serangan jantung (Bill dan Lakshmi, 2024). Selain penyakit jantung koroner, contoh lain dari PKV adalah penyakit pembuluh darah otak atau stroke, penyakit jantung bawaan, dan gagal jantung (Christina Magnussen dkk., 2023).

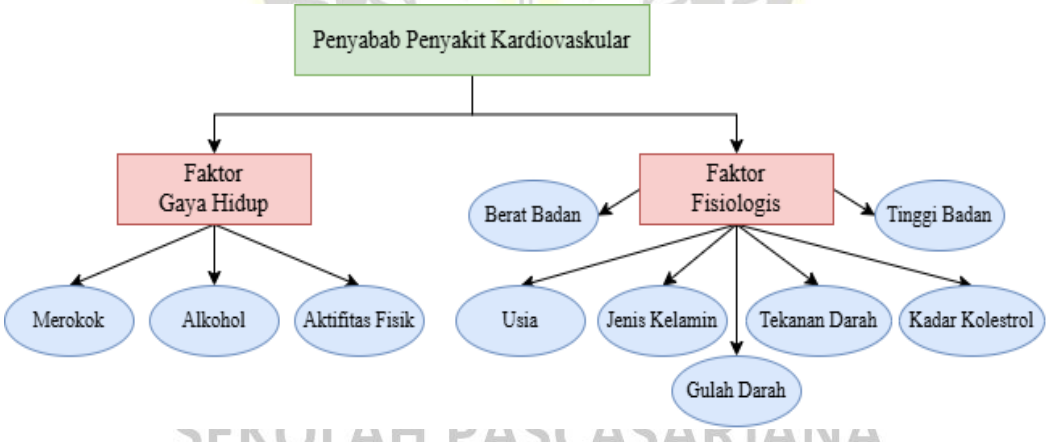
Penyebab utama PKV adalah gaya hidup yang tidak sehat, seperti merokok, sering meminum minuman beralkohol, serta kurangnya aktivitas fisik (Ciumărnean dkk., 2022). Selain gaya hidup, faktor fisiologis juga memiliki pengaruh besar terhadap perkembangan PKV, seperti hipertensi (tekanan darah tinggi) yang dapat menyebabkan kerusakan

pembuluh darah dalam jangka panjang, meningkatkan risiko serangan jantung dan stroke (Teo dan Rafiq, 2021). Selain itu, risiko terkena PKV juga akan meningkat seiring bertambahnya usia, dengan peningkatan risiko mencapai lebih dari 75% pada usia 60-80 tahun (Ciumărnean dkk., 2022).

Mengingat dampak besar yang ditimbulkan oleh PKV, salah satu hal yang sangat penting adalah pencegahan. Pencegahan PKV dapat dilakukan dengan menerapkan pola hidup sehat, seperti menjaga pola makan seimbang, rutin berolahraga, berhenti merokok, serta mengontrol tekanan darah, kadar gula darah, dan kadar kolesterol secara rutin (Bill dan Lakshmi, 2024).

2.2.2 Parameter Gaya Hidup dan Fisiologis

Faktor risiko PKV terbagi menjadi dua kategori, yaitu parameter gaya hidup (Kaminsky dkk., 2022) dan parameter fisiologis (Ciumărnean dkk., 2022). Kedua faktor risiko ini berperan besar dalam menentukan kemungkinan seseorang memiliki PKV atau tidak (Ciumărnean dkk., 2022). Visualisasi kategori faktor risiko PKV dapat dilihat pada Gambar 2.2



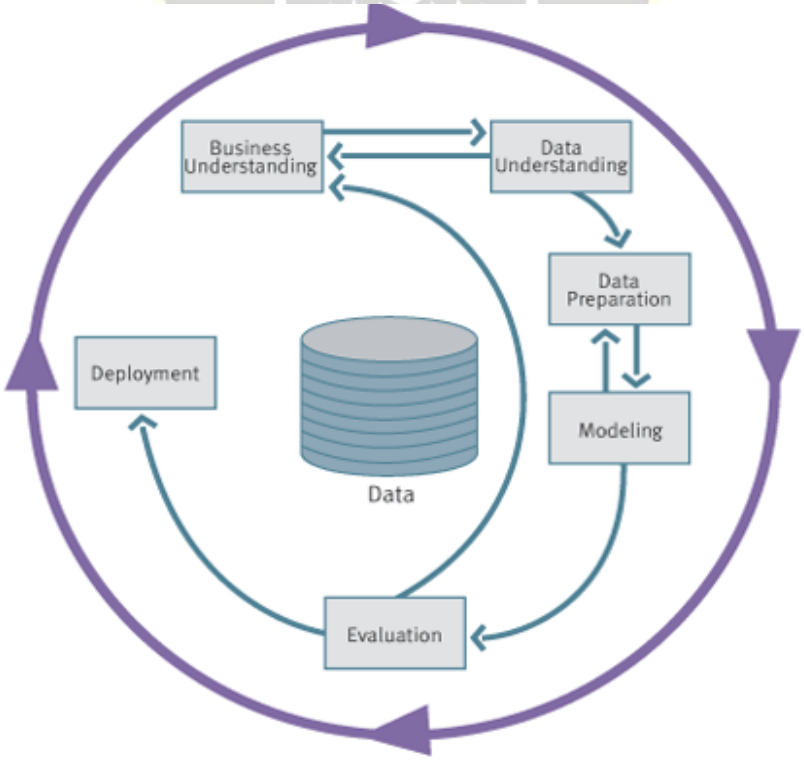
Gambar 2.2 Faktor Risiko Penyakit Kardiovaskular

Penyebab PKV dengan kategori gaya hidup adalah kurangnya aktivitas fisik, mengkonsumsi alkohol dan merokok. Aktivitas fisik yang kurang baik dapat menyebabkan otot jantung melemah dan rentan menyebabkan obesitas, sementara aktivitas fisik yang berlebihan juga dapat menyebabkan serangan jantung. Selanjutnya merokok, merokok dapat menyebabkan penyempitan pembuluh darah dan meningkatkan risiko hipertensi serta pembentukan gumpalan darah, yang berpotensi menyebabkan serangan jantung dan stroke. Sementara itu, mengkonsumsi alkohol dalam jumlah berlebihan dapat memicu tekanan darah tinggi, meningkatkan kadar trigliserida, dan mempercepat perkembangan penyakit jantung (Kaminsky dkk., 2022).

Selain faktor gaya hidup, parameter fisiologis juga digunakan untuk memprediksi risiko PKV pada penelitian ini. Kategori parameter fisiologis diantaranya, usia, jenis kelamin, tekanan darah, kadar kolesterol, gula darah, berat dan tinggi badan. Tekanan darah tinggi dapat menyebabkan kerusakan pada dinding pembuluh darah yang mengakibatkan komplikasi serius, seperti serangan jantung dan gagal jantung. Kadar kolesterol dalam darah juga menjadi indikator penting dalam prediksi risiko PKV. Selain itu, gula darah juga berpengaruh karena kadar gula darah yang tinggi dapat merusak pembuluh darah dan meningkatkan proses *aterosklerosis* (Ciumărnean dkk., 2022).

2.2.3 CRISP-DM

CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan standar proses umum yang digunakan dalam penyelesaian masalah bisnis maupun penelitian melalui pendekatan *data mining*. *Data mining* adalah proses untuk mengidentifikasi pola dan tren yang bermanfaat dari kumpulan data berukuran besar. Dalam penerapannya, para analis data sering kali harus mengulang tahap pemahaman dan pengolahan data, sehingga menyulitkan untuk melangkah ke tahap berikutnya. Oleh karena itu, CRISP-DM berperan sebagai panduan yang membantu menjaga proses data mining agar tetap terstruktur dan terarah dalam mencapai tujuan penelitian (Larose and Larose 2014). Siklus CRISP-DM dapat dilihat pada Gambar 2.3.



Gambar 2.3 Siklus Hidup CRISP-DM

a) *Business Understanding*

Tahap ini bertujuan untuk memahami secara mendalam tujuan dari penelitian. Di tahap ini dilakukan penentuan tujuan bisnis, kriteria keberhasilan, penilaian situasi, identifikasi sumber daya, risiko, serta penyusunan rencana proyek dan tujuan data mining.

b) *Data Understanding*

Pada tahap ini, dilakukan pengumpulan data awal dan melakukan eksplorasi data. Tujuan utamanya adalah memastikan data yang tersedia relevan dan cukup berkualitas untuk dianalisis lebih lanjut.

c) *Data Preparation*

Tahapan ini bertujuan untuk menyiapkan data agar siap digunakan pada tahap modeling. Contohnya dalam penelitian ini, proses ini mencakup penghapusan kolom yang tidak relevan, transformasi data (seperti konversi umur dan gender), tekanan darah, penanganan outlier dan missing values, serta normalisasi menggunakan *Min-Max Scaler*. normalisasi jika perlu, serta rekayasa atribut tambahan seperti BMI.

d) *Modeling*

Pada tahapan ini dilakukan pemilihan teknik pemodelan yang sesuai dan membangun model berdasarkan data yang telah dipersiapkan. Parameter model diatur dan pengujian model dilakukan. Contohnya pada penelitian ini, pengembangan model menggunakan algoritma RF. Beberapa Eksperimen eksperimen dijalankan, termasuk *baseline*, seleksi atribut, balancing data dan optimasi model dengan teknik *hyperparameter tuning* (*Random Search*) dan validasi menggunakan *k-fold cross-validation* untuk menghindari overfitting.

e) *Evaluation*

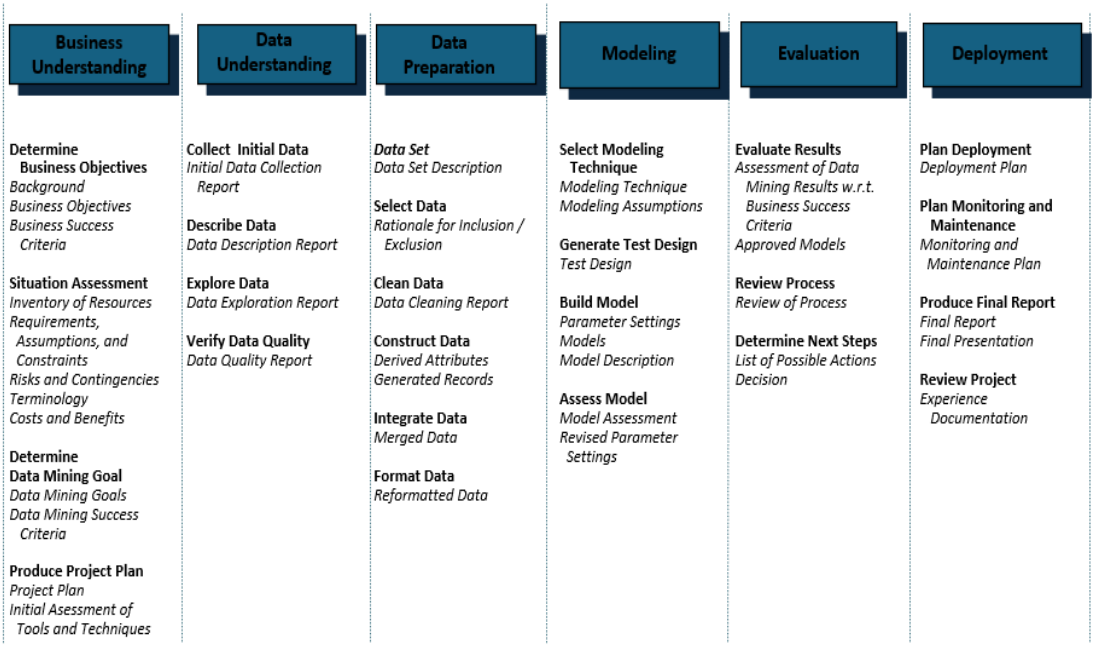
Setelah model dibangun, perlu dilakukan evaluasi secara menyeluruh untuk memastikan model telah memenuhi tujuan bisnis yang ditetapkan di awal. Contohnya pada penelitian ini, model dievaluasi menggunakan metrik seperti *accuracy, precision, recall, F1-score*, dan ROC-AUC. Evaluasi dilakukan melalui *cross-validation* pada data public serta validasi eksternal menggunakan data private untuk mengukur generalisasi model.

f) *Deployment*

Tahap akhir adalah penerapan hasil analisis ke dalam praktik nyata. Bentuk deployment bisa sangat beragam, mulai dari sekadar pembuatan laporan hingga pengembangan

sistem otomatis yang dapat digunakan secara berkelanjutan. Misalnya implementasi model ke halaman web secara real-time.

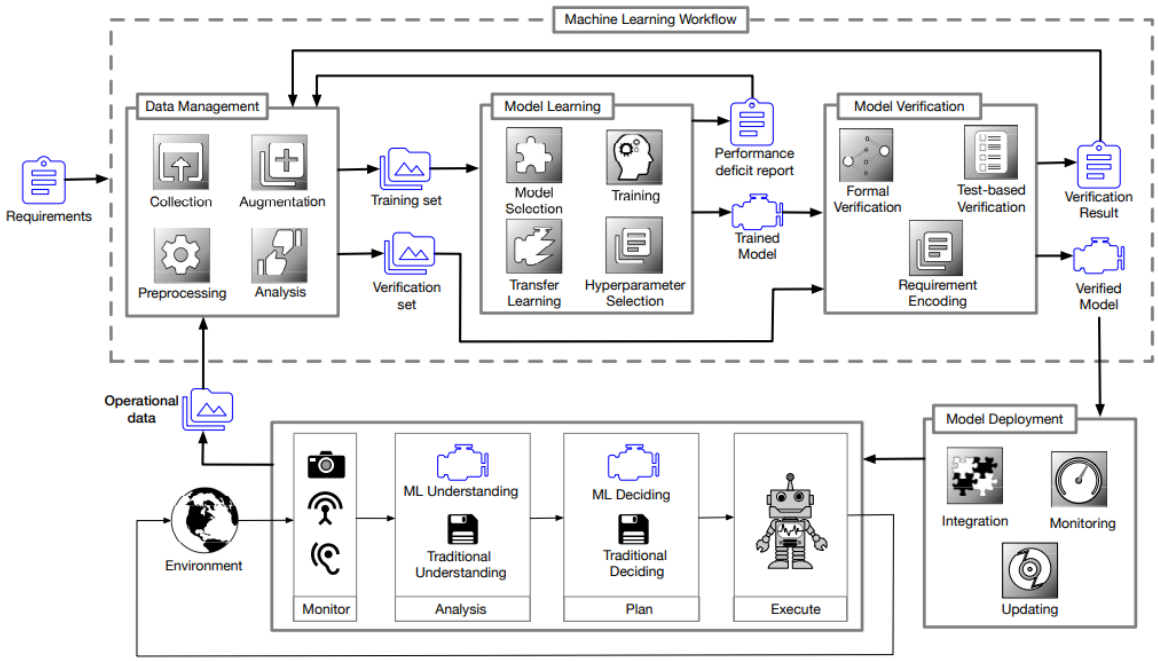
Agar lebih jelas penjabaran tiap fase CRISP-DM bisa dilihat pada Gambar 2.4



Gambar 2.4 Penjabaran Elemen Tiap Fase CRISP-DM

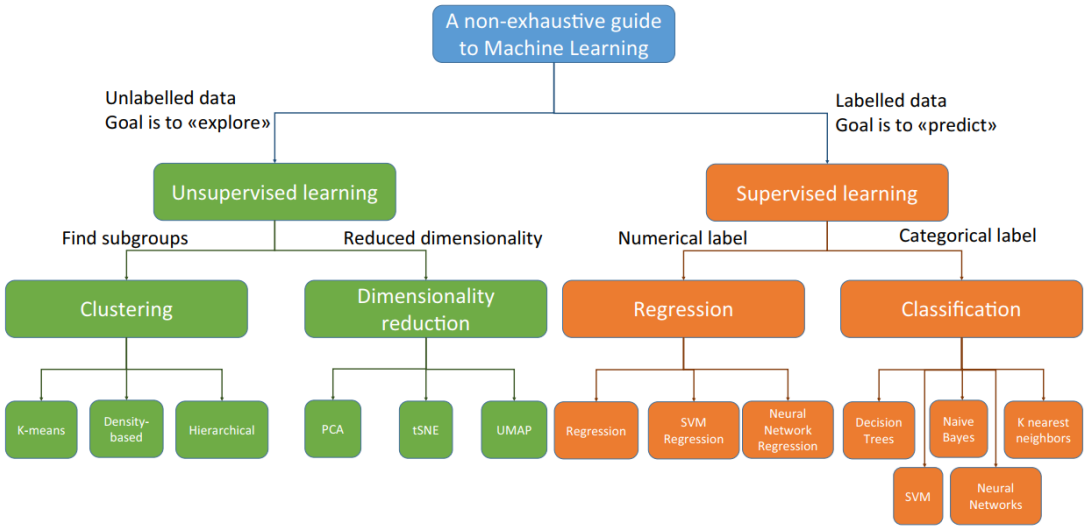
2.2.4 Machine Learning

Machine learning (ML) adalah cabang kecerdasan buatan yang memungkinkan komputer untuk belajar dari data dan membuat prediksi atau keputusan tanpa diprogram secara eksplisit (Balakrishnan dkk, 2021). Dalam penerapannya, proses pengembangan model ML menerapkan suatu siklus (*lifecycle*) yang terdiri dari empat tahap utama, yaitu *Data Management*, *Model Learning*, *Model Verification*, dan *Model Deployment* (Ashmore dkk, 2019). Tahap *data management* berfokus pada pengumpulan, augmentasi, pra-pemrosesan, dan analisis data untuk menghasilkan data training dan data verifikasi yang berkualitas. Kemudian tahap selanjutnya adalah *model learning, model learning* yang berfokus untuk pemilihan jenis model, *training*, *transfer learning* serta proses pelatihan untuk menghasilkan model yang optimal. Setelah itu, masuk ke tahap *model verification* yang bertujuan untuk memastikan model yang dihasilkan telah memenuhi kebutuhan dan persyaratan melalui proses verifikasi formal maupun pengujian berbasis data. Tahap terakhir adalah *model deployment*, yaitu proses mengintegrasikan model ke dalam sistem operasional, disertai dengan pemantauan dan pembaruan model agar tetap andal ketika digunakan di lingkungan nyata. Visualisasi ML *lifecycle* dapat dilihat pada Gambar 2.5.



Gambar 2.5 Machine Learning Lifecycle

Terdapat 2 jenis ML yang dapat digunakan dalam memprediksi PKV, yaitu *supervised learning* dan *unsupervised learning (Badillo dkk, 2020). Supervised learning* adalah pendekatan pembelajaran mesin di mana model dilatih menggunakan dataset yang telah diberi label (Badillo dkk, 2020). Dalam kasus prediksi penyakit, dataset memiliki label berupa kelas risiko, misalnya 0 = tidak berisiko dan 1 = berisiko tinggi. Sedangkan *unsupervised learning* adalah pendekatan di mana model tidak membutuhkan data berlabel, misalnya dalam *clustering* pasien berdasarkan karakteristik tertentu (Badillo dkk, 2020). Ilustrasi *supervised learning* dan *unsupervised learning* dapat dilihat pada Gambar 2.6.



Gambar 2.6 Supervised Learning dan Unsupervised Learning

Dalam bidang kesehatan, ML telah berkembang pesat dan digunakan dalam berbagai aplikasi, termasuk deteksi dini penyakit dan klasifikasi pasien berdasarkan tingkat risiko. Salah satu penerapan ML dalam prediksi penyakit adalah dengan menganalisis pola dari berbagai faktor risiko dan memberikan prediksi yang lebih akurat (Effati dkk., 2024). Cara kerja ML dengan memanfaatkan data untuk mempelajari pola dari data tersebut dan membangun model prediksi untuk mengidentifikasi kemungkinan seseorang mengalami PKV di masa depan (Ghosh dkk., 2021). Proses pembelajaran ini dilakukan dengan membagi data menjadi data training (*training* data) dan data uji (*testing* data), di mana model dilatih menggunakan data training dan kemudian diuji dengan data uji untuk dievaluasi kinerjanya (Anshori dkk., 2022).

Keunggulan ML dalam memprediksi penyakit dibandingkan dengan metode konvensional adalah kemampuannya dalam menganalisis data dalam jumlah besar, mengidentifikasi pola yang kompleks dan memberikan prediksi yang lebih akurat serta personalisasi diagnosis berdasarkan karakteristik individu (Azizan dkk., 2023). Namun, tantangan utama dalam penerapan ML dalam bidang kesehatan adalah ketersediaan dan kualitas data karena model ML sangat bergantung pada data yang digunakan untuk pelatihan.

2.2.5 Preprocessing Data

Preprocessing data merupakan proses untuk mengonversi data mentah menjadi data yang lebih terstruktur dan dioptimalkan untuk pelatihan model, atau dengan kata lain tahap *preprocessing* adalah tahap penyesuaian data yang akan diproses pada setiap *file* atau dokumen dengan tidak mengubah makna yang terkandung dalam data tersebut (Hakim, 2021). Tahapan ini mencakup berbagai aspek, seperti transformasi data, pembersihan, augmentasi, dan penskalaan, dengan penerapan yang disesuaikan berdasarkan jenis data serta tujuan penelitian (Golazad dkk., 2024). Secara keseluruhan, *preprocessing* data bertujuan untuk meningkatkan kualitas data sehingga model *machine learning* dapat memberikan prediksi yang lebih akurat dan andal dalam mendeteksi risiko PKV.

2.2.6 Random Forest

Random Forest (RF) merupakan salah satu algoritma pembelajaran mesin berbasis pohon keputusan (*decision tree*) yang menggunakan teknik *bagging* (*boostrap aggregating*) pada *ensemble learning* dengan membangun banyak *decision trees* untuk meningkatkan akurasi dan stabilitas prediksi sembari mengurangi risiko *overfitting* dengan memanfaatkan *bootstrap sampling* dan *Aggregating* (Y. Yang dan Wang, 2025). RF

bertujuan untuk mengurangi varian dan meningkatkan stabilitas tanpa menaikkan bias secara signifikan dengan cara melatih pohon-pohon yang berbeda pada subset data dan mengagregasi hasilnya (Khan dkk, 2024).

2.2.6.1 Decision Tree

Decision tree (DT) merupakan model prediktif berbentuk struktur hierarkis yang tersusun dari beberapa komponen utama, yaitu *root node, internal node,* dan *leaf node. Root node* merupakan titik awal dari seluruh proses pemisahan (*splitting*). Pemilihan split pada setiap node umumnya dilakukan dengan metrik *impurity* seperti Gini atau Entropy. Split yang menghasilkan impurity yang kecil dianggap optimal (Sun dkk, 2024). *Entropy* merupakan ukuran ketidakpastian variabel acak. Rumus yang digunakan untuk menghitung *entropy* :

$$\text{Entropy}(D) = -\sum_{i=1}^{N} P_i \log_2 (P_i), \tag{2.1}$$

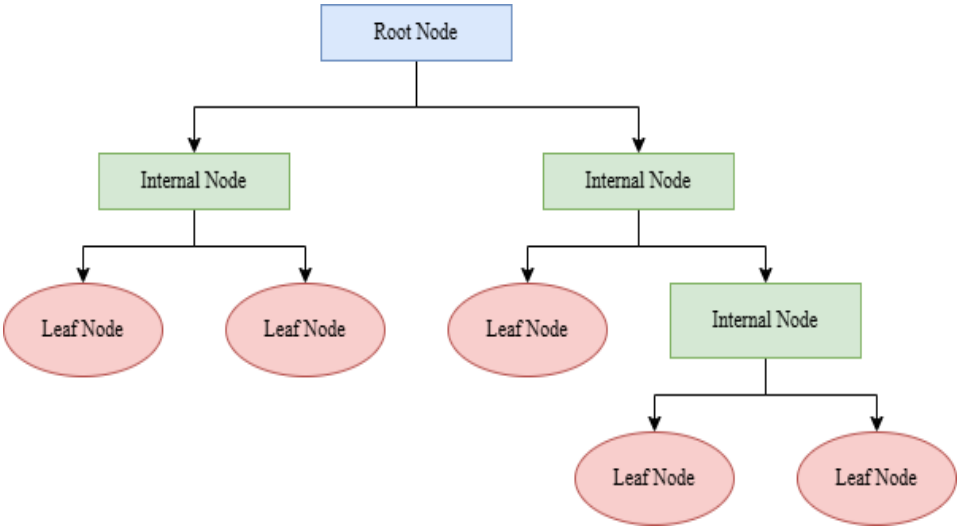
Keterangan:

D = Jumlah keseluruhan data

N = Jumlah partisi

P_i = Proporsi dari D_i terhadap D

Pada tahap *root node*, algoritma akan memilih atribut yang memberikan penurunan impurity paling besar sehingga *root node* menjadi pemisahan awal yang paling informatif. Setelah mendapatkan *root node* langkah selanjutnya akan terbentuk beberapa *internal node.* Internal node berisi subset data yang diperoleh dari pemisahan sebelumnya dan berfungsi sebagai titik evaluasi untuk menentukan *atribut* terbaik berikutnya yang digunakan untuk memisahkan data tersebut. Internal node akan terus berkembang hingga mencapai batas kedalaman atau kondisi berhenti tertentu. Struktur pohon kemudian berakhir pada *leaf node* atau terminal node, yaitu node yang tidak lagi dipecah dan menyajikan hasil prediksi akhir. Visualisasi DT dapat dilihat pada Gambar 2.7.

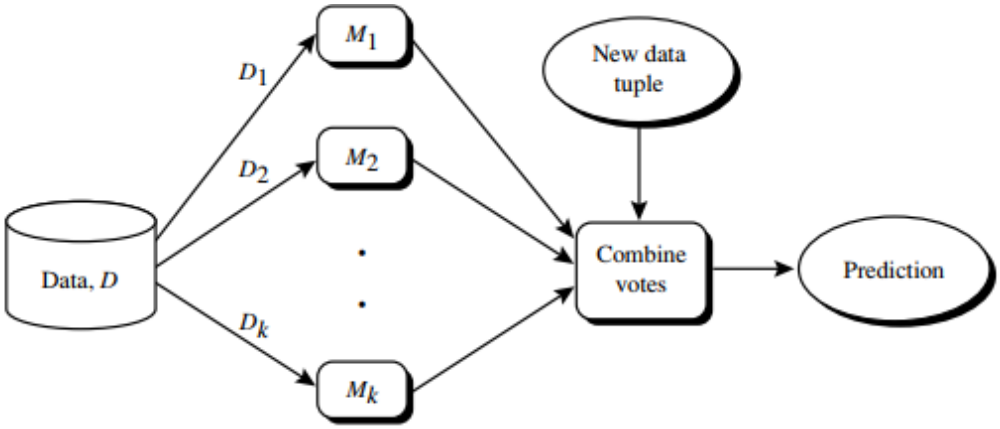


Gambar 2.7 Decision Tree

2.2.6.2 Ensemble Learning

Ensemble Learning merupakan salah satu metode ML yang menggabungkan beberapa model dasar (*base learners*) untuk menghasilkan prediksi yang lebih akurat dan stabil dibandingkan dengan satu model (Alqahtani dkk, 2022). *Ensemble learning* memiliki beberapa metode, diantaranya *bagging*, *boosting*, dan *stacking* yang memiliki cara yang berbeda-beda dalam membangun dan menggabungkan model (Mohammed dan Kora, 2023).

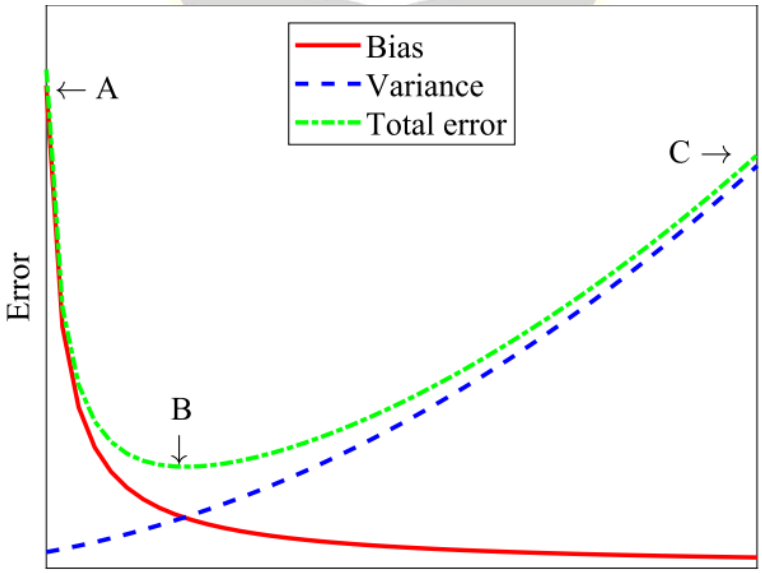
Bagging (*Bootstrap Aggregating*) bekerja dengan membangun beberapa model yang dilatih secara paralel menggunakan beberapa sampel data berbeda yang berasal dari dataset yang sama melalui teknik *bootstrap*, lalu hasil prediksi di aggregasi melalui voting atau rata-rata, untuk menurunkan *variance* model. Sementara itu, *boosting* bekerja dengan cara membangun model secara berurutan, namun setiap model berikutnya fokus untuk memperbaiki kesalahan yang dilakukan oleh model sebelumnya, sehingga bias bisa dikurangi dan akurasi keseluruhan meningkat. Sedangkan *stacking* bekerja dengan cara menggabungkan model-model yang berbeda (heterogen) dengan menggunakan satu *meta-learner* yang belajar dari *prediksi base learners* untuk memberikan prediksi akhir, sehingga *ensemble* bisa memanfaatkan kekuatan setiap model dasar. Visualisasi *Ensemble learning* dapat dilihat pada Gambar 2.8.



Gambar 2.8 Ensemble Learning

2.2.6.3 Bias dan Variance

Dalam pembelajaran mesin, kinerja model sangat dipengaruhi oleh *trade-off* antara *bias* dan *variance*. *Trade-off bias–variance* menggambarkan keseimbangan antara kesalahan akibat asumsi model yang terlalu sederhana (*bias*) dan kesalahan akibat sensitivitas model yang berlebihan terhadap data training (*variance*) (Guan dan Burton, 2022). *Bias* merupakan kesalahan yang muncul karena model terlalu sederhana sehingga tidak mampu untuk menangkap pola data yang sebenarnya sehingga berpotensi menyebabkan *underfitting*. Sebaliknya, *variance* merupakan kesalahan model yang terlalu sensitif terhadap data, sehingga menyebabkan model *overfitting*. Ilustrasi *Trade-off bias–variance* dapat dilihat pada Gambar 2.9.



Gambar 2.9 Trade Off Bias–Variance

Gambar 2.9 menggambarkan *trade-off* antara bias dan variance terhadap kompleksitas model. Pada titik A, model masih sangat sederhana sehingga bias tinggi dan

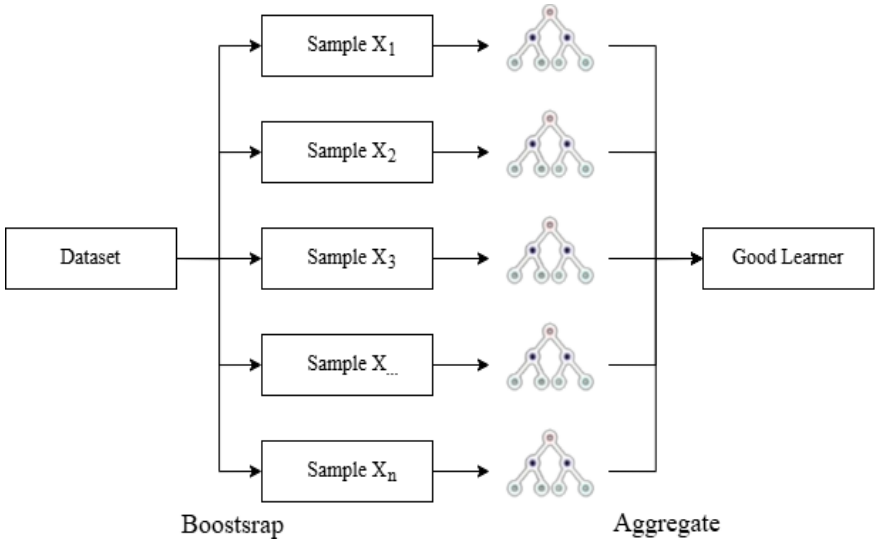
model tidak mampu menangkap pola data yang sebenarnya dan berpotensi menyebabkan *underfitting*. Seiring meningkatnya kompleksitas model, bias menurun namun variance mulai meningkat, hingga mencapai titik B sebagai kondisi optimal, keseimbangan bias dan variance menghasilkan *error* total paling rendah. Pada tiitk C, model menjadi terlalu kompleks sehingga *variance* tinggi, model sangat sensitif terhadap data, dan berisiko mengalami *overfitting* (Guan dan Burton, 2022)

Pada *Decision Tree*, model cenderung memiliki bias yang rendah karena mampu mempelajari pola data yang kompleks melalui pemisahan atribut secara berulang. Namun, fleksibilitas ini menyebabkan *variance* yang tinggi, di mana struktur pohon sangat dipengaruhi oleh perubahan kecil pada data pelatihan. Akibatnya, *Decision Tree* rentan mengalami *overfitting*, terutama pada data berukuran kecil, berdimensi tinggi, atau mengandung noise (Ibrahim, 2023). Pada *ensemble learning*, khususnya metode *bagging* dikembangkan untuk menurunkan *variance* dengan menggabungkan beberapa model yang dilatih pada subset data yang berbeda (Ibrahim, 2023).

2.2.6.4 Algoritma Random Forest

Random Forest adalah metode *ensemble learning* yang menggabungkan banyak *decision tree* (DT) menggunakan teknik *bagging*. Tujuannya adalah meningkatkan akurasi, mengurangi *variance*, dan mencegah *overfitting* (Y. Yang dan Wang, 2025). *Random Forest* secara khusus dirancang untuk mengurangi *variance* secara signifikan tanpa meningkatkan bias melalui penerapan *bootstrap sampling* pada setiap pohon keputusan. Pendekatan ini menghasilkan model yang lebih stabil, tahan terhadap *overfitting*, serta memiliki kemampuan generalisasi yang lebih unggul dibandingkan Decision Tree, terutama pada data berdimensi tinggi (data yang memiliki jumlah atribut yang banyak dalam 1 sampel) dan bersifat non-linier (Ibrahim, 2023).

Pada penerapannya, setiap pohon dilatih pada sampel data yang diambil menggunakan teknik *bootstrap* (data yang diambil dengan pengembalian) sehingga sampel data memiliki variasi struktur. Setelah semua pohon terbentuk, prediksi akhir ditentukan melalui voting mayoritas (klasifikasi) atau rata-rata (regresi). Algoritma RF bekerja dengan cara membangun banyak pohon keputusan melalui proses *bagging* dan pemilihan atribut secara acak agar diperoleh model yang lebih akurat dan stabil. Proses RF dapat dilihat pada Gambar 2.10.



Gambar 2.10 Proses Random Forest

Dapat dilihat pada Gambar 2.10, proses RF dimulai dengan mengambil sebanyak X sample data dari dataset asli menggunakan teknik *bootstrap,* sehingga setiap sampel data yang terbentuk menjadi subset memiliki komposisi yang berbeda. Setiap subset tersebut kemudian digunakan untuk melatih satu pohon keputusan. Pada tahap pembentukan pohon, RF hanya mempertimbangkan sebagian atribut yang dipilih secara acak pada setiap *node* untuk menghasilkan keragaman antar *tree* dan mencegah satu atribut dominan mempengaruhi seluruh model. Setelah semua pohon selesai dilatih, seluruh prediksinya digabungkan (*aggregate*) menggunakan *majority voting* (klasifikasi) atau *averaging* (regresi). Kombinasi banyak pohon yang berbeda inilah yang menghasilkan *good learner*, model yang lebih kuat, lebih tahan terhadap *noise*, dan lebih baik dalam generalisasi dibandingkan satu DT tunggal.

2.2.7 Teknik Optimasi Model

Teknik Optimasi model merupakan proses untuk meningkatkan performa model dengan menyesuaikan parameter pelatihan agar dapat menghasilkan prediksi yang lebih akurat dan efisien (Rosidah dan Prathivi, 2025). Teknik optimasi dilakukan melalui beberapa tahap, seperti *tuning hyperparameter* dan teknik validasi model (Dhanka dan Maini, 2025).

2.2.7.1 Tuning Hyperparameter

Tuning hyperparameter merupakan proses mencari kombinasi nilai hyperparameter terbaik untuk memaksimalkan performa model di data validasi (Nematzadeh dkk, 2022). Pada model RF, *hyperparameter* yang sering disesuaikan seperti jumlah pohon keputusan

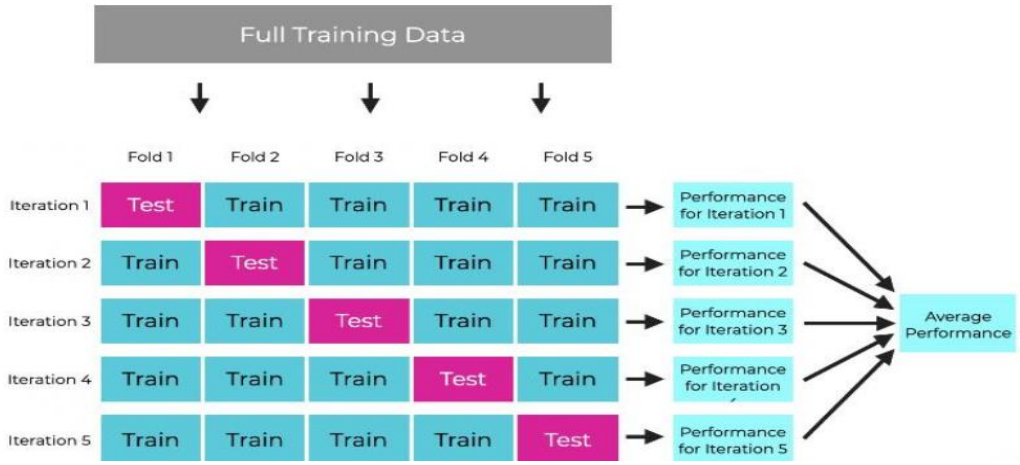
(n_estimators), kedalaman maksimum pohon (*max_depth*), jumlah atribut yang dipertimbangkan pada setiap split (*max_features*), dan kriteria pemisahan node (*criterion*). Terdapat beberapa metode tuning hyperparameter, diantaranya *grid search*, *random search*, *model-based* (*sequential*) seperti *Bayesian optimization* (Bischl dkk, 2023).

Grid search merupakan metode yang mencari kombinasi terbaik dengan mengevaluasi seluruh nilai dalam *grid* yang telah ditentukan, sehingga mudah diterapkan pada dataset yang tidak terlalu besar namun membutuhkan sumber daya komputasi yang sangat besar pada ruang pencarian yang luas. *Random search* merupakan metode yang lebih efisien karena memilih kombinasi *hyperparameter* secara acak dari distribusi nilai sehingga mampu menjelajahi ruang pencarian yang lebih luas tanpa harus menguji semua kemungkinan. Sementara itu, *Bayesian optimization* menggunakan pendekatan probabilistik dengan membangun *surrogate model* (model pengganti) untuk memprediksi kinerja setiap kombinasi *hyperparameter* dan memilih kandidat terbaik secara iteratif melalui fungsi akuisisi, sehingga jauh lebih hemat komputasi dan efektif terutama ketika proses pelatihan model membutuhkan biaya tinggi.

2.2.8 K-Fold Validation

K-fold cross-validation merupakan salah satu teknik validasi model yang digunakan untuk mengevaluasi performa model dengan membaginya menjadi beberapa subset data training dan data uji secara bergantian untuk memastikan proporsi kelas pada setiap lipatan tetap seimbang (Wilimitis dan Walsh, 2023). Validasi model berperan untuk memastikan bahwa model tidak mengalami *overfitting* atau *underfitting* (Dhanka dan Maini, 2025).

Proses *training* dan evaluasi dilakukan sebanyak x kali, dengan setiap *fold* secara bergantian berperan sebagai data uji (*test set*), sementara *fold* lainnya digunakan sebagai data training (*training set*). Visualisasi K-Fold Validation dapat dilihat pada Gambar 2.11.



Gambar 2.11 K-Fold Validation dengan K=5

2.2.9 Evaluasi Model Prediksi

Evaluasi model prediksi merupakan proses penting untuk menilai sejauh mana model mampu menghasilkan prediksi yang tepat dan untuk memahami kekuatan dan kelemahan model dalam menyelesaikan tugas (mengklasifikasikan data atau menemukan pola tertentu berdasarkan dataset) yang diberikan (Fadli dan Saputra, 2023). Tahapan umum dalam evaluasi model diantaranya persiapan data, pelatihan model, evaluasi dengan metrik klarifikasi (Fadli dan Saputra, 2023).

Pada penelitian ini, proses evaluasi model terhadap prediksi risiko PKV menggunakan metrik evaluasi yang terdiri dari akurasi, presisi, recall, dan F1-Score. Matriks evaluasi dapat dilihat pada Tabel 2.5. Berikut penjelasan masing-masing metrik :
Metrik akurasi berfungsi untuk mengukur seberapa sering model benar memprediksi label positif dan negatif penyakit kardiovaskular.
Berikut persamaan metrik akurasi :

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

Metrik presisi berfungsi untuk mengukur seberapa sering model dalam memprediksi positif PKV dengan benar sesuai label, berikut persamaan metrik presisi :

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

Metrik recall berfungsi untuk mengukur seberapa sering model mendeteksi seluruh kasus penyakit yang positif di antara semua kasus positif yang tersedia. Berikut persamaan metrik recall :

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

Metrik F1-Scores merupakan nilai rata-rata dari metrik presisi dan recall, metrik ini memberikan hasil yang lebih baik tentang keseimbangan antara kedua metrik tersebut. Berikut persamaan metrik presisi :

$$\text{F1} - \text{Scores} = 2\ x\ \frac{Presisi \text{ x } Recall}{Presisi + Recall} \quad (2.6)$$

Keterangan Persamaan:
TP = *True Positive*
TN = *True Negative*
FP = *False Positive*
FN = *False Negative*.

Tabel 2.5 Metrik Evaluasi

Metrik	Fokus Penilaian	Rumus	Kegunaan
Akurasi	Prediksi benar secara keseluruhan	(TP+TN) / (TP+TN+FP+FN)	Mengukur proporsi prediksi benar dari seluruh prediksi
Presisi	Ketepatan prediksi positif	TP / (TP+FP)	Mengukur seberapa banyak prediksi positif yang benar
Recall	Kemampuan menemukan positif	TP / (TP+FN)	Mengukur seberapa banyak kasus positif berhasil terdeteksi
F1-Score	Keseimbangan presisi dan recall	2 ×(Presisi × Recall) / (Presisi + Recall)	Nilai rata-rata dari metrik presisi dan recall

2.2.10 Explainable Artificial Intelligence

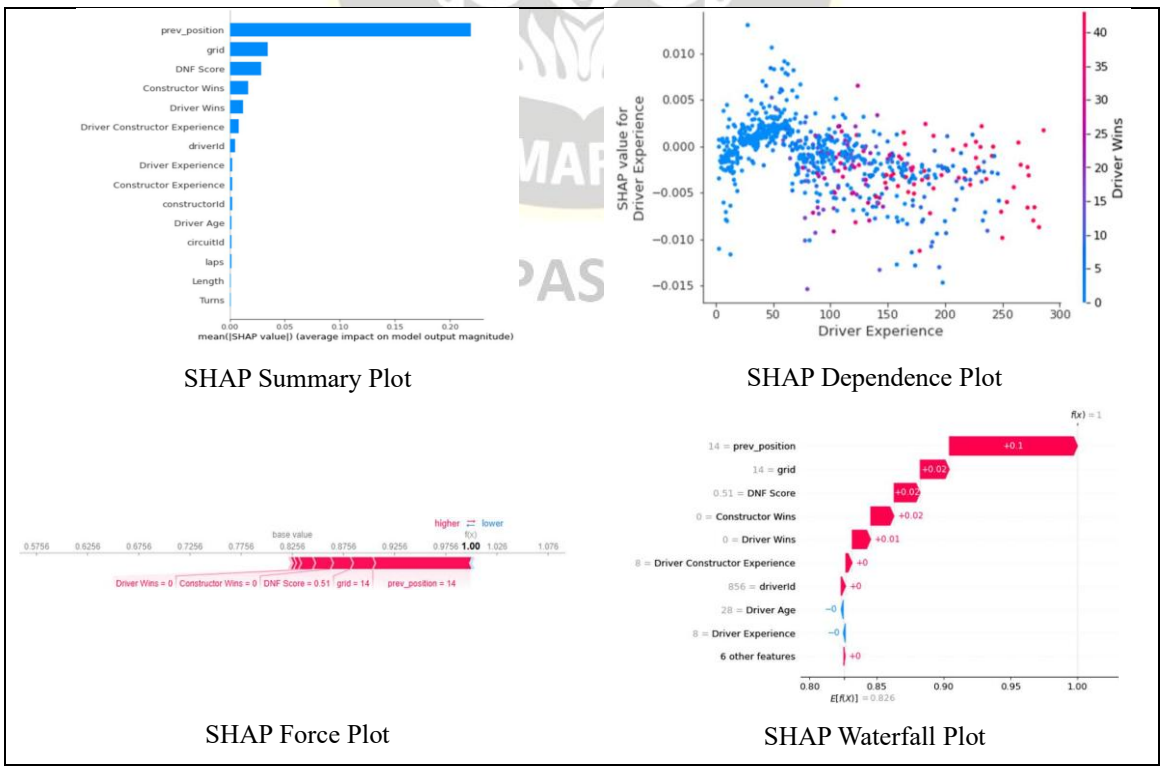
Explainable Artificial Intelligence (XAI) merupakan cabang dari kecerdasan buatan (AI) yang berfokus pada peningkatan transparansi dan intepretabilitas dari model AI sehingga proses pengambilan keputusannya dapat dipahami oleh manusia (Sharma dkk, 2024). Secara umum, XAI terdiri dari dua jenis pendekatan, yaitu *intrinsic explainability* dan *post-hoc explainability.* Metode *intrinsic* menggunakan model yang memang mudah dijelaskan sejak awal, seperti regresi linear, *decision tree*, atau *rule-based models*. Sebaliknya, metode *post-hoc* digunakan untuk menjelaskan model kompleks setelah dilatih, seperti *deep neural networks* atau *ensemble trees*. Contoh metode *post-hoc*, yaitu LIME, SHAP, Grad-CAM, dan visualisasi atribut. Namun, LIME dan SHAP merupakan metode yang paling banyak digunakan dalam bidang kesehatan karena dapat memberikan penjelasan mengenai kontribusi atribut terhadap prediksi (Vimbi dkk, 2024).

Sebagai contoh, (Vimbi dkk, 2024) melakukan penelitian menggunakan dua metode XAI, yaitu LIME dan SHAP untuk menjelaskan hasil prediksi penyakit Alzheimer. Hasilnya adalah LIME mampu menyoroti atribut pasien seperti *cognitive scores*, citra MRI lokal, dan biomarker tertentu, namun hasil ini cenderung bervariasi dan kurang stabil antar percobaan. Sementara itu, SHAP mengidentifikasi atribut penting seperti *volume hippocampus, atrofi kortikal*, atau nilai *kognitif* tertentu yang konsisten memengaruhi prediksi Alzheimer dan memberikan penjelasan yang lebih komprehensif. Kedua hasil tersebut menunjukkan bahwa SHAP lebih unggul dalam menunjukkan kontribusi setiap atribut secara konsisten di seluruh dataset dan dapat disimpulkan bahwa SHAP dianggap lebih informatif dibandingkan LIME.

2.2.10.1 Shapley Additive Explanations

Shapley Additive Explanations (SHAP) merupakan metode *Explainable* AI (XAI) yang menjelaskan kerelevanan hasil prediksi model ML berdasarkan teori yang dikembangkan oleh Lloyd Shapley pada tahun 1953. Visualisasi SHAP berfungsi untuk menggambarkan kontribusi setiap atribut terhadap prediksi model. Terdapat beberapa jenis visualisasi SHAP, diantaranya *SHAP Summary Plot, SHAP Dependence Plot, SHAP Force Plot dan SHAP Waterfall Plot* yang masing-masing memiliki gambaran analisis yang berbeda (El Haber dkk, 2025).

SHAP Summary Plot digunakan untuk memberikan gambaran mengenai pentingnya setiap atribut secara keseluruhan dan menunjukkan atribut paling berpengaruh terhadap *output* model. *SHAP Dependence Plot* digunakan untuk memberikan gambaran lebih detail terkait hubungan antar nilai suatu atribut dan kontribusinya terhadap prediksi. *SHAP Force Plot* digunakan untuk menggambarkan atribut mana yang meningkatkan atau menurunkan nilai prediksi, sehingga berguna untuk analisis kasus spesifik. *SHAP Waterfall Plot* digunakan untuk memberikan gambaran mengenai perhitungan kontribusi atribut secara berurutan dari nilai dasar (*expected value*) hingga menghasilkan prediksi akhir dan membantu memahami alur logika model untuk satu observasi. Contoh masing-masing visualisasi dapat dilihat pada Gambar 2.12.



Gambar 2.12 Jenis-Jenis Visualisasi SHAP

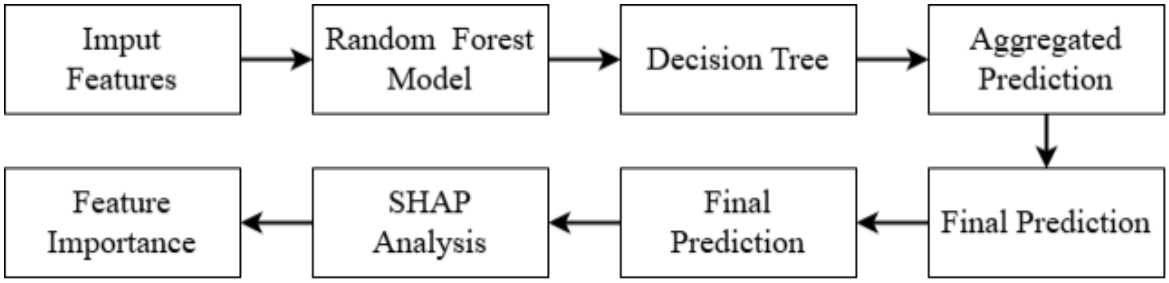
SHAP dapat menjelaskan secara kuantitatif seberapa besar kontribusi setiap atribut terhadap prediksi yang dihasilkan oleh model (Alkhanbouli dkk, 2025). Namun dalam menganalisis SHAP, belum ada aturan ambang batas (*fixed threshold*) yang benar-benar pasti (seperti *p-value* dalam statistik inferensial) untuk menentukan apakah suatu atribut "signifikan" atau "tidak signifikan" secara persentase. Oleh karena itu, penentuan *fixed threshold* pada penelitian ini dilakukan dengan cara mengelompokkan atribut berdasarkan besarnya pengaruh terhadap hasil prediksi model. *fixed threshold* ditetapkan secara *heuristik* dengan melihat sebaran nilai SHAP pada model terbaik, sehingga atribut dengan kontribusi yang besar dikategorikan sebagai sangat signifikan, pengaruh sedang sebagai cukup signifikan, dan pengaruh kecil sebagai tidak signifikan. Kategori SHAP dapat dilihat pada Tabel 2.6 Rumus yang digunakan untuk menghitung nilai SHAP adalah sebagai berikut :

$$Persentase = \frac{Nilai\ SHAP\ Atribut}{Total\ Nilai\ SHAP}\ x\ 100\% \quad (2.2)$$

Tabel 2.6 Kategori SHAP

Kategori	Kriteria/Batas Persentase	Arti
Sangat Signifikan	>= 10	Faktor Utama Model
Cukup Signifikan	>= 1% – 9%	Faktor Pendukung
Tidak Signifikan	< 1%	Faktor Pengaruh Minimal

Fixed threshold ditentukan dengan mempertimbangkan pola dominasi kontribusi atribut pada model, beberapa atribut menunjukkan kontribusi yang jauh lebih besar dibanding atribut lainnya, sehingga diperlukan pengelompokan untuk membedakan faktor utama, pendukung, dan faktor dengan pengaruh minimal. Alur kerja algoritma RF dengan analisis SHAP dapat dilihat pada Gambar 2.13.



Gambar 2.13 Alur kerja algoritma RF dengan analisis SHAP