

# Prediksi Penyakit Hipertensi Menggunakan metode Decision Tree dan Random Forest

Oleh :

M. Fahrul Rizki Aditya

Nuril Lutvi Azizah

Program Studi Teknik Informatika

Universitas Muhammadiyah Sidoarjo

Agustus, 2023

# Pendahuluan

- Hipertensi merupakan isu kesehatan masyarakat yang penting dimana jarang menyebabkan gejala nyata pada kesehatan fungsional pasien. Hipertensi merupakan faktor resiko utama pada penyakit jantung coroner, gagal jantung serta stroke. Faktor penyebab terjadinya hipertensi adalah umur, jenis kelamin, Riwayat keluarga, genetic, kebiasaan merokok, obesitas, kurang aktivitas fisik, stress, dan penggunaan esterogen.
- Hipertensi atau tekanan darah tinggi adalah kondisi ketika tekanan darah mencapai 140/90 mmHg atau lebih. Hipertensi merupakan gangguan pada pembuluh darah yang dapat mengakibatkan suplai oksigen dan nutrisi yang dibawa oleh darah terhambat sampai ke jaringan tubuh yang membutuhkan. Penyakit hipertensi ini dapat menjadi ancaman karena dengan tiba-tiba seseorang dapat divonis menderita darah tinggi

# Pendahuluan

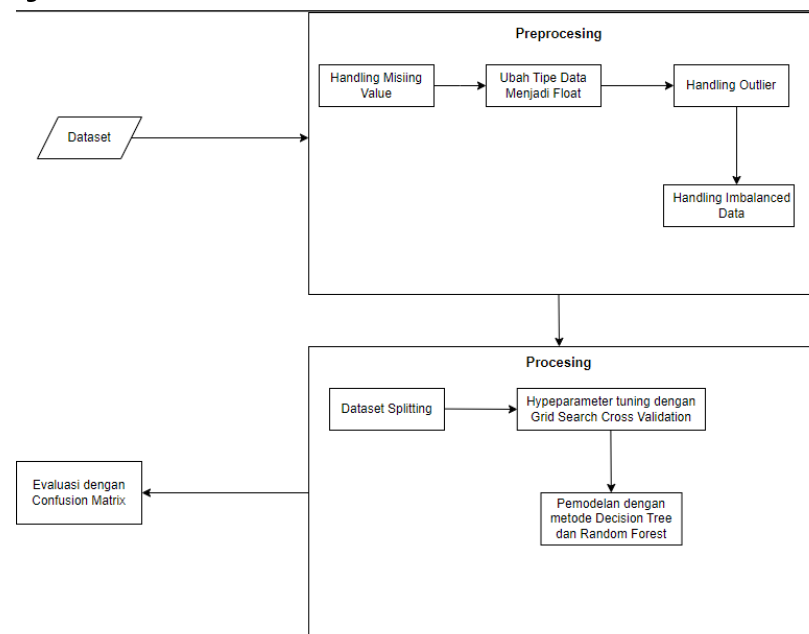
- Gejala utama yang dirasakan oleh penderita penyakit hipertensi dibagi menjadi 2 tipe, yaitu penderita hipertensi ringan dan hipertensi berat. Penderita hipertensi ringan akan merasakan sakit kepala, mimisan, nyeri dada, dan sesak nafas. Sedangkan, penderita hipertensi berat akan merasakan nyeri pada dada, tremor otot, dan mual
- Saat ini teknologi semakin berkembang. Komunitas medis sangat terbantu dengan perkembangan teknologi. Salah satunya adalah program yang dapat digunakan untuk mendeteksi penyakit hipertensi menggunakan kecerdasan buatan. Adapun bidang kecerdasan yang dapat digunakan adalah machine learning.
- Machine learning merupakan cabang ilmu bagian dari kecerdasan buatan (artificial intelligence), dengan pemrograman untuk memungkinkan komputer menjadi cerdas berperilaku seperti manusia, dan dapat meningkatkan pemahamannya melalui pengalaman secara otomatis

# Pendahuluan

- Machine learning memiliki fokus pada pengembangan sistem yang mampu belajar sendiri untuk memutuskan sesuatu tanpa harus berulang kali diprogram oleh manusia. Hal ini menjadikan mesin tidak hanya mampu berperilaku mengambil keputusan, namun juga dapat beradaptasi dengan perubahan yang terjadi. Machine learning menggunakan teknik untuk menangani data besar (big data) dengan cara yang cerdas untuk memberikan hasil yang tepat.
- penelitian ini bertujuan untuk mengoptimalkan kinerja model dan memberikan wawasan mengenai keunggulan kedua metode dalam hal akurasi dan kemampuan prediksi.

# Metode Penelitian

- Tahapan penelitian merupakan gambaran umum yang mencakup rangkaian alur penelitian yang akan dilakukan dalam pengerjaan penelitian ini dari awal hingga akhir menggunakan Bahasa pemrograman python dan menggunakan software jupyter notebook. Tahapan penelitian ini dapat dipaparkan melalui diagram alir seperti yang ditunjukkan dalam ilustrasi dibawah ini:



# Metode Penelitian

## Procesing

Data mentah yang ada pada bab sebelumnya tidak dapat diproses oleh mesin secara langsung. Perlu dilakukan modifikasi pada data mentah agar menjadi data yang siap diproses oleh mesin. Adapun beberapa tahapan dari preprocessing adalah sebagai berikut:

- Handling Missing Value

pada tahapan awal dari preprocessing adalah dengan melakukan pengecekan missing value atau data kosong yang ada pada data mentah. Data kosong ini biasanya terjadi karena adanya kesalahan dalam input data atau memang data tersebut sengaja tidak diisi. Mesin tidak dapat memproses data yang tidak bernilai. Oleh sebab itu perlu dilakukan impute missing value atau pengisian data kosong. Terdapat beberapa cara untuk melakukan pengisian data kosong diantaranya adalah mengambil nilai rata-rata suatu kolom atau melakukan penghapusan pada kolom tersebut. Namun penghapusan merupakan opsi terakhir apabila kolom yang terdapat data kosong tersebut adalah kolom target atau terlalu banyak data kosong yang ada pada kolom tersebut.

# Metode Penelitian

- Ubah Tipe Data Menjadi Float

Data yang digunakan pada penelitian ini memiliki tipe data float dan integer. Agar data memiliki satu tipe data yang sama, maka peneliti mengkonversi tipe data semua kolom menjadi float.

- Handling Outlier

Dalam analisis data, istilah outlier merujuk pada nilai yang secara signifikan berbeda dari nilai-nilai lain dalam dataset. Outlier dapat muncul karena kesalahan pengukuran, variasi yang ekstrem secara alami, atau situasi-situasi yang tidak biasa dalam data. Outlier dapat memberikan dampak negatif pada analisis statistik dan kinerja model prediksi karena bisa mempengaruhi perkiraan parameter dan merusak kinerja model.

# Metode Penelitian

- Handling Imbalanced Data

Tahap berikutnya yaitu melakukan proses Handling Imbalanced Data yang bertujuan untuk Menyeimbangkan Jumlah Kelas dengan menggunakan pendekatan oversampling. Teknik oversampling digunakan untuk memperluas ukuran sampel yang kurang umum dengan menerapkan metode SMOTE (Sintetis Minoritas Sampling Teknik).

## Procesing

Setelah melalui tahap preprocessing, selanjutnya adalah melakukan tahap processing atau pemrosesan data dengan menerapkan metode Decision tree dan Random Forest. Adapun tahapan processing pada penelitian ini adalah sebagai berikut.



# Metode Penelitian

- Data Splitting

Pada tahap awal pemrosesan, dataset dibagi menjadi dua bagian yaitu data latih dan data uji, dengan perbandingan sekitar 90% untuk data latih dan 10% untuk data uji. Data latih akan digunakan dalam proses pembelajaran mesin, sementara data uji akan dimanfaatkan untuk menguji model yang telah dihasilkan.

- Hyperparameter Dengan Grid Search Cross Validation

Grid search merupakan salah satu metode yang sering kali digunakan untuk melakukan pencarian nilai paling optimal dari sebuah parameter dalam proses pembentukan model. Proses ini biasa disebut dengan hyper parameter tuning. Tujuan proses GridSearch beserta Cross-Validation adalah untuk mengidentifikasi kombinasi hyper-parameter terbaik sehingga model dapat memprediksi data dengan akurat dan optimal

# Metode Penelitian

## Klasifikasi Dengan Metode Decision Tree dan Random Forest

Langkah pertama dalam tahap klasifikasi dengan metode decision tree dan random forest adalah membuka data yang diekstraksi di notebook Jupyter menggunakan library pandas. Setelah data berhasil dimuat, dilakukan pemisahan data antara data X dan data Y, dimana data X adalah kolom atribut dan data Y adalah kolom target.

## Tahap Evaluasi

Tahapan ini digunakan untuk mengukur performa dari model machine learning yang telah dibuat. terdapat tiga metrik evaluasi yang dapat digunakan yaitu precision, recall dan confusion matrix. Terdapat 4 istilah sebagai representasi hasil proses klasifikasi pada confusion matrix. Dengan adanya perhitungan dari confusion matrix maka dapat diperoleh accuracy, sensitivity, dan specificity. Agar mendapatkan hasil terbaik, beberapa perbandingan data train dan data test yang ada pada table di atas nantinya akan dilakukan percobaan pada masing-masing perbandingan.

# Hasil dan Pembahasan

- **Decision Tree**

Hyperparameter	Value
Criterion	gini
Max_depth	None
Mean_sampels_leaf	1
Min_sampels_split	2

Dari nilai parameter terbaik yang dihasilkan melalui grid search cross validation seperti yang ditunjukkan dalam Tabel 2, tercatat bahwa skor pelatihan (train) mencapai 100%, sementara skor pengujian (test) juga berhasil mencapai 100%.

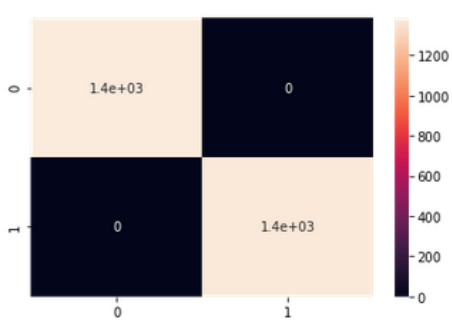
# Hasil dan Pembahasan

- **Random Forest**

Hyperparameter	Value
Criterion	gini
Max_depth	none
Mean_sampels_leaf	1
Min_sampels_split	10
n_estimator	200

Dari nilai parameter terbaik yang dihasilkan melalui grid search cross validation seperti yang ditunjukkan dalam Tabel 3, tercatat bahwa skor pelatihan (train) mencapai 100%, sementara skor pengujian (test) juga berhasil mencapai 100%.

# asil dan Pembahasan



algoritma	Confusion matrix
Decision tree	
Random Forest	

algoritma	Classification Report
Decision tree	<pre>Classification Report:   precision    recall  f1-score   support     0.0         1.00         1.00         1.00        1385    1.0         1.00         1.00         1.00        1369   accuracy          1.00          1.00          1.00        2754  macro avg         1.00          1.00          1.00        2754  weighted avg      1.00          1.00          1.00        2754  Confusion Matrix: [[1385   0]  [   0 1369]]</pre>
Random Forest	<pre>Classification Report:   precision    recall  f1-score   support     0.0         1.00         1.00         1.00        1385    1.0         1.00         1.00         1.00        1369   accuracy          1.00          1.00          1.00        2754  macro avg         1.00          1.00          1.00        2754  weighted avg      1.00          1.00          1.00        2754  Confusion Matrix: [[1385   0]  [   0 1369]]</pre>

Dari tabel 4 dan 5 yang tertera di atas, terlihat bahwa metode decision tree memberikan hasil akurasi, presisi, recall, dan F1-score sebesar 100%. Begitu juga, metode random forest juga menunjukkan hasil yang sama dengan semua nilai akurasi, presisi, recall, dan F1-score sebesar 100%.

# Kesimpulan

- Melalui penggunaan model yang dibangun dalam penelitian ini, peneliti berhasil melakukan prediksi penyakit hipertensi. Dengan menerapkan metode decision tree menghasilkan nilai akurasi sebesar 100%, dan metode random forest juga menghasilkan akurasi yang sama, yaitu 100%.

# Daftar Pustaka

- [1] B. L. Yudha, L. Muflikhah, and R. C. Wihandika, “Klasifikasi Risiko Hipertensi Menggunakan Metode Neighbor Weighted K- Nearest Neighbor (NWKNN),” J. Pengemb. Teknol. Inf. dan Ilmu Komput., vol. 2, no. 2, pp. 897–904, 2018.
- [2] R. E. Putri, “Implementasi Expert System Diagnosa Penyakit Hipertensi Menggunakan Metode Dempster Shafer,” JATISI (Jurnal Tek. Inform. dan Sist. Informasi), vol. 9, no. 2, pp. 1557–1567, 2022, doi: 10.35957/jatisi.v9i2.2100.
- [3] A. C. Telaumbanua and Y. Rahayu, “Penyuluhan Dan Edukasi Tentang Penyakit Hipertensi,” J. Abdimas Sainatika, vol. 3, no. 1, p. 119, 2021, doi: 10.30633/jas.v3i1.1069.
- [4] A. N. Syahrudin and T. Kurniawan, “Input dan Output pada Bahasa Pemrograman Python,” J. Dasar Pemrograman Python STMIK, no. January, pp. 1–7, 2018, [Online]. Available: <https://www.researchgate.net/publication/338385483>.
- [5] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, “Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN),” J. Inform., vol. 11, no. 1, p. 36, 2017, doi: 10.26555/jifo.v11i1.a5452.

# Daftar Pustaka

- [6] E. Retnoningsih and R. Pramudita, “Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python,” *Bina Insa. Ict J.*, vol. 7, no. 2, p. 156, 2020, doi: 10.51211/biict.v7i2.1422.
- [7] B. Kriswantara and R. Sadikin, “Used Car Price Prediction with Random Forest Regressor Model,” *J. Inf. Syst. Informatics Comput. Issue Period*, vol. 6, no. 1, pp. 40–49, 2022, doi: 10.52362/jisicom.v6i1.752.
- [8] A. Muzakir and R. A. Wulandari, “Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree,” *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016, doi: 10.15294/sji.v3i1.4610.
- [9] G. A. Sandag, “Prediksi Rating Aplikasi App Store Menggunakan Algoritma Random Forest,” *CogIto Smart J.*, vol. 6, no. 2, pp. 167–178, 2020, doi: 10.31154/cogito.v6i2.270.167-178.
- [10] M. M. Santoni, N. Chamidah, and N. Matondang, “Prediksi Hipertensi menggunakan Decision Tree, Naïve Bayes dan Artificial Neural Network pada software KNIME,” *Techno.Com*, vol. 19, no. 4, pp. 353–363, 2020, doi: 10.33633/tc.v19i4.3872.



# Daftar Pustaka

- [11] P. Purwono et al., “Model Prediksi Otomatis Jenis Penyakit Hipertensi Dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network,” *Insect (Informatics Secur. J. Tek. Inform.*, vol. 7, no. 2, pp. 82–90, 2022, [Online]. Available: <http://www.ejournal.um-sorong.ac.id/index.php/insect/article/view/1828>.
- [12] A. Wantoro, A. Syarif, K. N. Berawi, K. Muludi, S. R. Sulistiyanti, and S. Sutyarso, “Implementasi Metode Pembobotan Berbasis Aturan Dan Metode Profile Matching Pada Sistem Pakar Medis Untuk Prediksi Risiko Hipertensi,” *J. Teknoinfo*, vol. 15, no. 2, p. 134, 2021, doi: 10.33365/jti.v15i2.1523.
- [13] S. Choerunnisa Nurzanah, S. Alam, and T. Iman Hermanto, “Analisis Association Rule Untuk Identifikasi Pola Gejala Penyakit Hipertensi Menggunakan Algoritma Apriori (Studi Kasus: Klinik Rafina Medical Center),” *JIKO (Jurnal Inform. dan Komputer)*, vol. 5, no. 2, pp. 132–141, 2022, doi: 10.33387/jiko.v5i2.4792.
- [14] F. Supriyono, Erida, “SISTEM PAKAR PENEGAKAN DIAGNOSA PENYAKIT HIPERTENSI DENGAN INFÉRENSI FORWARD CHAINING MENGGUNAKAN METODE SUPPORT VECTOR MACHINE (SVM),” vol. 08, no. 02, pp. 207–221, 2022.
- [15] S. A. Nugraha, H. N. Palit, and H. Juwiantho, “Aplikasi Analisa Sentimen Bilingual dan Emoji pada Komentar Media Sosial Instagram Menggunakan Metode Support Vector Machine.”

# UMSIDA

DARI SINI PENCERAHAN BERSEMI



[www.umsida.ac.id](http://www.umsida.ac.id)



[umsida1912](#)



[umsida1912](#)



universitas  
muhammadiyah  
sidoarjo



[umsida1912](#)

