

**Prediksi Tahap Awal Penyakit Jantung Menggunakan Algoritma Random Forest
(Studi Kasus RSIJ)**

Tony Raymond Nangon¹, Allan Desi Alexander²

^{1,2}Universitas Bhayangkara

frizki192@gmail.com¹

ABSTRACT

*The dataset contains 303 samples and 14 attributes are taken for the dataset features. Then processed using open access Python software in a Jupyter notebook. The data set is classified and processed using the Random forest machine learning algorithm. The results of the data set are expressed in terms of accuracy, sensitivity and specificity in percentages. Using the random forest algorithm, an accuracy of 86.9% was obtained for predicting heart disease with a sensitivity value of 90.6% and a specificity value of 82.7%. From the receiver operating characteristics, we obtained a diagnosis rate for heart disease prediction using random forest of 93.3%. Random forest algorithm has been proven to be the most efficient algorithm for heart disease classification and hence it is used in the proposed system. **Keywords** : decision support system, vikor method, rubber tree seeds.*

ABSTRAK

Dataset berisi 303 sampel dan 14 atribut diambil untuk fitur dataset. Kemudian diolah menggunakan software python open access di notebook jupyter. Kumpulan datanya adalah diklasifikasikan dan diproses menggunakan algoritma pembelajaran mesin Random forest. Hasil dari kumpulan data dinyatakan dalam akurasi, sensitivitas dan spesifisitas dalam persentase. Menggunakan acak algoritma forest, diperoleh akurasi sebesar 86,9% untuk prediksi penyakit jantung dengan nilai sensitivitas 90,6% dan nilai spesifisitas 82,7%. Dari karakteristik pengoperasian receiver, kami memperoleh tingkat diagnosis untuk prediksi penyakit jantung menggunakan random forest adalah 93,3%. Hutan acak algoritma telah terbukti menjadi algoritma yang paling efisien untuk klasifikasi penyakit jantung dan karenanya itu digunakan dalam sistem yang diusulkan.

PENDAHULUAN

Klasifikasi dalam data mining adalah teknik untuk memisahkan data ke dalam kelas berbeda. Salah satu algoritma klasifikasi adalah random forest. Klasifikasi dengan random forest dilakukan dengan membangun beberapa pohon keputusan yang menghasilkan satu suara prediksi dari suara mayoritas (1). Pohon keputusan dibangun dengan menggunakan Classification and Regression Trees (CART). Random forest dapat menghasilkan akurasi yang tinggi dibandingkan dengan metode klasifikasi lainnya. Hanya saja saat random forest dijalankan pada imbalanced data akan mengakibatkan kemampuan klasifikasi menurun. Dalam algoritma pengklasifikasian termasuk random forest menganggap bahwa kelas minoritas ini sebagai noise atau outlier yang akan diabaikan, sehingga hasil prediksi klasifikasi cenderung pada kelas mayoritas. Akibatnya akurasi dari prediksi untuk kelas

minoritas akan jauh lebih kecil dibandingkan kelas mayoritas. Pembobotan dilakukan untuk setiap kelas dengan bobot tertinggi diberikan pada kelas minoritas. Jantung adalah salah satu organ tubuh yang berperan sangat penting bagi manusia karena berfungsi memompa darah ke seluruh tubuh. Menurut data organisasi Kesehatan dunia (WHO) tahun 2019 bahwa lebih dari 17 juta orang di dunia meninggal akibat penyakit jantung dan pembuluh darah yang merupakan penyebab kematian tertinggi di dunia. Terjadinya penyakit jantung karena penyempitan dan penghambatan pembuluh darah jantung sehingga penyaluran oksigen dan nutrisi ke jantung berkurang. Jika gejala penyakit jantung diketahui maka dapat dilakukan antisipasi (2). Dengan memprediksi resiko penyakit jantung dapat mengetahui seseorang terkena penyakit jantung atau tidak dari gejala yang ditimbulkan. Penelitian di bidang ilmu data dan kecerdasan buatan telah memberikan kontribusi yang signifikan dalam pemahaman dan prediksi penyakit, termasuk penyakit jantung. Dalam sebuah penelitian sebelumnya telah dilakukan penelitian tentang prediksi tahap awal penyakit jantung menggunakan teknik data mining dengan algoritma klasifikasi Random Forest mendapatkan hasil 97,88%, lalu ada juga yang menggunakan perbandingan metode Preprocessing pada Algoritma Support Vector Machine (SVM) dan Modified Balanced Random Forest (MBRF) pada penyakit diabetes dan mendapatkan Hasil (SVM) dengan angka 91,48% sedangkan (MBRF) sampai 97,8% (3). Maka dari itu pada penelitian skripsi dengan teknik data mining, penulis ingin membuat sistem prediksi tahap awal penyakit Jantung menggunakan algoritma Random Forest. diharapkan pada penelitian ini bisa mendapatkan nilai akurasi yang lebih tinggi daripada penelitian sebelumnya.

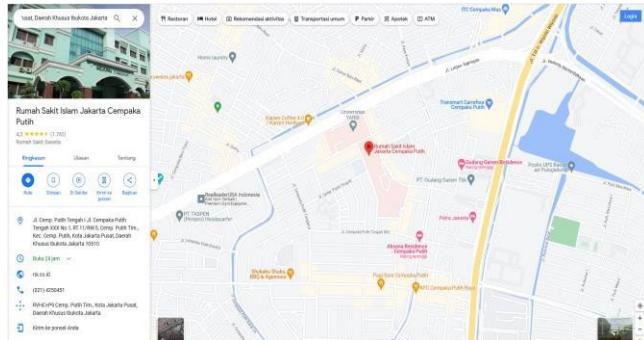
STATE OF THE ART

No	Judul	Hasil
1	Metode Weighted Random Forest dalam Klasifikasi Prediksi Kelangsungan Hidup Pasien Gagal Jantung [4]	Hasil penelitian menunjukkan bahwa metode klasifikasi Weighted Random Forest lebih baik daripada random forest karena dapat meningkatkan accuracy sebesar 1,61%, F-Measure sebesar 0,92%, dan nilai AUC (area under the curve) sebesar 0,89%. Hasil evaluasi untuk klasifikasi Weighted Random Forest untuk accuracy sebesar 90,32%, precision sebesar 93,22%, recall sebesar 96,49%, F-Measure sebesar 94,82%, dan nilai AUC sebesar 0,5825.

		<p>algoritma random forest dapat melakukan prediksi penyakit diabetes dengan kinerja yang baik. Nilai- nilai evaluasi kinerja algoritma random forest untuk prediksi penyakit diabetes adalah: akurasi sebesar 99.3 %, recall sebesar 99.5%, presisi sebesar 99.1%, F1-score sebesar 99.%, dan AUC sebesar 100%.</p>
3.	<p>Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst [6].</p>	<p>Pada penelitian kali ini menggunakan seleksi fitur menggunakan bestfirst menghasilkan 4 fitur yang sangat berpengaruh yaitu age, enjection fraction, serum creatinene dan time, serta penanganan imbalance class menggunakan model class balancer. Algoritma random forest dengan metode percentage split 80% yang menghasilkan accuracy 91,45%, mean absolute error 0.1874, incorrectly classified instances 8.55%, precision 0.915, recall 0.914, AUC 0.953.</p>

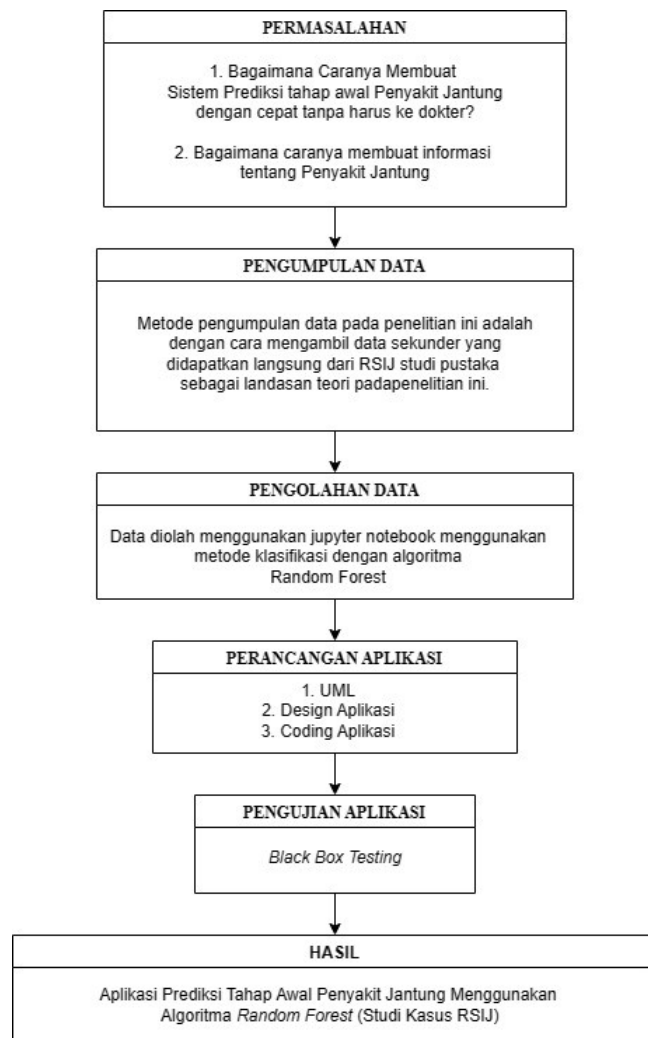
METODOLOGI PENELITIAN

Pada penelitian ini dilakukan di Rumah Sakit Islam Jakarta. Agar penelitian ini berjalan dengan apa yang diharapkan penulis, maka ruang lingkup penelitian hanya di Rumah Sakit Islam Jakarta Cempaka Putih. Waktu Penelitian ini di laksanakan pada 10 Oktober 2023 - 10 Desember 2023.



Gambar 1 Lokasi Penelitian

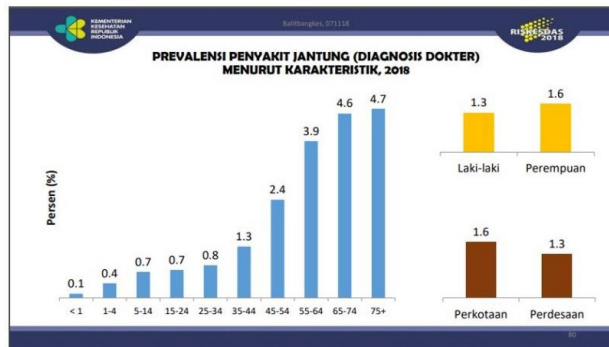
Kerangka Pemikiran



Gambar 2 Kerangka Pemikiran

Teknik Pengumpulan Data

Tahapan pengumpulan data ini menggunakan jenis data sekunder. Data sekunder yang digunakan diperoleh dari website Kaggle atau pasien dengan 9 features. Keseluruhan features dataset penyakit Jantung dapat dilihat pada tabel dibawah ini :



HASIL DAN PEMBAHASAN

A. Pengolahan Data

Projek ini membahas tentang penerapan Random Forest yaitu metode Classification. Tujuan dari projek ini adalah ensemble, yaitu mencari model prediksi klasifikasi dari algoritma Random Forest. Data yang digunakan pada projek ini adalah data pasien penyakit jantung. Dataset ini terdiri dari 767 baris dan 9 kolom.

B. Analisis Sistem

Insight : Korelasi dapat dilihat bahwa yang memiliki nilai positif maka antara feature-feature yang terkait memiliki korelasi positif yang memiliki arti berbanding lurus. Kemudian untuk yang memiliki nilai negatif maka antara feature-feature yang terkait memiliki korelasi negatif memiliki arti berbanding terbalik. Sementara untuk yang nilainya 1 memiliki arti berkorelasi dengan dirinya sendiri.

```
[77] print('Variabel independen:')
print(x.head())

print('\nVariabel dependen:')
print(y.head())

Variabel independen:
  age  sex  cp  trestbps  chol  fbs  restecg  thalach  exang  oldpeak  slope  \
0  63   1   3    145    233   1         0    150     0     2.3     0
1  37   1   2    130    250   0         1    187     0     3.5     0
2  41   0   1    130    204   0         0    172     0     1.4     2
3  56   1   1    120    236   0         1    178     0     0.8     2
4  57   0   0    120    354   0         1    163     1     0.6     2

   ca  thal  total_risk_factor
0  0    1             1
1  1    2             1
2  0    2             0
3  0    2             0
4  0    2             1

Variabel dependen:
0  1
1  1
2  1
3  1
4  1
Name: target, dtype: int64
```

Scaling Data pada x

KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan penerapan algoritma Random Forest dalam prediksi penyakit jantung maka dapat disimpulkan bahwa:

1. Pengimplementasian algoritma Random Forest dapat memprediksi penyakit jantung dengan nilai akurasi sebesar 78%, nilai precision 0.70, recall sebesar 0.52 dan f1-score 0.77
2. Secara umum aplikasi yang dibuat sudah memenuhi kebutuhan untuk memprediksi penyakit sesuai dengan fitur-fitur yang terdapat pada aplikasi tersebut.
3. Aspek Sistem atau Program
 - a. Ditambah level aktor ada yang bertindak sebagai admin dan user
 - b. Ditambah fitur pada sisi user yang harus memiliki menu home, pengertian,
4. Aspek Penelitian Selanjutnya
 - a. Ditambah Algoritma nya minimal 2 algoritma agar mendapat nilai akurasi yang lebih tinggi dan dijadikan sebagai landasan untuk pengembangan sistem.
 - b. Aplikasi ini masih sederhana untuk penelitian selanjutnya diperlukan inovasi ini agar lebih baik lagi.

DAFTAR PUSTAKA

- A. Kadir, Logika Pemrograman Python. Elex Media Komputindo, 2019.
- Cutler, A., Cutler, D. R. & Stevens, J., (2021). Random forests. Machine Learning, 45(1), 157-176.
- Cutler, A., Cutler, D. R. & Stevens, J., (2021). Random forests. Machine Learning, 45(1), 157-176..
- D. Suyanto, "Data Mining untuk klasifikasi dan klusterisasi data," Bandung Inform. Bandung, 2017.
- Direktorat Jenderal Pelayanan Kesehatan. Accessed: Nov. 16, 2023. [Online]. Available: https://yankes.kemkes.go.id/view_artikel/754/ayocari-tahu-apa-itu
- Han, J., Kamber, M., & Pei, J. (2021). Data Mining Concepts and Techniques Edition 3. Burlington: Elsevier.
- Kementerian Kesehatan Republik Indonesia. (2019). Hari Jantung Sedunia (World Heart Day): Your Heart is Our Heart Too. Direktorat Pencegahan dan Pengendalian Penyakit Tidak Menular, (Online). (<http://p2ptm.kemkes.go.id/kegiatan-p2ptm/pusat-/harijantung-sedunia-world-heart-day-your-heart-is-our-heart-too> diakses pada tanggal 2 Maret 2022).
- M. Driscoll, Jupyter Notebook 101. Amerika Serikat, 2018.
- N. C. Aziz and D. A. Anggora, "Implementasi algoritma knn untuk memprediksi potensi penyakit jantung dengan python flask." Universitas Muhammadiyah Surakarta, 2021.
- R. Abdullah, "Pemrograman web untuk pemula, jakarta: PT," Elex Media Komputindo, 2018.
- R. As and M. Shalahudin, "Rekayasa perangkat lunak terstruktur dan berorientasi objek,"

2021.

- T. K. Ho, "Random decision forests," in Proceedings of 3rd international conference on document analysis and recognition, 1995, vol. 1, pp. 278–282.
- Tan, P.N., Steinbach, M. & Kumar, V. (2022). Introduction to Data Mining (First Edition). Pearson Addison Wesley, Boston..
- Utami, Andi Nur Fadhilah, Suwanda. (2021). Penggunaan Estimator Robust Reweighted Minimum Covariance Determinant pada Diagram Kontrol T2 Hotelling untuk Monitoring Penyebaran Covid-19 di Korea Selatan, Jurnal.
- World Health Organization. (2020). Cardiovascular diseases. WHO World Health Organization, (Online), (<https://www.who.int/health-topics/cardiovascular-diseases> diakses pada tanggal 2 Maret 2022).
- Y. Yuliani, "Algoritma Random Forest Untuk Prediksi Kelangsungan Hidup Pasien Gagal Jantung Menggunakan Seleksi Fitur Bestfirst," Infotek J. Inform. dan Teknol., vol. 5, no. 2, pp. 298–306, 2022.
- Zhu, M., Xia, J., Jin, X. & Yan, M. (2017). Class Weights Random Forest Algorithm for Processing Class Imbalanced Medical Data. IEEE Access, 6:4641-4652..