**EKSTRAKSI DAN KLASIFIKASI SINYAL MIT-BIH ARRHYTHMIA DATABASE DENGAN MODEL RANDOM FOREST**

**LAPORAN PROGRAM INTERNSHIP I**

Diajukan untuk memenuhi kelulusan matakuliah Internship

Tahun akademik **2022/2023**

Disusun Oleh:

| Nama | : **M. RIZKY** |
| --- | --- |
| NPM | : 1194021 |



**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA**

**UNIVERSITAS LOGISTIK & BISNIS INTERNASIONAL**

**BANDUNG**

**2022**

***EXTRACTION AND CLASSIFICATION OF MIT-BIH ARRHYTHMIA DATABASE SIGNAL WITH RANDOM FOREST MODEL***

***INTERNSHIP PROGRAM REPORT I***

Submitted to allow Internship course graduation

Academic year **2022/2023**

Created By:

| Name | : **M. RIZKY** |
| --- | --- |
| NPM | : 1194021 |



***APPLIED BACHELOR PROGRAM OF INFORMATICS ENGINEERING***

***UNIVERSITY OF INTERNATIONAL LOGISTICS & BUSINESS***

***BANDUNG***

***2022***

**SURAT PERNYATAAN TIDAK PLAGIAT**

Bandung, 13 November 2022

Sodara yang bertanda tangan dibawah ini:

Nama : **M. RIZKY**

NPM : 1.19.4.021

Judul Laporan : EKSTRAKSI DAN KLASIFIKASI SINYAL MIT-BIH ARRHYTHMIA DATABASE DENGAN MODEL RANDOM FOREST

Menyatakan bahwa saya tidak melakukan tindakan peniru, menjiplak ataupun menyalin skripsi karya ilmiah yang telah ada. Apabila kita terbukti melakukan tindakan tersebut, maka kita bersedia menerima sanksi yang diberikan sesuai dengan ketentuan yang ditetapkan dan berlaku di Program Studi Diploma IV Teknik Informatika Universitas Logistik & Bisnis Internasional (ULBI).





**ABSTRAK**

Di era modern ini, banyak sekali kegiatan – kegiatan operasional maupun kegiatan lainnya yang melibatkan atau menggunakan *Artificial Intelligence (AI).* Tidak bisa di pungkiri lagi bahwa teknologi yang berkembang pesat seperti sekarang ini tentu di buat untuk mempermudah pekerjaan manusia bahkan menggantikan peran manusia. Salah satu bagian dari *Artificial Intelligence (AI)* adalah *Machine Learning*.

*Machine Learning* sebagai metode dalam system kecerdasan buatan yang mampu mengklasifikasikan data yang dimasukkan untuk keperluan dan kebutuhan masing - masing. Banyak aplikasi yang menerapkan *Machine Learning* untuk keperluan klasifikasi data, memprediksi hubungan antar data, membaca pola data, dan banyak implementasi lainnya.

Dalam laporan ini akan dibahas penggunaan keluaran struktur dari *Machine Learning* digunakan untuk mendeteksi sinyal ECG apakah sudah sesuai dengan data yang sudah di tetapkan pada anotasi. Berdasarkan hasil penelitian menunjukkan pemodelan struktur dari *Machine Learning* dapat digunakan untuk mendeteksi kesesuaian dari sinyal baik pemodelan secara terpisah ataupun pemodelan secara gabungan.

**Kata Kunci**: *Artificial Intelligence, Machine Learning*, Model, Klasifikasi, ECG

***ABSTRACT***

*In this modern era, there are many operational activities and other activities that involve or use Artificial Intelligence (AI). It is undeniable that technology that is developing rapidly as it is today is certainly made to facilitate human work and even replace the human role. One part of Artificial Intelligence (AI) is Machine Learning.*

*Machine Learning as a method in an artificial intelligence system that is able to classify the data entered for their respective needs and requirements. Many applications implement Machine Learning for data classification purposes, predicting relationships between data, reading data patterns, and many other implementations.*

*In this report, we will discuss the use of structured output from Machine Learning to detect whether the ECG signal matches the data set in the annotation. Based on the results of the research, it shows that structural modeling from Machine Learning can be used to detect the suitability of signals, either modeling separately or modeling as a combination.*

***Keywords****: Artificial Intelligence, Machine Learning, Model, Classification, ECG*

DAFTAR ISI

[**BAB I PENDAHULUAN**](#_heading=h.30j0zll) **1**

[1.1. Latar Belakang Masalah](#_heading=h.too64mpac2kq) 1

[1.2. Rumusan Masalah](#_heading=h.yj14m1ekaik) 2

[1.3. Tujuan](#_heading=h.r5suew7ver6m) 2

[1.4. Manfaat](#_heading=h.o0k5b6rsza0p) 2

[**BAB II LANDASAN TEORI**](#_heading=h.mlwyr2ii3bnb) **3**

[2.1. MIT-BIH Arrhythmia Database](#_heading=h.rh6w8c1h94mu) 3

[2.2. Electrocardiogram (ECG)](#_heading=h.7eze30a4vyvs) 3

[2.3. Random Forest](#_heading=h.xpt6zioft3c) 4

[**BAB III Analisis Organisasi Perusahaan**](#_heading=h.u0f65ftsanzu) **5**

[**BAB IV. Metodologi Penelitian**](#_heading=h.v1686jjtbpj6) **6**

[**BAB V. Analisis dan Perancangan Sistem**](#_heading=h.4ih2rij7ykfm) **6**

[**BAB VI. Pengkajian dan Evaluasi**](#_heading=h.nsfdh88wlyzj) **6**

[**BAB VII. Penutup**](#_heading=h.kcurrubh86ui) **6**

**DAFTAR ISI**

[**ABSTRAK**](#_heading=h.1fob9te)

[***ABSTRACT***](#_heading=h.3znysh7)

[**KATA PENGANTAR**](#_heading=h.2et92p0)

[**BAB I**](#_heading=h.30j0zll)

[1.1. Latar Belakang](#_heading=h.tyjcwt)

[1.2. Identifikasi Masalah](#_heading=h.3dy6vkm)

[1.3. Tujuan dan Manfaat](#_heading=h.1t3h5sf)

[1.4. Ruang Lingkup](#_heading=h.4d34og8)

[1.5. Penelitian Sebelumnya](#_heading=h.2s8eyo1)

[1.6. Sistematika Penulisan](#_heading=h.17dp8vu)

[**BAB II**](#_heading=h.3rdcrjn)

[2.1. Sistem Rekomendasi](#_heading=h.26in1rg)

[2.1.1. Definisi Sistem](#_heading=h.lnxbz9)

[2.1.3. Definisi Content Based Filtering](#_heading=h.35nkun2)

[2.2. Website](#_heading=h.1ksv4uv)

[2.2.1. Definisi Website](#_heading=h.44sinio)

[**BAB III**](#_heading=h.4i7ojhp)

[3.1. Sejarah Perusahaan](#_heading=h.2xcytpi)

[3.4. Struktur Organisasi dan Job Description Perusahaan](#_heading=h.1ci93xb)

[3.4.1. Bagan Struktur Organisasi](#_heading=h.3whwml4)

[3.5. Deskripsi dan Ruang Lingkup Internship](#_heading=h.qsh70q)

[**BAB IV**](#_heading=h.3as4poj)

[4.1. Diagram Alur Metodologi Penelitian](#_heading=h.1pxezwc)

[4.2. Tahapan – tahapan Diagram Alur Metodologi Penelitian](#_heading=h.49x2ik5)

[**BAB V**](#_heading=h.2p2csry)

[5.1. Kesimpulan](#_heading=h.147n2zr)

[5.2. Saran](#_heading=h.3o7alnk)

[**DAFTAR PUSTAKA**](#_heading=h.23ckvvd)

# BAB I

# PENDAHULUAN

## 1.1. Latar Belakang Masalah

Tidak bisa dipungkiri lagi bahwa perkembangan Artificial Intelligence begitu sangat cepat. Seiring berkembangnya teknologi yang sangat cepat, banyak sekali jenis - jenis teknologi yang bermunculan untuk membantu bahkan menggantikan pekerjaan manusia salah satunya di bidang kesehatan.Salah satu cabang Artificial Intelligence yang sekarang banyak sekali diminati adalah Machine Learning dan sampai saat ini masih terus berkembang pesat di kalangan programmer atau khususnya di dunia IT. Machine Learning sendiri terdiri atas 2 bagian yaitu Machine Learning Supervised dan Machine Learning Unsupervised.

Machine Learning Supervised adalah struktur dari suatu data yang hendak dianalisis telah ditentukan dahulu dan Machine Learning mencari data di struktur tersebut, sedangkan Machine Learning Unsupervised struktur dari suatu data dicari oleh Machine Learning itu sendiri (Pojon, 2017). Salah satu algoritma Supervised Learning yang sering digunakan pada proses klasifikasi adalah algoritma Random Forest (RF) . RF adalah teknik bagging yang memiliki karakteristik signifikan yang berjalan efisien pada dataset besar. Random forest dapat menangani ribuan variabel masukan tanpa penghapusan variabel dan memperkirakan fitur penting untuk klasifikasi [1].

Di dalam dunia medis, teknologi - teknologi banyak sekali diterapkan untuk memenuhi kebutuhan medis itu sendiri seperti AI pendeteksi pasien positif Covid atau tidak dengan memanfaatkan hembusan nafas dari pasien tersebut. Dan masih banyak lagi hal - hal yang bisa kita manfaatkan untuk mengembangkan teknologi di bidang kesehatan salah satunya adalah dengan memanfaatkan sinyal ECG atau Electrocardiography untuk mengklasifikasi penyakit gagal jantung pada pasien. ECG merupakan sebuah informasi sinyal yang digambarkan dalam bentuk diagram yang menampilkan informasi penting mengenai keadaan jantung manusia. Electrocardiography atau ECG adalah rekaman aktivitas listrik yang dihasilkan melalui siklus detak jantung [2]. Pada kegiatan internship yang penulis lakukan akan berfokus pada Ekstraksi dan klasifikasi sinyal MIT-BIH Arrhythmia menggunakan model Random Forest

## 1.2. Rumusan Masalah

1. Bagaimana proses ekstraksi fitur pada sinyal ECG?
2. Algoritma yang cocok untuk proses klasifikasi pada sinyal ECG?
3. Bagaimana proses implementasi model yang akan digunakan untuk mengklasifikasi data pada sinyal ECG?

## 1.3. Tujuan

Dari penjelasan pada bagian latar belakang di atas, dapat kita pahami bahwa penulisan laporan ini memiliki tujuan antara lain:

1. Dapat memahami bagaimana proses ekstraksi fitur pada sinyal ECG
2. Dapat menentukan model yang terbaik untuk proses klasifikasi sinyal ECG
3. Dapat memahami bagaimana implementasi model yang dipilih dalam proses klasifikasi

## 1.4. Manfaat

Manfaat yang bisa didapatkan pada penelitian ini adalah kita dapat memahami bagaimana proses dari kita melakukan ekstraksi ciri dari sinyal ECG lalu kemudian melakukan preprocessing data yang dimana proses tersebut harus kita lakukan agar data yang diproses klasifikasikan menghasilkan output yang akurat dengan hasil yang seharusnya.

# BAB II

# LANDASAN TEORI

## 2.1. *MIT-BIH Arrhythmia Database*

*MIT-BIH* *Arrhythmia* *database* adalah rangkaian uji standar yang umumnya tersedia untuk mengevaluasi aritmia deteksi. Sejak 1980, *database* ini telah digunakan untuk dasar penelitian untuk dinamika jantung di sekitar 500 lokasi di seluruh dunia. Basis data ini sebagian besar digunakan untuk tujuan medis dan penelitian dari deteksi dan analisis aritmia jantung yang berbeda. Basis data ini mencoba menyediakan informasi yang tepat untuk mendeteksi aritmia ventrikel [3]. Aritmia adalah perubahan detak jantung yang tidak normal karena detak jantung yang tidak tepat yang menyebabkan kegagalan dalam pemompaan darah. Aritmia dapat menyebabkan kematian jantung mendadak. Gejala aritmia yang umum adalah denyut prematur, jantung berdebar, pusing, kelelahan, dan pingsan. Aritmia lebih sering terjadi pada orang yang menderita tekanan darah tinggi, diabetes dan arteri koroner penyakit [4]. Sinyal *Electrocardiogram* yang akan digunakan pada kegiatan internship ini diambil dari *MIT-BIH Arrhythmia database*.

## 2.2. *Electrocardiogram (ECG)*

*Electrocardiogram (ECG)* adalah tes medis yang mengukur aktivitas listrik jantung. *ECG* digunakan untuk mendiagnosis dan memantau berbagai kondisi jantung, seperti serangan jantung, aritmia, dan gagal jantung. *ECG* akan merekam aktivitas listrik kecil yang dihasilkan oleh jantung selama periode waktu tertentu dengan menempatkan elektroda pada tubuh pasien [5]. Rekaman *ECG* berisi *noise* dan amplitudo yang bervariasi dari setiap orang sehingga sulit dalam proses mendiagnosis [6]. *Electrocardiogram* *(ECG)* memberikan informasi penting tentang berbagai kondisi manusia [7]. Untuk melakukan *ECG*, petugas kesehatan akan menempelkan tambalan kecil dan lengket yang disebut elektroda ke dada, lengan, dan kaki pasien. Elektroda terhubung ke mesin *ECG*, yang merekam sinyal listrik yang dihasilkan oleh jantung saat bergerak ke seluruh tubuh. Mesin tersebut menghasilkan jejak aktivitas listrik jantung, yang disebut strip *ECG*, yang kemudian diinterpretasikan oleh petugas kesehatan.

*Machine Learning* dapat diterapkan pada analisis data *ECG* untuk meningkatkan akurasi dan efisiensi diagnosis dan pengobatan. Misalnya, algoritma pembelajaran mesin dapat dilatih untuk mengenali pola dalam data *ECG* yang menunjukkan kondisi jantung tertentu. Algoritma ini kemudian dapat digunakan untuk menganalisis data *ECG* secara otomatis dan memberikan rekomendasi diagnosis atau pengobatan. Algoritma pembelajaran mesin dapat dilatih untuk mengklasifikasikan data *ECG* ke dalam kategori yang berbeda, seperti normal atau abnormal, atau untuk mengidentifikasi kondisi jantung tertentu.

Biasanya, klasifikasi sinyal *ECG* memiliki empat fase: preprocessing, segmentasi, ekstraksi fitur, dan klasifikasi. Fase preprocessing terutama ditujukan untuk mendeteksi dan melemahkan frekuensi sinyal *ECG* yang terkait dengan artefak, yang juga biasanya melakukan normalisasi dan peningkatan sinyal. Setelah preprocessing, segmentasi akan membagi sinyal menjadi segmen yang lebih kecil, yang dapat mengekspresikan aktivitas listrik jantung dengan lebih baik [6].

## 2.3. Machine Learning

Pembelajaran Mesin atau Machine Learning merupakan kemajuan teknologi yang penting karena dapat membantu dalam mengambil keputusan dengan mekanisme prediksi dan klasifikasi berdasarkan data yang ada [8]. Berfokus pada performance yang tinggi, teknik pembelajaran mesin atau machine learning diterapkan pada bisnis dengan data yang berkembang pesat. Karena pendekatan desain cocok untuk komunikasi komputasi paralel dan terdistribusi yang berevolusi atau data bisnis yang dinamis dan berkembang kedalam model Machine Learning [9].

Teknologi berbasis komputer modern banyak yang telah menggunakan Pembelajaran Mesin atau Machine Learning. Machine Learning merupakan cabang dari Kecerdasan Buatan atau Artificial Intelligence yang luas dan sudah berkembang pesat saat ini yang memungkinkan komputer untuk belajar dan berkembang secara otomatis tanpa harus diprogram secara eksplisit. Teknologi ini berasal dari mempelajari pengenalan pola dan teori pembelajaran komputasi. Secara umum, metode pembelajaran yang umum digunakan oleh Machine Learning dapat diklasifikasikan menjadi Supervised, Unsupervised, dan Reinforcement Learning [10].

## 2.3.1. Supervised Learning

Supervised Learning merupakan metode Machine Learning untuk menyimpulkan fungsi dari data *train* ada. Algoritma Supervised Learning biasanya berisi kumpulan sampel input (*feature*) dan label yang berkaitan dengan kumpulan data tersebut. Tujuan dari pengklasifikasian adalah untuk menemukan batas yang sesuai yang dapat memprediksi label yang benar pada data *test*. Secara singkat, dalam Supervised Learning memiliki setiap contoh data yang berpasangan yang terdiri dari objek masukan (*input*) dan objek keluaran (*output*) yang diinginkan. Algoritma Supervised Learning menganalisis data *train* dan menghasilkan fungsi (*model*) [11]. Beberapa contoh metode algoritma pada Supervised Learning:

## 2.3.1.1. Random Forest

Random Forest adalah metode Machine Learning yang diperkenalkan pada tahun 2001 oleh Leo Breiman. Metode ini menggunakan serangkaian besar dari Decision Tree dengan korelasi timbal balik yang rendah dan fitur yang dipilih secara acak menggunakan metode bagging (Bootstrap AGGregatING) [12].

Random Forest merupakan salah satu metode pengklasifikasian terbaik dan banyak digunakan untuk regresi dan juga klasifikasi. Random Forest memiliki algoritma yang sederhana sehingga menjadi salah satu pilihan yang menarik untuk mengklasifikasi teks. Selain itu, Random Forest juga memiliki kemampuan untuk mengolah data berdimensi tinggi dan memiliki performa yang tinggi walaupun menggunakan data yang banyak sehingga menjadi salah satu keuntungan menggunakan model ini dibandingkan dengan model Machine Learning lainnya [13].

Pemilihan model ini didasarkan karena pada faktanya bahwa Random Forest secara luas dianggap sebagai salah satu metode Machine Learning yang paling sukses dan banyak digunakan hingga saat ini [14].

## 2.3.1.2. Naive Bayes

Naive Bayes adalah salah satu pengklasifikasi terkemuka yang telah banyak dikutip oleh banyak peneliti dan digunakan di banyak aspek karena kesederhanaannya dan kinerja dari klasifikasi yang nyata [15]. Diantara bermacam-macam teknik atau metode klasifikasi saat ini, pengklasifikasi Naive Bayes (*NB*) berperan penting karena kesederhanaan, traktabilitas dan efisiensinya [16].

Naive Bayes juga merupakan pengklasifikasi yang sangat kompeten dalam banyak aplikasi di dunia nyata. Meskipun Naive Bayes telah menunjukan akurasi klasifikasi yang luar biasa, namun output yang dihasilkan jarang benar dalam kenyataan [17]. Terlepas dari hal itu, pada kenyataanya Naive Bayes bekerja dengan baik diimplementasikan di dunia ini seperti memprediksi waktu, pemfilteran spam, prakiraan cuaca, dan diagnosis medis [18].

Pengklasifikasian Naive Bayes didasarkan pada kombinasi Teorema Bayes dan asumsi independensi atribut. Pengklasifikasi Naive Bayes didasarkan pada asumsi yang disederhanakan bahwa nilai atribut bersifat independen secara kondisional, berdasarkan asumsi nilai target yang diberikan. Pendekatan Bayes untuk klasifikasi kasus baru terdiri dari penetapan nilai target yang paling mungkin, dengan asumsi bahwa ada [19].

## 2.3.1.3. K-Nearest Neighbor

## 2.3.1.4. Support Vector Machine (SVM)

## 2.3.2. Unsupervised Learning

## 2.3.2.1. K-Means

# BAB III Analisis Organisasi Perusahaan

# BAB IV. Metodologi Penelitian

# BAB V. Analisis dan Perancangan Sistem

# BAB VI. Pengkajian dan Evaluasi

# BAB VII. Penutup

DAFTAR PUSTAKA

* + 1. Nayak, D. R., Dash, R., & Majhi, B. (2016). Brain MR image classification using two-dimensional discrete wavelet transform and AdaBoost with random forests. *Neurocomputing*, *177*, 188-197.
    2. Kuila, S., Dhanda, N., & Joardar, S. (2020). Feature Extraction and Classification of MIT-BIH Arrhythmia Database. In Proceedings of the 2nd International Conference on Communication, Devices and Computing (pp. 417-427). Springer, Singapore.
    3. Apandi, Z. F. M., Ikeura, R., & Hayakawa, S. (2018, August). Arrhythmia detection using MIT-BIH dataset: A review. In *2018 International Conference on Computational Approach in Smart Systems Design and Applications (ICASSDA)* (pp. 1-5). IEEE.
    4. Li, T., & Zhou, M. (2016). ECG classification using wavelet packet entropy and random forests. *Entropy*, *18*(8), 285.
    5. Kumar, R. G., & Kumaraswamy, Y. S. (2012). Investigating cardiac arrhythmia in ECG using random forest classification. *International Journal of Computer Applications*, *37*(4), 31-34.
    6. Wasimuddin, M., Elleithy, K., Abuzneid, A. S., Faezipour, M., & Abuzaghleh, O. (2020). Stages-based ECG signal analysis from traditional signal processing to machine learning approaches: A survey. *IEEE Access*, *8*, 177782-177803.