

BAB 1

PENDAHULUAN

1.1. Latar Belakang

Penyakit gagal jantung dapat disebabkan oleh beberapa kondisi penyakit jantung. Menurut WHO, penyakit jantung atau yang disingkat sebagai CVD (*Cardiovascular disease*) adalah kelompok gangguan pada jantung dan pembuluh darah yang termasuk di antaranya: *coronary heart disease*, *cerebrovascular disease*, *rheumatic heart disease*, dan kondisi jantung lainnya. Lebih dari 4 dari 5 kematian atas CVD disebabkan oleh gagal jantung dan *strokes*, dan sepertiga dari angka kematian tersebut merupakan orang yang mati premature dibawah umur 70. Faktor resiko penyakit ini seringkali disebabkan oleh diet yang tidak sehat, kurang berolahraga, serta penggunaan rokok dan alkohol yang berlebih. Sifatnya yang menyerang mendadak menyebabkan dikembangkannya teknologi yang dapat memprevensi. EKG sebagai salah satu teknologi hasil yang dikembangkan berkontribusi memberikan tanda dan informasi akan aktifitas aliran listrik pada jantung.

Pada penelitian ini, tim peneliti memutuskan untuk memprediksi probabilitas penyakit gagal jantung menggunakan Random Forest. Data yang dipakai untuk melatih algoritma Random Forest yang dipakai diambil dari kompilasi beberapa observasi yang mana total akhirnya berjumlah 918 observasi dengan 12 atribut[1]. (fedesoriano. (September 2021). Heart Failure Prediction Dataset. Retrieved [Date Retrieved] from <https://www.kaggle.com/fedesoriano/heart-failure-prediction>). Integrasi 11 fitur yang sama menghasilkan dataset untuk penelitian jantung dengan jumlah data terbesar saat ini. Kecukupan jumlah data ini ditargetkan agar penelitian yang dilakukan akan dataset tersebut memperoleh hasil yang objektif dan dapat diandalkan sebagai model prediksi.

Pemilihan algoritma Random Forest dikarenakan algoritma tersebut adalah salah satu dari peringkat 10 terbesar algoritma *Machine Learning*. Kualitas algoritma yang dipilih diharapkan mampu menghasilkan hasil prediksi yang aktual, akurat dan dapat diandalkan. Penelitian ini juga diharapkan sebagai jembatan antar penelitian deteksi dan ekstraksi fitur EKG yang sebelumnya sering dilakukan yang diharapkan pada akhirnya dapat tercipta suatu model lengkap dari alat hingga klasifikasi dengan desain

user interface yang menarik sehingga mampu dipakai dan diterapkan oleh para tenaga medis maupun untuk *personal use*.

1.2. Rumusan Masalah

Adapun rumusan masalah yaitu: Bagaimana implementasi algoritma Random Forest dalam melakukan prediksi penyakit gagal jantung dan bagaimana nilai peformanya?

1.3. Tujuan dan Manfaat

1.3.1. Tujuan

Adapun tujuan dari penelitian ini adalah untuk menunjukkan kemampuan Random Forest dalam memprediksi penyakit gagal jantung dengan hasil peforma yang baik serta membuat model klasifikasi yang sederhana namun beperforma baik.

1.3.2. Manfaat

Manfaat dari penelitian ini adalah untuk menjembatani penelitian deteksi dan ekstraksi fitur EKG sebelumnya sehingga mampu dimanfaatkan dan dikembangkan untuk tahap selanjutnya hingga produk siap pakai. Hasil penelitian diharapkan mampu diandalkan sebagai pengembangan tahap berikutnya dari penelitian sebelumnya dan bahan acuan penelitian lainnya di masa depa.

1.4. Batasan Masalah

Dalam penelitian ini, terdapat batasan masalah dari permasalahan yang ada, yaitu :

1. Dataset terdiri dari 918 observasi dengan 12 atribut yang merupakan integrasi antar 11 fitur umum dari beberapa observasi, diantaranya: 303 observasi Cleveland, 294 observasi Hungarian, 123 observasi Switzerland, 200 observasi Long Beach VA, dan 270 observasi Stalog (Heart) Data Set. Total keseluruhan terdapat 1190 observasi namun diantaranya terdapat 272 data observasi duplikat / sama.
2. Algoritma yang digunakan adalah Random Forest.
3. Bahasa pemrograman yang digunakan adalah bahasa Python dengan memanfaatkan beberapa *library* yang telah ada, yaitu: *pandas*, *numpy*, *seaborn*, *matplotlib*, *sklearn*, *yellowbrick*, *colorama*, dan *termcolor*.

4. Hasil penelitian difokuskan pada penerapan algoritma Random Forest yang kemudian akan dilakukan validasi dengan teknik *cross validation* dan GridCV. Data independen terdiri dari 11 kolom hasil integrasi fitur penyakit jantung yang umum dan sama. Dataset yang dipakai tidak memiliki data hilang dan data dependen yang akan dijadikan sebagai target adalah kolom HeartDisease yang sudah bertipe data Boolean. Dari keseluruhan dataset terdapat 5 kolom bertipe data kategori dan sisanya bertipe integer. Data kategori tersebut nantinya akan diberi data *dummy* (*dummy variables*) terlebih dahulu.

1.5. Keterbaruan

Penelitian yang berkaitan sebelumnya dimulai dengan penelitian dengan judul “*Abnormalities State Detection from P-Wave, QRS Complex, and T-Wave in Noisy ECG*”. Tujuan penelitian tersebut adalah membangun sistemasi algoritma yang dapat diterapkan kedalam alat portabel. Penelitian tersebut secara spesifik meneliti tentang proses filterasi EKG hingga ekstraksi fitur yang dilakukan pada subjek dengan tiga kondisi (duduk-istirahat, berjalan dan jalan cepat). Hasil penelitian tersebut yang menggunakan FIR dan Butterworth filter menunjukkan data sinyal mentah yang direkam berhasil di filter dengan baik dan desain model ekstraksi fitur QRS complex dapat dengan akurat menandai puncak QRS. Kesimpulan penelitian yang didapat menunjukkan bahwa data EKG dan ekstraksi fitur yang dilakukan belum diterapkan algoritma klasifikasi dan belum diimplementasi kedalam alat portable siap pakai [1].

Penelitian berikutnya dari tahapan yang sama mengubah penerapan dari subjek manusia remaja-dewasa menjadi fetus dalam kandungan (ibu hamil) dengan judul “*An Application of Modified Filter Algorithm Fetal Electrocardiogram Signals with Various Subjects*”. Penelitian tersebut berfokus pada proses deteksi sinyal fetus dalam kandungan untuk mendapat hasil EKG dini. Hasil penelitian menunjukkan proses filterasi noise terhadap FEKG yang direkam namun tidak meneruskannya kedalam tahap ekstraksi fitur maupun klasifikasi EKG [2].

Penelitian yang ditulis peneliti ini diharapkan dapat dijadikan sebagai penyelesaian tahapan proses sinyal EKG yaitu proses klasifikasi EKG berdasarkan data aktual dan data medis dengan diagnosa yang telah dilakukan oleh profesionalnya. Hasil penelitian ini dapat juga digunakan sebagai bahan acuan untuk melanjutkan penelitian

sebelumnya ke tahap selanjutnya. Proses klasifikasi pada proses sinyal EKG adalah yang terpenting untuk dapat memberikan informasi yang dapat digunakan baik pengguna maupun tenaga medis.

Penelitian sebelumnya mengenai klasifikasi EKG masih banyak yang menggunakan dataset aritmia dari MIT-BIH. Salah satu dari penelitian tersebut yang dijadikan sebagai referensi penelitian ini berjudul “Deteksi Arritmia pada Sinyal EKG dengan Deep Neural Network”. Penelitian tersebut menyajikan hasil validasi data akhir sebesar 99.62% dan sensitivitas 98.92%. Algoritma yang digunakan adalah *Dense Neural Network* dengan beberapa nilai parameter yang dicoba. Sumber referensi penelitian tersebut menunjukkan bahwa tidak sedikit penelitian sebelumnya dengan database yang sama telah memperoleh hasil signifikan diatas 95% tersebut. Confusion Matriks hasil penelitian tersebut namun menunjukkan kelemahan dataset yang dipakai dimana lebih banyak data normal dari abnormal pada dataset MIT-BIH [3]. Tujuan dari seluruh penelitian diatas pada intinya adalah menciptakan sistem prevensi terhadap penyakit gagal jantung. Berdasarkan penelitian yang dikaji diatas peneliti berharap dapat menyajikan sistemasi algoritma klasifikasi siap pakai yang dapat diintegrasikan dan diimplementasikan kedalam alat ekg protabel siap pakai.