



HYBRID MACHINE LEARNING MODEL UNTUK MEMPREDIKSI PENYAKIT JANTUNG DENGAN METODE LOGISTIC REGRESSION DAN RANDOM FOREST

Silmi Ath Thahirah Al Azhima¹, Dwicky Darmawan², Nurul Fahmi Arief Hakim³, Iwan Kustiawan⁴, Mariya Al Qibtiya⁵, Nendi Suhendi Syafei⁶

^{1,2,3,4,5} Universitas Pendidikan Indonesia,

⁶ Universitas Padjadjaran

Bandung, Jawa Barat, Indonesia

silmithahirah@upi.edu, dwickydarmawan22@upi.edu, nurulfahmi@upi.edu, iwan_kustiawan@upi.edu,
marialqibtiya@upi.edu, n.suhendi@unpad.ac.id

Abstract

The heart is the main organ that must work properly and regularly. If there is interference, it will be fatal, namely the onset of a heart attack. Heart attack is included in the 10 diseases with a high risk of death. This is caused by stress factors, blood pressure, excessive work, blood sugar, and others. The purpose of this study is to predict heart disease using Machine Learning (ML) algorithms as an early preventive measure on desktop-based information systems. With Machine Learning models, the hybrid model can increase the accuracy value of an ML method that is added to other ML methods. The accuracy value obtained from the Hybrid Model Machine Learning using the Random Forest and Logistic Regression methods is 84.48%, which is an increase of 1.32%.

Keywords: Heart Disease, Hybrid Model, Information System, Logistic Regression, Random Forest

Abstrak

Jantung adalah organ utama yang harus bekerja dengan benar dan teratur. Jika terjadi gangguan, akan berakibat fatal yaitu timbulnya serangan jantung. Serangan jantung termasuk dalam 10 penyakit dengan resiko kematian yang tinggi. Hal itu disebabkan oleh faktor stress, tekanan darah, kerja berlebihan, gula darah dan lainnya. Tujuan penelitian ini adalah untuk memprediksi penyakit jantung menggunakan algoritma *Machine Learning* (ML) sebagai langkah preventif dini pada sistem informasi berbasis desktop. Dengan *Machine Learning models* yaitu *hybrid model* dapat meningkatkan nilai akurasi dari sebuah metode ML yang ditambahkan dengan metode ML lainnya. Nilai akurasi yang didapatkan dari *Hybrid Model Machine Learning* dengan menggunakan metode *Random Forest* dan *Logistic Regression* sebesar 84,48% yang meningkat sebesar 1,32%.

Kata kunci: Hybrid Model, Logistic Regression, Penyakit Jantung, Random Forest, Sistem Informasi

1. PENDAHULUAN

Penyakit jantung *coroner* merupakan salah satu penyakit tidak menular namun berdampak besar. Menurut data WHO, penyakit jantung merupakan salah satu dari 10 penyakit yang menyebabkan kematian terbanyak di dunia sebesar 16% [1] dan sekitar 1 juta orang menderita penyakit jantung [2]. Penyakit jantung mengakibatkan kematian disebabkan oleh tekanan darah, stress, kerja berlebihan, gula darah dan banyak penyebab lainnya. Jantung adalah organ utama yang harus bekerja dengan benar karena berfungsi untuk memompa darah ke seluruh tubuh sehingga oksigen dan zat-zat gizi dapat tersalurkan. Jika jantung tidak bekerja dengan benar, akan sangat mengganggu fungsi organ tubuh

lainnya bahkan beresiko menyebabkan gagal jantung. Gagal jantung ini dapat dicegah sedini mungkin [3].

Dengan adanya perkembangan teknologi, banyak hal yang dapat dilakukan untuk memberikan kemudahan kepada manusia, diantaranya perkembangan bidang ilmu *Artificial Intelligence* (AI). AI dapat dimanfaatkan pada segala bidang seperti *computer vision* yang dapat mendeteksi penyakit dan sistem *autopilot* pada transportasi [4]. Sehingga pada penelitian kali ini akan dibahas mengenai bidang AI yakni *Machine Learning* (ML) yang dapat memprediksi potensi penyakit jantung berdasarkan data-data terkait. Beberapa metode *Machine Learning* yang digunakan untuk

memprediksi penyakit jantung diantaranya *Random Forest classifier* dengan nilai akurasi yang dihasilkan sekitar 83% [5]. Sedangkan pada penelitian lainnya didapatkan hasil akurasi sebesar 86.5% untuk *Logistic Regression*, 80.9% untuk *Random Forest*, 84.3% untuk *Naïve Bayes*, 84.3% algoritma *Gradient Boosting*, dan 79.8% untuk *SVM* [6]. Jika dilihat dari hasil akurasi yang didapatkan, algoritma *Logistic Regression* yang memiliki tingkat akurasi tertinggi dan diikuti oleh algoritma *Random Forest*. Lalu pada penelitian lainnya membahas mengenai tiga metode ML diantaranya *Random Forest*, *Logistic Regression* dan *Decision Tree*. Hasil akurasi tertinggi sebesar 92% dihasilkan oleh algoritma *Logistic Regression* [7].

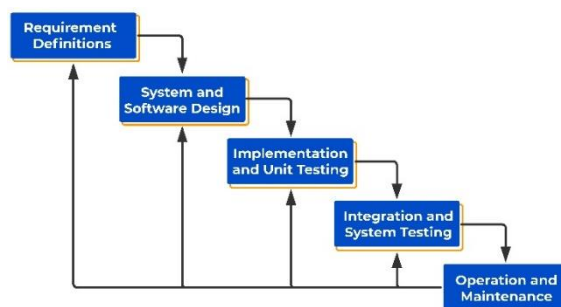
Metode *Machine Learning* digunakan untuk mendapatkan hasil akurasi dan pemahaman yang tinggi pada setiap keputusannya. Sehingga hasil dari prediksi yang dilakukan oleh metode tersebut dapat dipercaya. Untuk meningkatkan nilai akurasi, salah satunya dapat ditambahkan *hybrid model* pada *Machine Learning*. *Hybrid model* ini merupakan penggabungan antara dua metode yang dinilai memiliki tingkat akurasi yang tinggi. Terdapat penelitian yang membandingkan hasil akurasi dari *Machine Learning Explicit*, *Implicit*, dan *Hybrid model* yang diimplementasikan pada diagnosa medis, *e-commerce*, dan *financial decision*. Hasil dari penelitian tersebut dapat disimpulkan bahwa *hybrid model* dapat meningkatkan nilai akurasi prediksi [8]. Penggabungan antara dua metode pada *hybrid model* bertujuan untuk melengkapi hasil akhir berdasarkan hasil sebelumnya. Contoh jika sebelum metode *SVM* ini berjalan, terdapat proses perhitungan *FR*. Maka model tersebut merupakan *FR-SVM hybrid model* [9].

Penelitian yang membahas mengenai *hybrid model* diantaranya, memiliki nilai akurasi sebesar 88,7% dengan *Hybrid Random Forest* dan *Linear Model (HRFLM)* [10]. Selain itu penggunaan *hybrid model* pada *Random Forest* dan *Decision Tree* menghasilkan nilai akurasi 88% [11]. Berdasarkan kedua referensi mengenai *hybrid model* ini dapat disimpulkan bahwa nilai akurasi dapat meningkat sebesar 2,3% dan 7% Ketika metode *Random Forest* ditambahkan dengan metode lainnya. Sehingga pada penelitian ini akan dibahas mengenai implementasi *hybrid model*, namun metode ML yang digunakan pada *hybrid model* yakni *Random Forest* dan *Logistic Regression* untuk memprediksi penyakit jantung yang akan diimplementasikan pada sistem informasi rekam medis. Diharapkan dengan *hybrid model* ini ketika hasil akhir prediksi dari metode *Random Forest* belum lengkap, dapat dioptimalkan oleh metode *Logistic Regression*. Metode *Logistic Regression* dipilih karena pada beberapa penelitian yang sudah dikaji disimpulkan bahwa metode tersebut memiliki nilai akurasi tertinggi. Kelemahan dari metode *Random Forest* yakni komputasi yang tinggi, *overfitting*, pembelajaran yang lambat dan ukuran yang besar dari model yang dihasilkan [12-15]. Kelemahan dari metode *Logistic Regression* yakni tidak dapat memprediksi hasil

yang berkelanjutan, membutuhkan ukuran sampel yang besar agar hasil yang didapat stabil [16].

2. METODE PENELITIAN

Metode pengembangan sistem ini dilakukan dengan metode *Iterative Waterfall* yang dapat dilihat pada Gambar 1. *Model System Development Life Cycle (SDLC) Iterative Waterfall* merupakan pengembangan dari model klasik *waterfall* yang terbagi menjadi lima tahapan yakni, *Requirement*, *Design*, *Implementaion and Unit Testing*, *Integration and Testing*, dan *Operstion and Maintance*. Keuntungan dari menggunakan model ini adalah sederhana, mudah diatur dan ketika terjadi kesalahan atau kekurangan dapat dilakukan evaluasi dengan mengulang tahap sebelumnya sehingga meningkatkan efisiensi dalam pengembangan sistem [17].



Gambar 1. Model *Iterative Waterfall*

2.1 Tahapan Penelitian

Pada tahap *requirement* (kebutuhan), pengumpulan kebutuhan sistem informasi dirinci mulai dari konsep, data dan fitur seperti terlihat pada Tabel 1. Konsep dari sistem informasi yang dibuat menggunakan data rekam medis dan berbasis aplikasi desktop. Sedangkan, Penggunaan *Machine Learning* bertujuan untuk memprediksi penyakit serangan jantung pada seseorang sehingga dapat diprediksi apakah seseorang terindikasi memiliki penyakit serangan jantung atau tidak sebagai langkah preventif lebih dini. Model *Machine Learning* yang digunakan yaitu model *Random Forest* dan *Logistic Regression* dikarenakan berdasarkan studi literatur memiliki tingkat akurasi yang tinggi jika dibandingkan dengan model lainnya.

Tabel 1. Kebutuhan Sistem Informasi

Kebutuhan	Spesifikasi
Data	<ul style="list-style-type: none"> Sistem prediksi penyakit jantung memiliki 19 data terintegrasi dengan <i>database MySQL</i> Terdapat 19 data yang terdiri dari data pribadi dan data prediksi. Data pribadi terdiri dari informasi pasien seperti, nama, usia, tempat tanggal lahir, jenis kelamin, alamat dan nomor telepon. Data prediksi terdiri dari <i>cp</i>, <i>trestbps</i>, <i>chol</i>, <i>fb</i>, <i>restecg</i>, <i>thalach</i>, <i>exang</i>, <i>oldpeak</i>, <i>slope</i>, <i>ca</i>, <i>thal</i> Nama lengkap dan tempat lahir menggunakan tipe data <i>VARCHAR</i>

	<ul style="list-style-type: none"> Tanggal lahir menggunakan tipe data <i>DATE</i> dengan format penulisan tahun - bulan - tanggal Alamat menggunakan tipe data <i>MEDIUM TEXT</i> Usia, jenis kelamin, nomor <i>handphone</i>, <i>cp</i>, <i>trestbps</i>, <i>chol</i>, <i>fbs</i>, <i>restecg</i>, <i>thalach</i>, <i>exang</i>, <i>slope</i>, <i>ca</i>, <i>thal</i> menggunakan tipe data <i>INT</i> <i>Oldpeak</i> menggunakan tipe data <i>DECIMAL</i> Kelas <i>widget</i> yang dipakai ada tiga yaitu <i>entry</i>, <i>date entry</i> dan <i>radiobutton</i>
Fitur	<ul style="list-style-type: none"> Sistem prediksi penyakit jantung memiliki 3 fitur utama yaitu prediksi, memperbarui dan menghapus. Fitur prediksi digunakan untuk memprediksi data yang ada di <i>entry form</i>, menyimpan data ke <i>database</i> dan menyimpan data ke tabel <i>treeview</i>. Fitur memperbarui digunakan untuk memperbarui data yang tersedia Fitur menghapus digunakan untuk menghapus data yang tersedia

Dataset *heart disease* UCI dari Kaggle.com digunakan sebagai data sampel yang terbagi menjadi dua yaitu data latih dan data uji yang dapat dilihat pada Tabel 2 [18]. Dari 303 data yang tersedia dibagi menjadi 14 kategori yaitu:

1. *Age* : Umur;
2. *Sex* : Jenis kelamin (laki-laki dan perempuan);
3. *Cp* : Tipe nyeri dada (*typical angina*, *atypical angina*, *non-anginal pain*, dan *asymptomatic*). Berikut penjelasannya:
 - a. Gejala umum nyeri dada dengan kemungkinan penyumbatan arteri koroner adalah *Typical angina*.
 - b. Gejala tidak rinci, kemungkinan penyumbatan lebih rendah adalah *Atypical angina*.
 - c. Rasa menusuk dan sakit dalam jangka waktu panjang atau pendek adalah *Non-anginal pain*.
 - d. Tidak menunjukkan gejala penyakit adalah *Asymptomatic*.
4. *Trestbps* : dalam satuan mmHG tekanan darah istirahat;
5. *Chol* : serum kolesterol dalam satuan mg/dl;

6. *Fbs* : besar gula darah puasa dalam satuan mg/dl, lebih besar atau kurang dari 120 mg/dl;
7. *Restecg* : hasil elektrokardiografi (ECG) istirahat;
8. *Thalach* : detak jantung maksimum;
9. *Exang* : nyeri dada akibat olahraga;
10. *Oldpeak* : besar segmen ST dari olahraga relatif terhadap kondisi istirahat;
11. *Slope* : besar kemiringan segmen ST pada kondisi latihan puncak atau maksimum;
12. *Ca* : jumlah pembuluh darah utama (0-3);
13. *Thal* : status jantung yang dibagi menjadi 3 diantaranya, 3 = normal, 6 = cacat tetap, 7 = cacat reversibel;
14. *Target* : terindikasi penyakit serangan jantung atau tidak.

Tahap kedua adalah desain sistem yang terbagi menjadi dua yaitu desain sistem informasi dan desain sistem *Machine Learning*. Pada tahap perancangan sistem informasi prediksi penyakit jantung dibuat berbasis aplikasi desktop sehingga tidak membutuhkan jaringan internet untuk menjalankannya. Sistem yang dibuat hanya memiliki satu jendela utama dengan desain sebagai berikut:

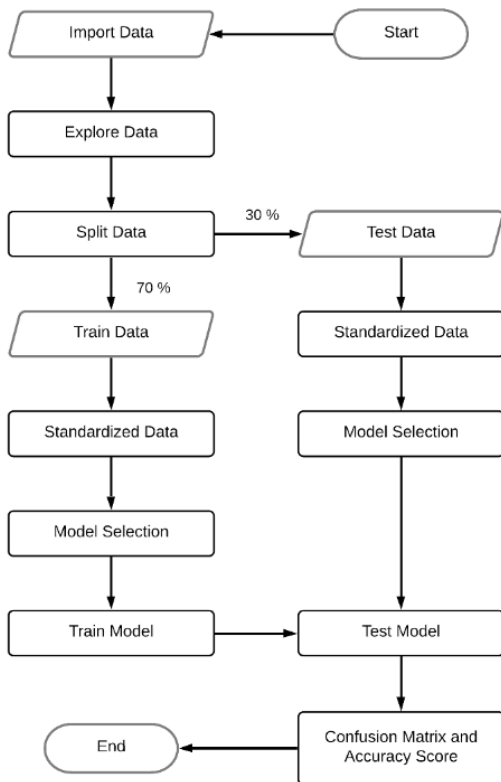
1. 22 label dibagi menjadi tiga yaitu label judul, label data dan label perintah.
2. *Entry form* dibagi menjadi tiga yaitu *entry* berjumlah 10, *date entry* berjumlah 1 dan *radiobutton* berjumlah 19 yang terbagi kedalam 7 grup.
3. 3 tombol klik yaitu *predict*, *update* dan *delete*.
4. Satu tabel *treeview*
5. Dimensi layar penuh yang ukurannya menyesuaikan devais yang dipakai.

Pada tahap desain sistem, *Machine Learning* dibuat dengan model *Random Forest*, *Logistic Regression* dan *Hybrid Model* dengan dataset yang dipakai adalah *Cleveland UCI Heart Disease*. *Hybrid Model* merupakan gabungan model dari *Random Forest* dan *Logistic Regression*. Data yang dipakai sebanyak 303 data dibagi kedalam dua jenis dengan perbandingan 70:30 untuk data latih dan data uji .

Tabel 2. Dataset *heart disease* UCI

Age	Sex	Cp	Trtbps	Chol	Fbs	Restecg	Thalach	Exng	Oldpeak	Slp	Caa	Thall	Output
41	1	1	120	157	0	1	182	0	0	2	0	2	1
38	1	2	138	175	0	1	173	0	0	2	4	2	1
38	1	2	138	175	0	1	173	0	0	2	4	2	1
67	1	0	160	286	0	0	108	1	1.5	1	3	2	0
67	1	0	120	229	0	0	129	1	2.6	1	2	3	0
62	0	0	140	268	0	0	160	0	3.6	0	2	2	0
63	1	0	130	254	0	0	147	0	1.4	1	1	3	0
.
.
.
41	1	1	120	157	0	1	182	0	0	2	0	2	1

Pada tahap ketiga yaitu pemrograman dibagi menjadi dua unit yaitu pemrograman GUI (*Graphical User Interface*) dan pemrograman *Machine Learning* dengan menggunakan bahasa python dan beberapa library yang tersedia. Pada tahap pemrograman *Machine Learning* seperti terlihat pada Gambar 2, langkah pertama yaitu mengimpor data dengan menggunakan library Pandas. Eksplorasi data digunakan untuk mengetahui dimensi dataset, contoh datanya dan mengelompokkan data. Pemisahan data (*Split Data*) berfungsi untuk mengelompokkan data kedalam dua jenis yaitu data latih dan data uji dengan perbandingan 70:30 untuk data latih dan data uji. Standardisasi data merupakan digunakan untuk menstandarisasi nilai fitur atau atribut dari rentang yang berbeda ke dalam rentang tertentu. Pemilihan model digunakan untuk memilih model yang ditentukan yaitu *Logistic Regression*, *Random Forest* dan *Hybrid Model*. Pelatihan dan pengujian dataset merupakan tahap yang digunakan untuk melatih dan menguji dataset kedalam model yang telah ditentukan sebelumnya. Evaluasi model berfungsi untuk mengukur tingkat akurasi dari masing-masing model yang telah ditentukan dimana model terbaik akan digunakan dalam penelitian ini.



Gambar 2. Flowchart pemrograman *Machine Learning*

Tahap keempat yaitu pengujian. Setiap bagian fungsi pada sistem akan diuji kesesuaiannya dengan *requirement* yang sudah ditetapkan. Pemrograman *Machine Learning* akan dimasukkan kedalam pemrograman GUI sebagai satu definisi fungsi yaitu *def predict_data(self)*. Pengujian ini

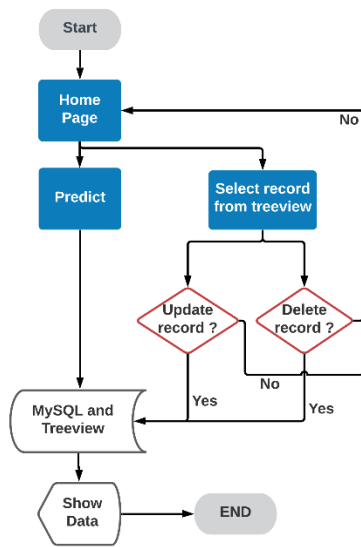
dilakukan dengan metode *blackbox testing*, yaitu pengujian yang hanya melihat *input* dan *output*-nya. Berikut penjelasan mengenai pengujian masing-masing fungsi pada sistem sesuai seperti pada Tabel 3.

Tabel 3. Pengujian Sistem

Data Tes	Tujuan	Hasil yang diharapkan	Hasil sebenarnya
Mengisi semua <i>entry form</i> untuk data pribadi dan data prediksi	Tes tombol <i>predict</i>	<ul style="list-style-type: none"> Boks pesan peringatan “The person has heart disease” akan muncul, data akan tersimpan ke <i>database MySQL</i> dan ditampilkan pada tabel <i>treeview</i> Boks pesan informasi “The person does not have heart disease” akan muncul, data akan tersimpan ke <i>database MySQL</i> dan ditampilkan pada tabel <i>treeview</i> 	Tes tombol <i>predict</i> dengan mengisi semua <i>entry form</i> pada data pribadi dan data prediksi berhasil dilakukan dan hasilnya sesuai yang diharapkan
Memperbarui data yang tersedia dengan memilih salah satu data pada tabel <i>treeview</i>	Tes tombol <i>Update</i>	Data secara otomatis muncul ke <i>entry form</i> . Ketika menekan tombol <i>update</i> akan muncul boks pesan pertanyaan iya dan tidak. Jika menekan iya, data akan diperbaharui dan jika menekan tidak, data tidak akan diperbaharui	Tes tombol <i>Update</i> berhasil dilakukan dan hasilnya sesuai yang diharapkan
Menghapus data yang tersedia dengan memilih salah satu data pada tabel <i>treeview</i>	Tes tombol <i>Delete</i>	Data secara otomatis muncul ke <i>entry form</i> . Ketika menekan tombol <i>delete</i> akan muncul boks pesan pertanyaan iya dan tidak. Jika menekan iya, data akan dihapus dan jika menekan tidak, data tidak akan dihapus	Tes tombol <i>Delete</i> berhasil dilakukan dan hasilnya sesuai yang diharapkan

Alur sistem dimulai dari pengguna menjalankan program dan akan muncul tampilan halaman utama. Fitur yang tersedia dalam sistem informasi yang dibuat yaitu *predict*, *update* dan *delete*. Ketika menekan tombol *predict*, maka sistem akan menjalankan fungsi *Machine Learning* yang terdapat dalam program dan data akan diproses kedalam *database* dan ditampilkan pada tabel *treeview*. Jika akan memperbarui atau menghapus data, maka harus memilih salah satu data dalam tabel *treeview*, data akan otomatis masuk ke *entry form* dan menekan tombol *update* atau *delete*. Data yang diperbarui atau dihapus akan diproses di *database*, dan tabel *treeview* akan menampilkan perubahan yang telah dilakukan seperti pada Gambar 3. Setelah selesai

menggunakan program, pengguna dapat menutup program dan program berakhir.



Gambar 3. Flowchart Sistem

3. HASIL DAN PEMBAHASAN

Pada bagian ini, hasil dari pengembangan sistem akan dibahas. Model *Machine Learning* yang telah dibuat dievaluasi klasifikasi benar dan salah, akurasi dan presisi untuk mendapatkan satu model yang efektif sehingga model tersebut dapat dipakai pada sistem informasi yang telah dibuat. Klasifikasi benar dan salah merupakan kuantitas dasar yang digunakan untuk mengevaluasi kinerja dari model *Machine Learning*. *Output* dari klasifikasi dibagi menjadi 4 (empat) fundamental dasar yaitu (1) contohnya positif dan pengklasifikasi dengan benar mengenalinya seperti itu dapat disebut juga sebagai *True Positive* (TP); (2) contohnya negatif dan pengklasifikasi dengan benar mengenalinya seperti itu dapat disebut juga sebagai *True Negative* (TN); (3) contohnya positif, tetapi pengklasifikasi melabelinya sebagai negatif dapat disebut juga sebagai *False Negative* (FN); dan (4) contohnya negatif, tetapi pengklasifikasi melabelinya sebagai positif dapat disebut juga sebagai *False Positive* (FP) seperti terlihat pada Tabel 4 mengenai klasifikasi benar dan salah.

Tabel 4. Klasifikasi Benar dan Salah

		Label <i>Machine Learning</i>	
		Negatif (0)	Positif (1)
Label Aktudal	Negatif (0)	TN	FP
	Positif (1)	FN	TP

Tingkat kesalahan dari klasifikasi atau E merupakan frekuensi dari kesalahan atau *error* yang dibuat oleh pengklasifikasi (Model *Machine Learning*) dari dataset yang diberikan. Tingkat kesalahan dapat dihitung dengan menggunakan persamaan (1).

$$E = \frac{FP + FN}{FP + FN + TP + TN} \tag{1}$$

Akurasi merupakan frekuensi klasifikasi yang benar dibuat oleh pengklasifikasi dari dataset yang diberikan. Akurasi dapat dihitung dengan cara $Acc = 1 - E$ atau menggunakan persamaan (2).

$$Acc = \frac{TP + TN}{FP + FN + TP + TN} \tag{2}$$

Pada persamaan (3), presisi merupakan persentase dari *True Positive* (TP) diantara semua dataset yang pengklasifikasi diberi label positif. Dengan kata lain, presisi merupakan probabilitas bahwa pengklasifikasi benar ketika memberi label contoh sebagai positif.

$$Pr = \frac{TP}{TP + FP} \tag{3}$$

		Predicted		
		0	1	Σ
Actual	0	104	34	138
	1	17	148	165
Σ		121	182	303

Gambar 4. Evaluasi Model *Random Forest*

Pada Gambar 4 menunjukkan bahwa model *Random Forest* memiliki TP sebanyak 148, TN sebanyak 104, FP sebanyak 34 dan FN sebanyak 17. Dari data diatas kita dapat hitung tingkat kesalahan, akurasi dan presisi untuk model *Random Forest* dengan perhitungan dibawah ini:

$$E = \frac{34 + 17}{34 + 17 + 148 + 104} = 0.1683$$

$$Acc = \frac{148 + 104}{34 + 17 + 148 + 104} = 0.8316$$

$$Pr = \frac{148}{148 + 34} = 0.8131$$

Setelah melalui perhitungan, model *Random Forest* memiliki tingkat kesalahan sebesar 16,83%, akurasi sebesar 83,16% dan presisi sebesar 81,31%.

		Predicted		
		0	1	Σ
Actual	0	89	49	138
	1	18	147	165
Σ		107	196	303

Gambar 5. Evaluasi Model *Logistic Regression*

Pada Gambar 5 menunjukkan bahwa model *Logistic Regression* memiliki TP sebanyak 147, TN sebanyak 89, FP sebanyak 49 dan FN sebanyak 18. Dari data diatas kita dapat hitung tingkat kesalahan, akurasi dan presisi untuk model *Logistic Regression* dengan perhitungan di bawah ini:

$$E = \frac{49 + 18}{49 + 18 + 147 + 89} = 0.2211$$

$$Acc = \frac{147 + 89}{49 + 18 + 147 + 89} = 0.7788$$

$$Pr = \frac{147}{147 + 49} = 0.75$$

Setelah melalui perhitungan, model *Logistic Regression* memiliki tingkat kesalahan sebesar 22,11%, akurasi sebesar 77,88% dan presisi sebesar 75%.

		Predicted		Σ
		0	1	
Actual	0	106	32	138
	1	15	150	165
Σ		121	182	303

Gambar 6. Evaluasi *Hybrid Model*

Pada Gambar 6. menunjukkan bahwa *hybrid model* memiliki TP sebanyak 150, TN sebanyak 106, FP sebanyak 32 dan FN sebanyak 15. Dari data diatas kita dapat hitung tingkat kesalahan, akurasi dan presisi untuk *hybrid model* dengan perhitungan dibawah ini:

$$E = \frac{32 + 15}{32 + 15 + 150 + 106} = 0.1551$$

$$Acc = \frac{150 + 106}{32 + 15 + 150 + 106} = 0.8448$$

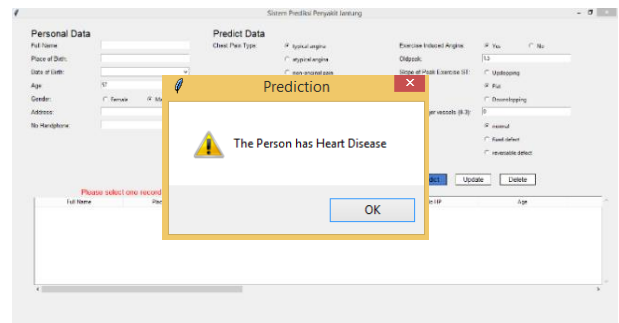
$$Pr = \frac{150}{150 + 32} = 0.8241$$

Setelah melalui perhitungan, *hybrid model* memiliki tingkat kesalahan sebesar 15,51%, akurasi sebesar 84,48% dan presisi sebesar 82,41%. Hasil evaluasi dari masing-masing metode tertera pada Tabel 5.

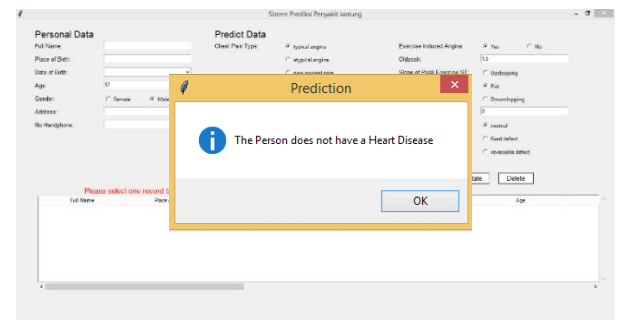
Tabel 5. Hasil Evaluasi Kinerja Model

Nama Model	Tingkat Kesalahan (%)	Akurasi (%)	Presisi (%)	Waktu (s)
<i>Hybrid Model</i>	15,51	84,48	82,41	0.635
<i>Random Forest</i>	16,83	83,16	81,31	0.230
<i>Logistic Regression</i>	22,11	77,88	75	0.100

Berdasarkan Tabel 5 *hybrid model* memiliki nilai akurasi dan presisi yang lebih besar dibandingkan dengan model *Random Forest* dan model *Logistic Regression*. Berdasarkan hasil tersebut dapat disimpulkan bahwa dengan menambah *hybrid model* dapat meningkatkan tingkat akurasi. Ketika hanya menggunakan metode *Random Forest* memiliki akurasi 83,16%, lalu ketika ditambahkan metode *Logistic Regression* melalui *hybrid model* menghasilkan nilai akurasi yang meningkat sebesar 1,32% dan nilai presisi meningkat sebesar 1,1%. Namun peningkatan nilai akurasi dan persisi tidak terlalu besar, dikarenakan hasil prediksi dari *Logistic Regression* lebih rendah dibandingkan dengan metode *Random Forest*. Selain itu, dalam lama waktu proses yang dibutuhkan *hybrid model* lebih lama. Hal tersebut disebabkan karena terdapat dua metode prediksi yang digunakan. Dari hasil analisis dengan menggunakan *hybrid model* membuktikan bahwa nilai akurasi dapat meningkat. Sehingga *hybrid model* dipilih untuk diintegrasikan dengan sistem informasi yang telah dibuat. Algoritma tersebut dijadikan satu definisi fungsi di dalam program sistem informasi yang dinamakan sebagai *predict*. Hasil dari prediksi ini akan ditampilkan dalam bentuk *message box* seperti pada Gambar 7 dan Gambar 8.



Gambar 7. Message Box Warning



Gambar 8. Message Box Info

4. KESIMPULAN

Setelah melakukan penelitian mengenai sistem informasi rekam medis berbasis aplikasi desktop dan Machine Learning untuk memprediksi penyakit serangan jantung dapat disimpulkan bahwa sistem informasi telah berhasil dibuat. Berdasarkan hasil yang diperoleh, sistem informasi prediksi penyakit jantung dapat memprediksi seseorang apakah terkena penyakit jantung atau tidak, menyimpan data pribadi dan data prediksi, menghapus dan

memperbaharui data. Seluruh fitur atau fungsi pada sistem sudah sesuai dengan *requirement* yang diinginkan setelah dilakukan pengujian *blackbox testing*. Sistem informasi ini juga dapat dijalankan di daerah yang tidak ada jaringan internet. Model *Machine Learning* yang digunakan dalam sistem informasi adalah *Hybrid Model* karena memiliki kinerja lebih baik dengan tingkat kesalahan sebesar 15,51%, akurasi sebesar 84,48% dan presisi sebesar 82,41%. Dibandingkan dengan model *Random Forest* dengan tingkat kesalahan 16,83%, akurasi 83,16%, presisi 81,31%. Peningkatan nilai akurasi saat hanya menggunakan metode *Random Forest* menjadi menggunakan *hybrid model* yaitu sebesar 1,32%. Dari hasil berikut disimpulkan bahwa *hybrid model* dapat meningkatkan nilai akurasi.

Ucapan Terima Kasih

Penulis ucapkan terima kasih kepada seluruh tim yang mendukung keberhasilan penelitian ini, terutama kepada Dwicky selaku mahasiswa yang sangat berperan banyak dalam pelaksanaan penelitian ini.

DAFTAR PUSTAKA

- [1] "The top 10 causes of death," *who.int*, 2020.
- [2] 2018 Kementerian Kesehatan RI, "Laporan Nasional RKD2018_FINAL.pdf," *Badan Penelitian dan Pengembangan Kesehatan*. p. 198, 2018.
- [3] "Ini Penyebab Gagal Jantung dan Faktor Risikonya," *alodokter.com*, 2022. .
- [4] I. Arifin, R. F. Haidi, and M. Dzalhaqi, "Penerapan Computer Vision Menggunakan Metode Deep Learning pada Perspektif Generasi Ulul Albab," *J. Teknol. Terpadu*, vol. 7, no. 2, pp. 98–107, 2021.
- [5] V. Chang, V. R. Bhavani, A. Q. Xu, and M. Hossain, "An artificial intelligence model for heart disease detection using machine learning algorithms," *Healthc. Anal.*, vol. 2, no. September 2021, p. 100016, 2022.
- [6] M. Balakrishnan, A. B. Arockia Christopher, P. Ramprakash, and A. Logeswari, "Prediction of Cardiovascular Disease using Machine Learning," *J. Phys. Conf. Ser.*, vol. 1767, no. 1, pp. 1–7, 2021.
- [7] M. Patel, R. Patange, C. Patil, and P. A. Kapoor, "Predicting Heart Disease Using Machine Learning Algorithms .," pp. 61–66, 2022.
- [8] V. Miškovic, "Machine Learning of Hybrid Classification Models for Decision Support," no. June, pp. 318–323, 2014.
- [9] A. Bhattacharya, "WHAT IS HYBRID MACHINE LEARNING AND HOW TO USE IT?," *analyticsinsight.net*, 2022. .
- [10] S. Mohan, C. Thirumalai, and G. Srivastava, "Effective heart disease prediction using hybrid machine learning techniques," *IEEE Access*, vol. 7, pp. 81542–81554, 2019.
- [11] M. Kavitha, G. Gnaneswar, R. Dinesh, Y. R. Sai, and R. S. Suraj, "Heart Disease Prediction using Hybrid machine Learning Model," *Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021*, pp. 1329–1333, 2021.
- [12] Nazarenko, E., Varkentin, V., & Polyakova, T. (2019). Features of Application of Machine Learning Methods for Classification of Network Traffic (Features, Advantages, Disadvantages). 2019 International Multi-Conference on Industrial Engineering and Modern Technologies (FarEastCon).
- [13] Ravishankar, A., Natarajan, S., & Malakreddy, B. (2020). Sylvian-Silva (SFORCE): An Ensembled Boost Approach Towards Machine Learning. IEEEExplore.
- [14] Prajwala, T. (2015). A Comparative Study on Decision Tree and Random Forest Using R Tool. International Journal of Advanced Research in Computer and Communication Engineering., Vol. 4, pp. 196-199,2015.
- [15] Kanervo, A. (2022). Random Forest Application to Tumour Classification. Thesis. University of Turku.
- [16] Choudhury, A., & Gupta, D. (2018). A Survey on Medical Diagnosis of Diabetes Using Machine Learning Techniques. Recent Developments in Machine Learning and Data Analytics, pp. 67–78.
- [12] M. Samadi Gharajeh, "Waterative Model: an Integration of the Waterfall and Iterative Software Development Paradigms," *Database Syst. J.*, vol. X, pp. 75–81, 2019.
- [18] M. R. Sony, "UCI Heart Disease Data," *kaggle.com*, 2020.