

**KLASIFIKASI TAHAP AWAL RISIKO PENYAKIT  
DIABETES MENGGUNAKAN ALGORITMA *RANDOM*  
*FOREST***

**SKRIPSI**

Oleh:

**BANGUN TEGAR SYAHBEKTI**

**201910225346**



**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BHAYANGKARA JAKARTA RAYA**

**2024**

## LEMBAR PERSETUJUAN PEMBIMBING

Judul Proposal : Klasifikasi Tahap Awal Risiko Penyakit Diabetes  
Menggunakan Algoritma *Random Forest*  
Nama Mahasiswa : Bangun Tegar Syahbecti  
Nomor Pokok Mahasiswa : 201910225346  
Program Studi/Fakultas : Informatika / Ilmu Komputer

Jakarta, 17 Februari 2024

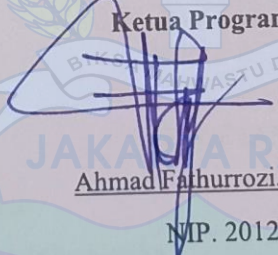
MENYETUJUI,

**Pembimbing I**

  
Ratna Salkiawati, S.T., M.Kom.

NIDN : 0310038006

**Ketua Program Studi**

  
Ahmad Fathurrozi, S.E., M.M.

NIP. 2012486

**Program Studi Informatika  
Fakultas Ilmu Komputer  
Universitas Bhayangkara Jakarta Raya**

2024

## LEMBAR PENGESAHAN

Judul Tugas Akhir : Klasifikasi Tahap Awal Risiko Penyakit Diabetes  
Menggunakan Algoritma *Random Forest*  
Nama Mahasiswa : Bangun Tegar Syahbekti  
Nomor Pokok Mahasiswa : 201910225346  
Program Studi/Fakultas : Informatika/Ilmu Komputer  
Tanggal Lulus Ujian Akhir : 17/02/2024

Jakarta, 21 Februari 2024

MENGESAHKAN,

Ketua Tim Penguji : Adi Muhajirin, S.Kom., M.Kom., M.M.

NIDN : 0318038501

Penguji I : Siti Setiawati, S.Pd., M.Pd.

NIDN : 0313107904

Penguji II : Ratna Salkiawati, S.T., M.Kom.

NIDN : 0310038006

Ketua

Dekan

Program Studi Informatika

Fakultas Ilmu Komputer

Ahmad Fathurrozi, S.E., M.M.

NIP. 2012486

Dr. Dra. Tyastuti Sri Lestari, M.M.

NIP. 140820



LEMBAR PERNYATAAN BUKAN PLAGIASI

Yang bertanda tangan dibawah ini :

Nama : Bangun Tegar Syahbecti  
NPM : 201910225346  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Judul Tugas Akhir : Klasifikasi Tahap Awal Risiko Penyakit Diabetes  
Menggunakan Algoritma *Random Forest*

Dengan ini menyatakan bahwa hasil penulisan skripsi yang telah saya buat ini merupakan **hasil karya saya sendiri dan benar keasliannya**. Apabila dikemudian hari penulisan skripsi ini merupakan plagiat atau penjiplakan terhadap karya orang lain, maka saya bersedia mempertanggungjawabkan sekaligus bersedia menerima sanksi berdasarkan tata tertib di Universitas Bhayangkara Jakarta Raya.

Demikian pernyataan ini saya buat dalam keadaan sadar dan tidak dipaksakan dari pihak manapun.

Jakarta, 21 Februari 2024

Penulis



Bangun Tegar Syahbecti

## ABSTRAK

**Bangun Tegar Syahbeki** – 201910225346, Klasifikasi Tahap Awal Risiko Penyakit Diabetes Menggunakan Algoritma *Random Forest*. Informatika Universitas Bhayangkara Jakarta Raya. 2024

Penyakit diabetes menurut Organisasi Kesehatan Dunia (WHO) gaya hidup yang tidak sehat dan pola makan yang tidak teratur merupakan awalnya tahap terkena risiko penyakit diabetes. Untuk mengurangi nilai pengidap atau kematian karena diabetes maka sangat penting untuk mengklasifikasikan individu untuk secepatnya melakukan tes medis untuk mengetahui positif atau negatif diabetes. Dengan metode *Random Forest* hasil nilai akurasi akan lebih akurat dengan terbuatnya banyak pohon lalu dilakukan majority vote untuk pengambilan Keputusan. Maka dari itu tujuan dari penelitian ini adalah untuk menjelaskan prediksi kemungkinan diabetes secara dini lalu mengklasifikasikan penyakit diabetes tahap awal menggunakan algoritma *Random Forest*. Data yang dikumpulkan melalui Kaggle untuk melakukan penelitian ini. Klasifikasi adalah Teknik data mining yang menetapkan kategori pada umpulan data. Dengan metode *Random Forest* nilai yang diraih dalam mencari akurasi sebesar 0,67%, dan untuk nilai presisi 0,5%, *recall* 1, lalu *F1 Score* 0,67%. Berdasarkan nilai nilai ini, Penelitian ini menunjukkan bahwa metode *Random Forest* memiliki nilai akurasi yang cukup baik.

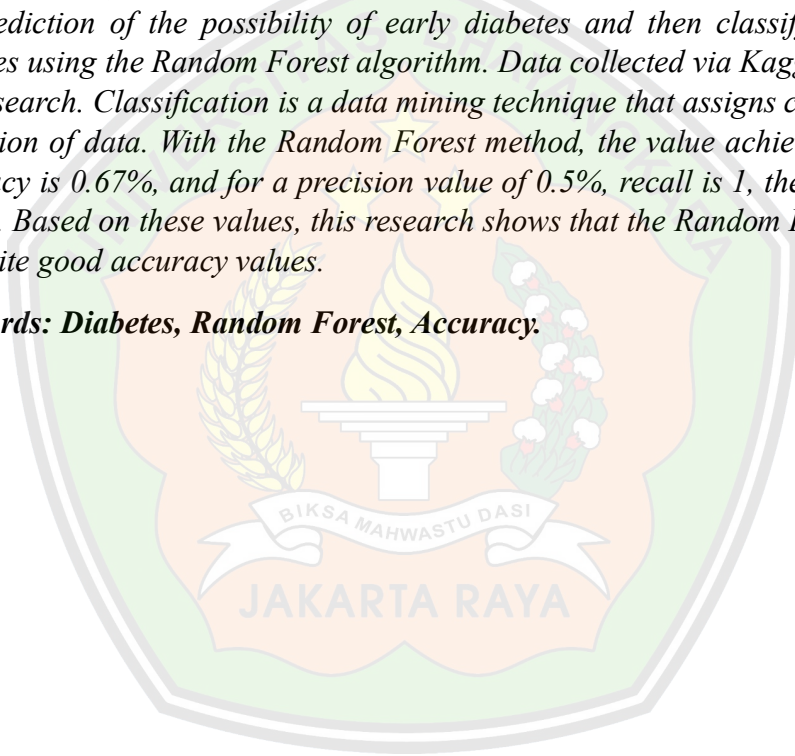
**Kata Kunci:** Diabetes, *Random Forest*, Akurasi.

## ***ABSTRACT***

**Bangun Tegar Syahbekti** – 201910225346, *Early Stage Classification of Diabetes Risk Using the Random Forest Informatics Algorithm, Bhayangkara University, Jakarta Raya.*

*According to the World Health Organization (WHO), an unhealthy lifestyle and irregular eating patterns are the initial stages of developing the risk of diabetes. To reduce the number of sufferers or deaths due to diabetes, it is very important to classify individuals to immediately carry out medical tests to determine whether they are positive or negative for diabetes. With the Random Forest method, the accuracy value results will be more accurate by creating many trees and then using a majority vote for decision making. Therefore, the aim of this research is to explain the prediction of the possibility of early diabetes and then classify early stage diabetes using the Random Forest algorithm. Data collected via Kaggle to conduct this research. Classification is a data mining technique that assigns categories to a collection of data. With the Random Forest method, the value achieved in finding accuracy is 0.67%, and for a precision value of 0.5%, recall is 1, then F1 Score is 0.67%. Based on these values, this research shows that the Random Forest method has quite good accuracy values.*

**Keywords:** *Diabetes, Random Forest, Accuracy.*



**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASIKARYA**  
**ILMIAH UNTUK KEPENTINGAN AKADEMIK**

---

Sebagai sivitas akademik Universitas Bhayangkara Jakarta Raya, saya yang bertanda tangan di bawah ini :

Nama : Bangun Tegar Syahbecti  
NPM : 201910225346  
Program Studi : Informatika  
Fakultas : Ilmu Komputer  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Bhayangkara Jakarta Raya **Hak Bebas Royalti Non-Esklusif (*Non- Exclusive Royalty-Free Right*)**, atas karya ilmiah saya yang berjudul:

**KLASIFIKASI TAHAP AWAL RISIKO PENYAKIT DIABETES  
MENGUNAKAN ALGORITMA *RANDOM FOREST***

beserta perangkat yang ada (bila diperlukan). Dengan hak bebas royalti non-eksklusif ini, Universitas Bhayangkara Jakarta Raya berhak menyimpan, mengalih mediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya dan mempublikasikannya di Internet atau media lain untuk kepentingan akademik tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemilik hak cipta.

Segala bentuk tuntutan hukum yang timbul atas pelanggaran hak cipta dalam karya ilmiah ini menjadi tanggung jawab saya pribadi

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Jakarta  
Pada Tanggal : 21 Februari 2024  
Yang Menandatangani

  
Bangun Tegar Syahbecti

## KATA PENGANTAR

Segala puji bagi Allah SWT yang telah memberikan anugerah dan kemurahan-Nya sehingga penulis dapat menyelesaikan proposal skripsi dengan judul "Klasifikasi Tahap Awal Risiko Penyakit Diabetes Menggunakan Algoritma *Random Forest*". Saat seorang penulis mengalami kesulitan dan rintangan. Tetapi sejumlah orang-orang yang memberikan doa, bantuan, dukungan dan selalu memberi inspirasi sehingga penyusunan proposal skripsi ini dapat diselesaikan. Karena itulah penulis ingin mengungkapkan rasa terima kasih Kepada:

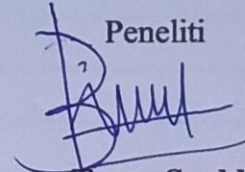
1. Bapak Irjen Polisi (Purn) Prof. Dr. Drs. Bambang Karsono, S.H., M.M. Selaku Rektor Universitas Bhayangkara Jakarta Raya.
2. Ibu Dr. Dra. Tyastuti Sri Lestari, M.M. Sebagai Dekan Fakultas Ilmu Komputer Universitas Bhayangkara Jakarta Raya.
3. Bapak Ahmad Fathurrozi, S.E., M.M.S.I dalam kapasitasnya sebagai Ketua Program Studi Informatika di Universitas Bhayangkara Jakarta Raya.
4. Ibu Ratna Salkiawati, S.T., M.Kom. Sebagai pembimbing saya telah memberikan arahan dan motivasi yang mudah dipahami selama sesi bimbingan. Dan ada juga banyak dosen lain yang telah memberi saya banyak pelajaran yang tidak dapat saya sebutkan satu per satu.
5. Bapak Adi Muhajrin M.Kom., M.M selaku Pembimbing akademik di Program Studi Informatika di Universitas Bhayangkara Jakarta Raya
6. Bapak dan Ibu dosen fakultas Ilmu Komputer atas ilmu dan pengalamannya selama menimba ilmu di prodi informatika.
7. Orang tua selalu memberi semangat dan dukungan serta mendoakan agar diberi kesehatan dan kemudahan menyelesaikan Seminar Proposal setiap hari.
8. Semua teman di Program Studi Informatika, terutama Al Anwari Abbas, Roni Hamonangan, Ditya Kamila Putri, dan Ridwan Wi utomo, selalu memberikan semangat, dukungan, dan bekerja sama dalam perkuliahan.



Penulis menyadari bahwa masih ada kelemahan dalam proposal kripsi ini, baik dari segi konten maupun presentasinya. Karena itu, penulis sangat menghargai kritik dan saran yang diberikan. Penulis berharap agar skripsi ini dapat memberikan manfaat yang signifikan.

Bekasi, 19 Februari 2024

Peneliti



**Bangun Tegar Syahbukti**

201910225346



## DAFTAR ISI

<b>LEMBAR PENGESAHAN .....</b>	<b>iii</b>
<b>ABSTRAK.....</b>	<b>v</b>
<b>ABSTRACT.....</b>	<b>vi</b>
<b>KATA PENGANTAR .....</b>	<b>viii</b>
<b>DAFTAR ISI.....</b>	<b>x</b>
<b>DAFTAR TABEL .....</b>	<b>xii</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR LAMPIRAN.....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1.1 Latar Belakang .....	1
1.2 Identifikasi Masalah .....	4
1.3 Rumusan Masalah .....	5
1.4 Batasan Masalah.....	5
1.5 Tujuan Dan Manfaat.....	5
1.5.1 Tujuan .....	5
1.5.2 Manfaat .....	5
1.6 Sistematika Penulisan .....	6
<b>BAB II LANDASAN TEORI.....</b>	<b>8</b>
2.1 Tinjauan Pustaka .....	8
2.2 Diabetes .....	9
2.3 Klasifikasi dan <i>Data mining</i> .....	10
2.4 <i>Random Forest</i> .....	13
2.5 <i>Confusion Matrix</i> .....	15
2.5 Weka.....	17
2.6 <i>Sample</i> Perhitungan.....	18
2.7 <i>Knowledge Discovery in Database</i> (KDD).....	20
<b>BAB III METODE PENELITIAN .....</b>	<b>23</b>
3.1 Objek penelitian .....	23
3.2 Kerangka Penelitian.....	23
3.4 Data Penelitian .....	24
3.3 Metode Pengumpulan Data .....	25

3.5	<i>Hardware dan Software</i> .....	27
3.6	Metode Penelitian.....	28
<b>BAB IV HASIL DAN PEMBAHASAN</b> .....		<b>29</b>
4.1	Proses KDD .....	29
1.	<i>Data Selection</i> .....	29
2.	<i>Preprocessing Data</i> .....	37
3.	<i>Transformation</i> .....	38
4.	<i>Data mining</i> .....	40
5.	<i>Evaluation and Interpretation</i> .....	66
6.	<i>Knowledge Presentation</i> .....	66
<b>BAB V PENUTUP</b> .....		<b>67</b>
5.1	Kesimpulan.....	67
5.2	Saran.....	67
<b>DAFTAR PUSTAKA</b> .....		<b>68</b>
<b>LAMPIRAN</b> .....		<b>73</b>
<b>UNIVERSITAS BHAYANGKARA JAKARTA RAYA</b> .....		<b>74</b>
<b>FAKULTAS ILMU KOMPUTER</b> .....		<b>74</b>
<b>BIODATA MAHASISWA</b> .....		<b>77</b>

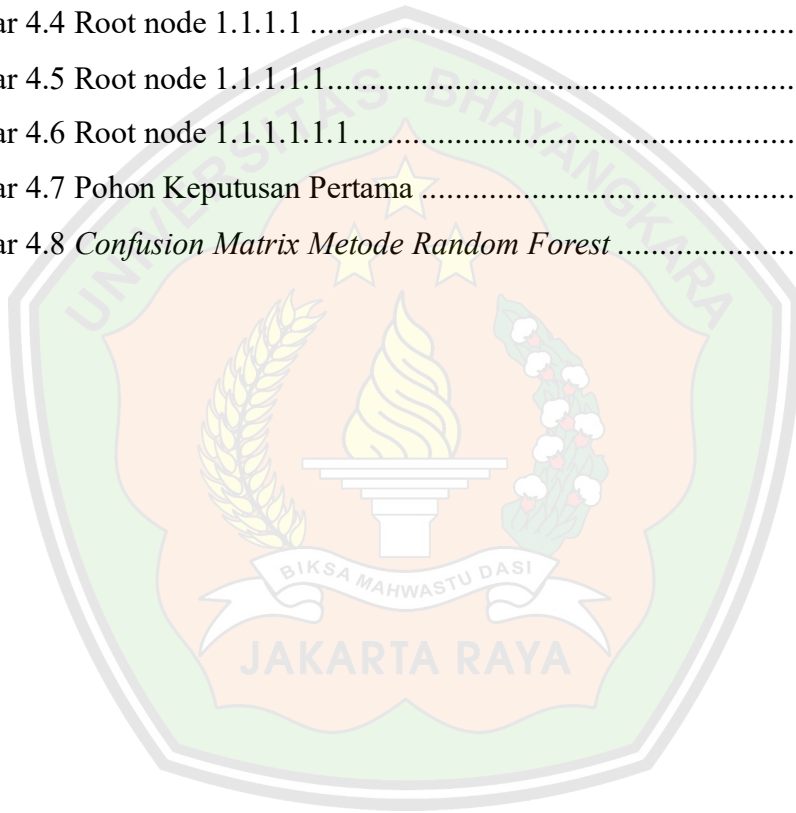


## DAFTAR TABEL

Tabel 2.1 Tinjauan Pustaka.....	8
Tabel 2.2 <i>Confusion Matrix</i> .....	17
Tabel 3.1 Narasumber.....	26
Tabel 3.2 Wawancara.....	26
Tabel 3.2 Spesifikasi <i>Hardware</i> .....	27
Tabel 3.3 Spesifikasi <i>Software</i> .....	28
Tabel 4.1 Tabel Dataset.....	29
Tabel 4.2 Tabel Deskripsi Variabel.....	38
Tabel 4.3 Tabel <i>Data Training</i> 70%.....	39
Tabel 4.4 Tabel <i>Data Testing</i> 30%.....	41
Tabel 4.5 <i>Bootstrap Sampling</i> Pohon Pertama.....	42
Tabel 4.6 Perhitungan <i>entropy</i> dan <i>Gain</i> untu variabel X1, X3, X7 dan X13.....	46
Tabel 4.7 Perhitungan <i>entropy</i> dan <i>Gain</i> untu variabel X5, X8, X10 dan X14.....	47
Tabel 4.8 Atribut yang ada dalam dataset.....	50
Tabel 4.9 Variabel X6, X9,X11,X15.....	50
Tabel 4.10 Bootstrap X7,X10,X12,X14.....	53
Tabel 4.11 Variabel X7,X10,12,14.....	53
Tabel 4.12 Bootstrap X4,X6,X7,X8.....	56
Tabel 4.13 Variabel X4, X6,X7,X8.....	56
Tabel 4.14 Bootstrap Variabel X2,X11,X2,X14.....	59
Tabel 4.15 Variabel X2,X11,X12,X14.....	60
Tabel 4.16 Bootstrap X1, X4, X7, X8.....	63
Tabel 4.17 Variabel X1,X4,X7,X8.....	53

## DAFTAR GAMBAR

Gambar 1.1 Dataset Penelitian .....	3
Gambar 2.1 Implementasi <i>Random Forest</i> .....	15
Gambar 2.2 Tahapan KDD.....	20
Gambar 3.1 Kerangka Penelitian` .....	23
Gambar 4.1 Data <i>Selection</i> .....	45
Gambar 4.2 Root node 1.1 .....	49
Gambar 4.3 Root node 1.1.1 .....	52
Gambar 4.4 Root node 1.1.1.1 .....	55
Gambar 4.5 Root node 1.1.1.1.1.....	59
Gambar 4.6 Root node 1.1.1.1.1.1.....	62
Gambar 4.7 Pohon Keputusan Pertama .....	65
Gambar 4.8 <i>Confusion Matrix Metode Random Forest</i> .....	66



## DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan Skripsi.....	74
Lampiran 2 Kartu Bimbingan .....	75
Lampiran 3 Hasil Plagiarsm.....	76
Lampiran 4 Biodata Mahasiswa.....	77

