

TESIS

**IMPLEMENTASI ALGORITMA RANDOM FOREST DAN K-NEAREST
NEIGHBOR PADA APLIKASI DETEKSI RESIKO TINGGI
PADA KEHAMILAN DI DESA BOLOPLERET
KABUPATEN KLATEN**



Disusun oleh:

Nama : Triana
NIM : 22.55.1200
Konsentrasi : Business Intelligence

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

TESIS

**IMPLEMENTASI ALGORITMA RANDOM FOREST DAN K-NEAREST
NEIGHBOR PADA APLIKASI DETEKSI RESIKO TINGGI
PADA KEHAMILAN DI DESA BOLOPLERET
KABUPATEN KLATEN**

**IMPLEMENTATION OF RANDOM FOREST AND K-NEAREST
NEIGHBOR ALGORITHMS IN HIGH RISK DETECTION
APPLICATIONS IN PREGNANCY IN BOLOPLERET
VILLAGE, KLATEN DISTRICT**

Diajukan untuk memenuhi salah satu syarat memperoleh derajat Magister



Disusun oleh:

Nama : Trlana
NIM : 22.55.1200
Konsentrasi : Business Intelligence

**PROGRAM STUDI S2 TEKNIK INFORMATIKA
PROGRAM PASCASARJANA UNIVERSITAS AMIKOM YOGYAKARTA
YOGYAKARTA**

2024

HALAMAN PENGESAHAN

**IMPLEMENTASI ALGORITMA RANDOM FOREST DAN K-NEAREST
NEIGHBOR PADA APLIKASI DETEKSI RESIKO TINGGI
PADA KEHAMILAN DI DESA BOLOPLERET
KABUPATEN KLATEN**

**IMPLEMENTATION OF RANDOM FOREST AND K-NEAREST NEIGHBOR
ALGORITHMS IN HIGH RISK DETECTION APPLICATIONS
IN PREGNANCY IN BOLOPLERET VILLAGE,
KLATEN DISTRICT**

Dipersiapkan dan Disusun oleh

Triana

22.55.1200

Telah Diujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari NamaHari, tanggal ujian tesis

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 02 Januari 2024
Rektor

Prof. Dr. M. Suyanto, M.M.
NIK. 190302001

HALAMAN PERSETUJUAN

IMPLEMENTASI ALGORITMA RANDOM FOREST DAN K-NEAREST
NEIGHBOR PADA APLIKASI DETEKSI RESIKO TINGGI
PADA KEHAMILAN DI DESA BOLOPLERET
KABUPATEN KLATEN

IMPLEMENTATION OF RANDOM FOREST AND K-NEAREST NEIGHBOR
ALGORITHMS IN HIGH RISK DETECTION APPLICATIONS
IN PREGNANCY IN BOLOPLERET VILLAGE,
KLATEN DISTRICT

Dipersiapkan dan Disusun oleh

Triana

22.55.1200

Telah Dujikan dan Dipertahankan dalam Sidang Ujian Tesis
Program Studi S2 Teknik Informatika
Program Pascasarjana Universitas AMIKOM Yogyakarta
pada hari Selasa, 02 Januari 2024

Pembimbing Utama

Anggota Tim Penguji

Prof. Dr. Ema Utami, M.Kom
NIK. 190302037

Alva Hendi Muhammad, M.Eng., Ph.D
NIK. 190302493

Pembimbing Pendamping

Hanafi, S.Kom., M.Eng., Ph.D
NIK. 190302024

Anggit Dwi Hartanto, M.Kom
NIK. 0513098601

Prof. Dr. Ema Utami, M.Kom
NIK. 190302037

Tesis ini telah diterima sebagai salah satu persyaratan
untuk memperoleh gelar Magister Komputer

Yogyakarta, 02 Januari 2024
Direktur Program Pascasarjana

Prof. Dr. Kusriani, M.Kom.
NIK. 190302106

HALAMAN PERNYATAAN KEASLIAN TESIS

Yang bertanda tangan dibawah ini :

Nama mahasiswa : Triana
NIM : 22.55.1200
Konsentrasi : Business Intelligence

Menyatakan bahwa Tesis dengan judul Implementasi Algoritma *Random Forest* Dan *K-Nearest Neighbor* Pada Aplikasi Deteksi Resiko Tinggi Pada Kehamilan Di Desa Bolopleret Kabupaten Klaten

Dosen Pembimbing Utama : Prof. Dr. Ema Utami, M.Kom
Dosen Pembimbing Pendamping : Anggit Dwi Hartanto, M.Kom

1. Karya tulis ini adalah benar-benar ASLI dan BELUM PERNAH diajukan untuk mendapatkan gelar akademik, baik Universitas AMIKOM Yogyakarta maupun di Perguruan Tinggi lainnya.
2. Karya tulis ini merupakan gagasan, rumusan dan penelitian SAYA sendiri, tanpa bantuan pihak lain kecuali arahan dari Tim Dosen Pembimbing
3. Dalam karya tulis ini tidak terdapat karya atau pendapat orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan disebutkan dalam Daftar Pustaka pada karya tulis ini
4. Perangkat lunak yang digunakan dalam penelitian ini sepenuhnya menjadi tanggung jawab SAYA, bukan tanggung jawab Universitas AMIKOM Yogyakarta
5. Pernyataan ini SAYA buat dengan sesungguhnya, apabila dikemudian hari terdapat penyimpangan dan ketidakbenaran dalam pernyataan ini, maka SAYA bersedia menerima SANKSI AKADEMIK dengan pencabutan gelar yang sudah diperoleh, serta sanksi lainnya sesuai dengan noema yang berlaku di Perguruan Tinggi.

Yogyakarta, 02 Januari 2024

Yang Menyatakan,



Triana

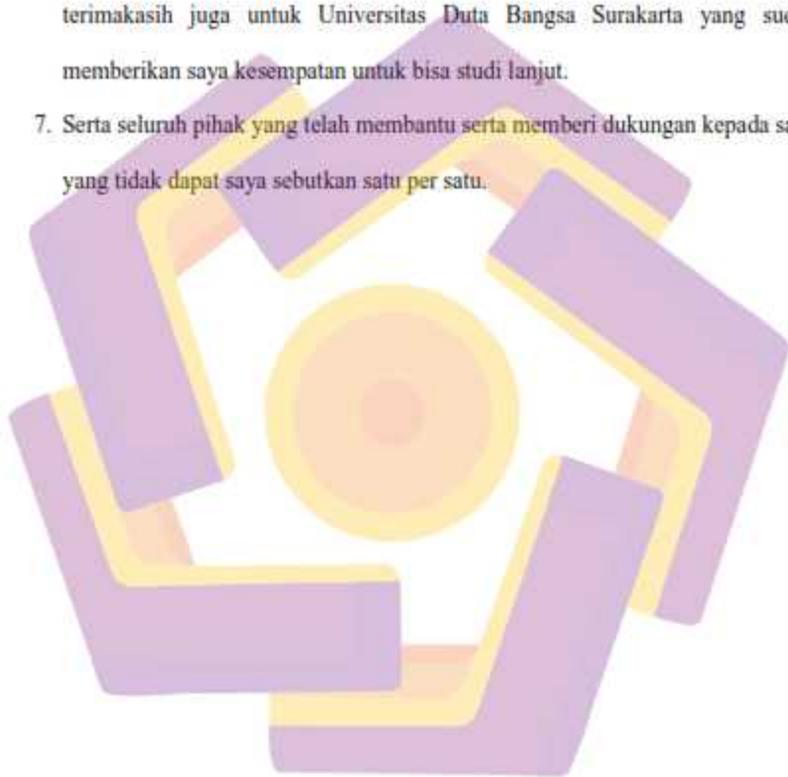
HALAMAN PERSEMBAHAN

Puji syukur saya panjatkan kepada Allah SWT atas limpahan berkat berupa rahmat dan ridho dari-Nya sehingga peneliti dapat menyelesaikan tesis ini sesuai dengan waktu dan hasil yang diharapkan. Saya juga berterima kasih kepada pihak-pihak yang baik secara langsung maupun tidak langsung telah membantu saya dalam menyelesaikan penelitian ini. Oleh karena itu, dengan selesainya penelitian ini, semata-mata saya persembahkan kepada.

1. Orang tua saya Bapak Sumarno dan Ibu Nuryati yang senantiasa mendukung, mendo'akan, memberi semangat, dan mendidik saya menjadi pribadi yang lebih baik.
2. Ibu Prof. Dr. Ema Utami S.Si., M.Kom. selaku dosen pembimbing pertama dan Bapak Anggit Dwi Hartanto, M.Kom selaku pembimbing kedua yang telah membimbing dan memberi saran-saran serta masukan positif yang membantu saya dalam menyusun dan menyelesaikan penelitian ini.
3. Bapak Kepala Desa Bolo Pleret Klaten yaitu Bapak Catur dan ibu Woro selaku bidan serta semua pihak telah membantu saya dalam melakukan penelitian ini.
4. Kepada Arif Eko Fitrianto yang sudah mendukung saya sampai dekat ini, Aghni Rizqi Ni'mal Abdu yang sudah menemani saya dan senantiasa mendengarkan keluh kesah saya saat proses pengerjaan Tesis ini.
5. Bapak dan Ibu Dosen Universitas Amikom Yogyakarta yang telah membekali ilmu serta mendidik sikap dan kedisiplinan saya selama masa perkuliahan saya

baik pada jenjang Sarjana maupun Pasca Sarjana di Universitas Amikom Yogyakarta.

6. TIM Lembaga Jaminan Mutu terkhusus bu Intan Oktaviani yang sudah mendukung dan selalu memberikan motivasi untuk menyelesaikan Studi S2 ini, terimakasih juga untuk Universitas Duta Bangsa Surakarta yang sudah memberikan saya kesempatan untuk bisa studi lanjut.
7. Serta seluruh pihak yang telah membantu serta memberi dukungan kepada saya yang tidak dapat saya sebutkan satu per satu.



HALAMAN MOTTO

1. *“Jangan pergi mengikuti kemana jalan akan berujung. Buat jalanmu sendiri dan tinggalkanlah jejak”.*
2. *“Sesungguhnya Allah tidak akan merubah keadaan suatu kaum sehingga mereka merubah keadaan yang ada pada diri mereka sendiri” (QS.Ar’Ra’d:11)*
3. *“Barangsiapa yang mempelajari ilmu pengetahuan yang seharusnya yang ditunjukkan untuk mencari ridho Allah bahkan hanya untuk mendapatkan kedudukan atau kekayaan duniawi, maka ia tidak akan mendapatkan baunya surgananti pada hari kiamat (riwayat Abu Hurairah radhiallahu anhu)”.*



KATA PENGANTAR

Assalamu'alaikum Wr. Wb.

Puji syukur saya panjatkan kepada Allah SWT atas limpahan berkat, rahmat dan hidayah dari-Nya sehingga peneliti dapat menyelesaikan penulisan tesis berjudul **"Implementasi Algoritma Random Forest Dan K-Nearest Neighbor Pada Aplikasi Deteksi Resiko Tinggi Pada Kehamilan Di Desa Bolopleret Kabupaten Klaten"** dengan tepat waktu.

Tesis ini disusun untuk memenuhi salah satu persyaratan bagi setiap mahasiswa dalam menyelesaikan Program Magister Teknik Informatika di Universitas Amikom Yogyakarta. Tesis ini juga digunakan sebagai bukti bahwa peneliti telah menyelesaikan jenjang studi Magister Teknik Informatika serta memperoleh gelar Magister Komputer.

Dengan selesainya tesis ini, peneliti tidak lupa untuk mengucapkan terima kasih sebesar-besarnya kepada.

1. Bapak Prof. Dr. M. Suyanto, MM., selaku Rektor Universitas Amikom Yogyakarta.
2. Ibu Prof. Dr. Kusriani, M.Kom., selaku Kepala Program Studi Fakultas Magister Ilmu Informatika, yang telah membantu dan membekali ilmu, masuk serta motivasi untuk memulai penelitian ini sebagai proyek tesis.
3. Ibu Prof. Dr. Ema Utami S.Si., M.Kom. selaku dosen pembimbing pertama dan Bapak Anggit Dwi Hartanto, M.Kom. selaku pembimbing kedua yang telah

membimbing dan memberi saran-saran serta masukan positif yang membantu saya dalam menyusun dan menyelesaikan penelitian ini.

4. Keluarga saya dan keluarga besar yang selalu mendukung, memotivasi dan mendo'akan saya dalam keadaan apapun.

Peneliti menyadari bahwa tesis ini masih memiliki banyak kelemahan dan kekurangan. Oleh karena itu, harapan peneliti kepada semua pihak agar berkenan menyampaikan kritik dan saran yang membangun untuk menambah kesempurnaan tesis ini. Semoga keberadaan tesis ini dapat bermanfaat bagi penelitian yang dilakukan berikutnya serta menambah wawasan kita.

Wasalamu'alaikum Wr. Wb.

Yogyakarta, 02 Januari 2024

Penulis

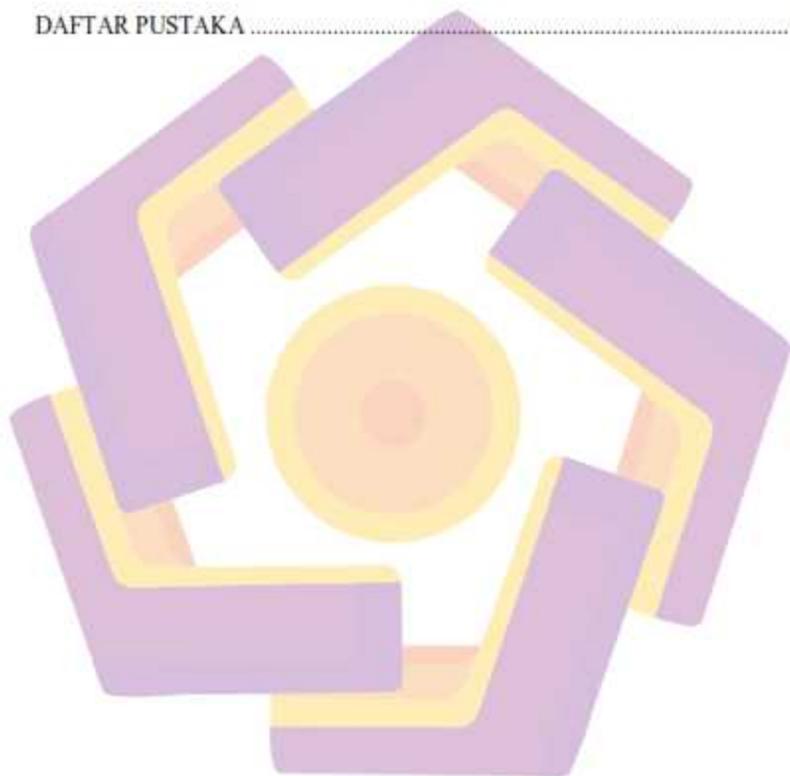
DAFTAR ISI

HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN.....	iii
HALAMAN PERSETUJUAN.....	iv
HALAMAN PERNYATAAN KEASLIAN TESIS.....	v
HALAMAN PERSEMBAHAN.....	vi
HALAMAN MOTTO.....	viii
KATA PENGANTAR.....	ix
DAFTAR ISI.....	xi
DAFTAR TABEL.....	xv
DAFTAR GAMBAR.....	xvi
INTISARI.....	xvii
<i>ABSTRACT</i>	xviii
BAB I PENDAHULUAN	1
1.1. Latar Belakang Masalah.....	1
1.2. Rumusan Masalah.....	6
1.3. Batasan Masalah.....	7
1.4. Tujuan Penelitian.....	8
1.5. Manfaat Penelitian.....	8
BAB II TINJAUAN PUSTAKA	9
2.1. Tinjauan Pustaka.....	9
2.2. Keaslian Penelitian.....	16

2.3. Landasan Teori.....	18
2.3.1. <i>Machine Learning</i>	18
2.3.2. <i>Random Forest</i>	18
2.3.3. Algoritma <i>K-nearest Neighbors</i>	21
2.3.4. <i>Cross Validation</i>	23
2.3.5. <i>Confusion Matrix</i>	24
2.3.6. Index Pengukuran.....	25
BAB III METODE PENELITIAN.....	27
3.1. Jenis, Sifat, dan Pendekatan Penelitian.....	27
3.1.1. Jenis Penelitian.....	27
3.1.2. Sifat Penelitian.....	27
3.1.3. Pendekatan Penelitian.....	27
3.2. Metode Pengumpulan Data.....	28
3.2. Metode Analisis Data.....	28
3.2.1. Alat.....	28
3.2.2. Bahan.....	29
3.3. Alur Penelitian.....	30
3.4. Data Pengumpulan Data	32
3.4.1. Pengumpulan Data.....	32
3.4.2. Pengujian Data.....	33
BAB IV HASIL PENELITIAN DAN PEMBAHASAN	34

4.1. Objek Penelitian.....	34
4.1.1. Deskripsi Singkat Instansi	34
4.1.2. Struktur Pemerintah Desa Bolo Pleret.....	36
4.1.3. Demografi Desa Bolo Pleret Klaten	37
4.2. Penelitian Eksperimental.....	39
4.3. Pengumpulan Data Uji.....	39
4.4. Proses <i>preprocessing</i> data.....	41
4.5. Pengujian Normalitas Data.....	42
4.6. Implementasi Algoritma.....	43
4.7. Jumlah Prediksi Risiko Tinggi dan Rendah	45
4.8. Evaluasi	47
4.8.1. Akurasi	47
4.8.2. <i>Precision</i>	47
4.8.3. <i>Recall</i>	50
4.8.4. <i>F1-Score</i>	52
4.9. Pembahasan.....	53
4.9.1. Identifikasi Faktor Risiko Tinggi	54
4.9.2. Akurasi Model.....	54
4.9.3. Evaluasi	54
4.9.4. Distribusi Hasil Prediksi.....	55
4.9.5. Implikasi dan Rekomendasi	55

4.9.6. Konklusi	56
BAB V PENUTUP	57
5.1. Kesimpulan	57
5.2. Saran	58
DAFTAR PUSTAKA	60



DAFTAR TABEL

Tabel 2.1. Matriks literatur review dan posisi penelitian Implementasi Algoritma Random Forest Dan K-Nearest Neighbor Pada Aplikasi Deteksi Resiko Tinggi Pada Kehamilan Di Desa Bolopleret Kabupaten Klaten.....	16
Tabel 2.2 <i>Confusion Matrix</i>	25
Tabel 2.3 <i>Confusion Matrix 3×3</i>	25
Tabel 4.1. Jumlah Penduduk Desa Bolo Pleret Yang Telah Menyelesaikan Pendidikan.....	35
Tabel 4.2. Jumlah penduduk Desa Bolo Pleret Klaten	38
Tabel 4.3. Jumlah dan Laju Pertumbuhan Penduduk Desa Bolo Pleret Klaten Tahun 2021-2022	38
Tabel 4.4. Jumlah Kepadatan dan Persebaran Penduduk Desa Bolo Pleret Klaten Tahun 2021.....	39
Tabel 4.5. Pengujian normalitas Data kehamilan Desa Bolo Pleret Klaten	43

DAFTAR GAMBAR

Gambar 2.1.	Diagram <i>Random Decision Forest</i>	19
Gambar 2.2.	Contoh <i>Random Forest</i>	20
Gambar 2.3.	Rumus <i>Random Forest</i>	21
Gambar 2.4.	Skema <i>Cross Validation</i>	24
Gambar 3.1.	Kerangka pemikiran	31
Gambar 4.1	Peta Desa Bolo Pleret Klaten Jawa Tengah	34
Gambar 4.2.	Struktur Organisasi Desa Bolo Pleret Klaten	37
Gambar 4.3.	Dataset SQL Kehamilan Desa Bolo Pleret Klaten	40
Gambar 4.4.	Analisis Deskriptif pada Data.....	41
Gambar 4.5.	<i>Classification Report Random Forest</i>	44
Gambar 4.6.	<i>Classification Report K-NN</i>	45
Gambar 4.7.	Hasil Prediksi Untuk Algoritma <i>Random Forest</i>	46
Gambar 4.8.	Hasil Prediksi Untuk Algoritma <i>KNN</i>	46
Gambar 4.9.	Perbandingan Akurasi Model.....	49
Gambar 4.10.	<i>Confusion Matrix</i> Algoritma <i>Random Forest</i> dan <i>K-Nearest Neighbor</i>	50
Gambar 4.11.	Grafik ROC Algoritma <i>Random Forest</i> dan <i>K-Nearest Neighbor</i>	51
Gambar 4.12.	Grafik <i>Precision – Recall Curve</i> Algoritma <i>Random Forest</i> dan <i>K-Nearest Neighbor</i>	53

INTISARI

Peningkatan deteksi dini risiko tinggi pada kehamilan merupakan aspek penting dalam upaya peningkatan kesehatan maternal di wilayah pedesaan. Penelitian ini mengusulkan implementasi algoritma *Random Forest* dan *K-Nearest Neighbor (KNN)* sebagai metode untuk mendeteksi risiko tinggi pada kehamilan di Desa Bolopleter, Kabupaten Klaten.

Data yang digunakan meliputi faktor-faktor risiko seperti usia ibu, riwayat kesehatan, dan kondisi lingkungan. Algoritma *Random Forest* digunakan untuk mengidentifikasi pola kompleks dan hubungan non-linear antara variabel, sementara KNN digunakan untuk mengklasifikasikan risiko berdasarkan kemiripan dengan data latih. Eksperimen dilakukan menggunakan dataset kehamilan di Desa Bolopleter.

Hasil evaluasi menunjukkan bahwa model yang dihasilkan mampu memberikan prediksi risiko tinggi pada kehamilan dengan akurasi yang memuaskan. Implementasi ini dapat menjadi dasar bagi pengembangan aplikasi deteksi dini risiko tinggi pada kehamilan di wilayah pedesaan untuk meningkatkan kesehatan maternal dan neonatal.

Kata kunci: *Random Forest*, Klaten, Resiko Kehamilan

ABSTRACT

Improving early detection of high-risk pregnancies is an essential aspect of enhancing maternal health in rural areas. This research proposes the implementation of Random Forest and K-Nearest Neighbor (KNN) algorithms as methods for detecting high-risk pregnancies in Bolopleter Village, Klaten District.

The data utilized include risk factors such as maternal age, health history, and environmental conditions. The Random Forest algorithm is employed to identify complex patterns and nonlinear relationships between variables, while KNN is utilized for risk classification based on similarity to the training data. Experiments are conducted using pregnancy datasets from Bolopleter Village.

Evaluation results demonstrate that the generated model can provide predictions of high-risk pregnancies with satisfactory accuracy. This implementation could serve as the foundation for developing early detection applications for high-risk pregnancies in rural areas, thereby improving maternal and neonatal health.

Keyword: Random Forest, Klaten, Pregnancy Risk

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Kehamilan beresiko merupakan suatu kehamilan yang memiliki resiko lebih besar dari biasanya (baik dari ibu maupun bayinya). Kehamilan beresiko ini menyebabkan terjadinya bahaya dan komplikasi yang lebih besar terhadap ibu atau janin selama kehamilan, persalinan maupun nifas jika dibandingkan dengan kehamilan, persalinan dan nifas normal. Banyaknya kasus kehamilan beresiko di Kabupaten Klaten, berakibat juga terhadap banyaknya angka kematian ibu dan anak. Dengan tingginya jumlah kehamilan beresiko menjadikan tugas tersendiri bagi para tenaga kesehatan di kabupaten Klaten. Dalam targetnya hanya 20% dari jumlah ibu hamil yang ada di wilayah kabupaten Klaten yang mengalami kehamilan beresiko. Akan tetapi kenyataan di lapangan tidak sesuai dengan data yang telah di targetkan, bisa lebih banyak atau bahkan juga bisa kurang dengan target yang ditetapkan.

Terbatasnya jumlah dokter kandungan yang berada di Dinas Kesehatan Kabupaten Klaten, dan banyaknya jumlah penduduk di Kabupaten Klaten menjadikan kesehatan Ibu dan Anak di Kabupaten Klaten tersebut hanya di layani oleh bidan – bidan yang ada di masing-masing wilayah kerja Dinas Kesehatan Kabupaten Klaten. Permasalahan kehamilan beresiko yang terjadi di wilayah tersebut menyebabkan konsultasi kehamilan beresiko dilakukan kepada dokter kandungan yang berada di pusat pemerintahan. Dengan adanya permasalahan di

atas penulis mencoba untuk mengklasifikasikan kehamilan ibu, apakah ibu tersebut akan masuk kedalam Kelompok 1, Kelompok 2, atau Kelompok 3 dan pada masa kehamilan resiko yang terjadi dalam penelitian ini diklasifikasikan menjadi dua yaitu pendarahan dan eklampsia. Sesuai dengan kriteria yang ada, akan di klasifikasikan menggunakan metode Random Forest dan K-Nearest Neighbor berdasarkan data sample ibu hamil yang telah mengalami kehamilan beresiko. Sehingga setelah proses klasifikasi tersebut diharapkan kehamilan tersebut dapat di awasi, jika di butuhkan akan diberikan rujukan untuk perawatan lebih lanjut sehingga dapat meminimalisir kematian ibu dan anak.

Nurcahyo (2020:42) menyebutkan bahwa resiko tinggi kehamilan adalah suatu kehamilan yang memiliki risiko lebih besar dari biasanya (baik bagi ibu maupun bayinya), dan dapat terjadinya penyakit atau kecacatan maupun kematian sebelum ataupun sesudah persalinan. Risiko tinggi kehamilan ini dapat membahayakan ibu maupun bayi yang sedang dikandung, karena dampaknya yang besar seperti kematian pada ibu atau bayi. Terdapat beberapa faktor penyebab terjadinya risiko tinggi pada ibu hamil seperti ibu hamil yang memiliki riwayat kurang baik pada kehamilan dan persalinan sebelumnya,tinggi badan kurang dari 145 cm, berat badan dibawah berat badan normal, usia dibawah 20 tahun atau lebih dari 35 tahun,memiliki tiga anak atau lebih,jarak antara dua kehamilan yang pendek, memiliki riwayat kurang darah, pendarahan pada kehamilan,tekanan darah,kelainan pada letak janin,dan riwayat penyakit kronik (Sofiyana & Himawan, 2018). Didapatkan data dari World bank, kasus kematian ibu hamil di Indonesia pada tahun 2017 yaitu sebesar 177 setiap 100.000 kelahiran. Jumlah ini lebih besar

jika dibandingkan negara ASEAN lainnya seperti Malaysia yaitu sebesar 29 setiap 100.000 kelahiran dan Thailand yaitu sebesar 37 setiap 100.000 kelahiran (World Health Organization, 2019). Tentu risiko tinggi pada kehamilan menjadi masalah di Indonesia karena telah memakan banyak korban.

Pemerintah Indonesia telah mengupayakan berbagai hal untuk mengurangi angka kematian ibu dan bayi. Upaya pemerintah dalam penurunan angka kematian ibu dan bayi terdapat dalam Permenkes RI no. 39 tahun 2016 tentang pedoman penyelenggaraan program Indonesia sehat. Di dalam pedoman umum disebutkan bahwa salah satu upaya pemerintah untuk mengurangi angka kematian bayi adalah penguatan kader Posyandu. Selain itu, dibentuknya kader di Posyandu juga untuk mencegah dan mengurangi kematian ibu. Adapun tugas dari kader adalah mengenali tanda bahaya pada ibu hamil lebih dini, mendorong ibu hamil untuk mendapatkan pelayanan kesehatan, mengenali tanda bahaya kehamilan, dan pengambilan keputusan untuk meminta tolong pada tenaga kesehatan.

Upaya penguatan kader juga dilaksanakan sampai di tingkat desa sebagai pelaksana kebijakan, salah satunya desa Bolopleret yang terletak di Kabupaten Klaten. Pemerintah desa Bolopleret bersama bidan desa melaksanakan program one client one kader (oce oke). Program one client one kader ini merupakan pelibatan kader dalam upaya pemantauan dan pendampingan ibu hamil yang dalam pelaksanaannya satu kader hanya mendampingi satu ibu hamil yang berbasis dalam kewilayahan. Tugas utama kader dalam program ini terdapat 4 hal yaitu mengingatkan jadwal pemeriksaan, pencatatan dan pendeteksian pada trimester I, pemantauan kesehatan janin, dan komitmen dalam pelaksanaan Program

Perencanaan Persalinan dan Pencegahan Komplikasi(P4K). Sebelumnya program one client one kader (oce oke) telah sukses dilaksanakan di kabupaten Banjarnegara dan masuk dalam Top 99 Inovasi Pelayanan Publik tahun 2019 yang dilaksanakan oleh Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi Republik Indonesia. Pada Puskesmas kecamatan Pejawaran kabupaten Banjarnegara menunjukkan keberhasilan program one client one kader. Pada tahun 2015 sebelum dilaksanakan program one client one kader terdapat 22 kasus kematian bayi di kecamatan Pejawaran, kemudian saat dilaksanakan program one client one kader pada tahun 2016, tidak terdapat kasus kematian bayi sama sekali sampai pada tahun 2017. Keberhasilan program one client one kader ini kemudian dilaksanakan di seluruh Banjarnegara (Khusnul Khotimah, 2021).

Di desa Bolopleret, pelaksanaan program one client one kader dilaksanakan dengan sangat baik. Namun pelaksanaan program ini masih dilakukan secara manual. Pendeteksian masalah dilakukan secara mandiri dan pencatatan masih berupa tulisan tangan. Oleh karena itu, dengan perkembangan teknologi diharapkan dapat membantu mempercepat proses pendeteksian risiko tinggi ibu hamil.

Saat ini di Desa Bolopleret masih dilakukan pendeteksian Risiko tinggi pada kehamilan dengan cara manual. Pendeteksian ini masih dilakukan oleh bidan sendiri. Kader hanya diberikan pelatihan untuk melakukan pencatatan saja. Kemudian hasil catatan yang sudah ditulis oleh kader diberikan kepada bidan. Setelah itu bidan akan membuat catatan hasil pendeteksian dan diserahkan kepada kader. Hal ini akan sangat lama jika kader memberikan hasil catatannya bersama-

sama sedangkan bidan yang bertugas hanya satu orang. Akan sangat cepat untuk pendeteksian jika kader memiliki pengetahuan seperti bidan.

Salah satu disiplin ilmu yang mempelajari metode untuk mengekstrak pengetahuan atau menemukan pola dari suatu data yang besar adalah *Data Mining*. Data mining adalah proses melakukan ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan sebelumnya tidak diketahui, dari suatu data (Witten et al., 2011). *Huge of interesting (non-trivial, implicit, previously unknown and potentially useful) patterns or knowledge from huge amount of data* (Han et al., 2011). *Data mining* sering dianggap sebagai bagian dari *Knowledge Discovery in Database* (KDD) yaitu sebuah proses mencari pengetahuan yang bermanfaat dari data. Selain itu *data mining* juga dikenal dengan nama *knowledge extraction*, *pattern analysis*, *information harvesting*, dan *Business intelligence*.

Ada 5 peranan utama data mining, yaitu: Estimasi, Prediksi, Klasifikasi, Klustering, dan Asosiasi. Algoritma *data mining* yang sering digunakan dalam klasifikasi diantaranya adalah *Random Forest*, *Naive Bayes*, *K-Nearest Neighbor*, C4.5, ID3, CART, *Linear Discriminant Analysis*, *Logistic Regression*, dan lain-lain. Namun, pada Tesis ini penulis hanya akan menggunakan algoritma *Random Forest* dan *K-Nearest Neighbor* untuk mengolah, mengklasifikasikan, serta *mining knowledge* dari *dataset* kehamilan.

Pemilihan penggunaan algoritma *Forest* dan *K-Nearest Neighbor* pada penelitian ini didasarkan pada beberapa alasan, yaitu: Selain kedua algoritma tersebut sama-sama mudah diimplementasikan dan sama-sama dapat memberikan hasil yang baik dalam kasus klasifikasi, kedua algoritma tersebut juga mempunyai

beberapa keunggulan masing-masing. Didasarkan pada pernyataan (Breiman, 2020) yang menyebutkan bahwa algoritma *random forest* dapat mengatasi data training dalam jumlah sangat besar secara efisien dan merupakan metode yang efektif dalam mengestimasi *missing data*. Algoritma *K-Nearest Neighbor* untuk mengklasifikasi objek baru berdasarkan atribut dan *training samples*. Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN (Krisandi et al, 2019).

Dalam *data mining*, penelitian mengenai klasifikasi kehamilan sudah pernah dilakukan oleh peneliti lain. Sebagian besar penelitian tersebut difokuskan pada identifikasi variabel prediktor. Ada banyak penelitian dalam literatur terdahulu yang menjelaskan faktor-faktor apa saja yang dapat mensukseskan proses pengklasifikasian kehamilan. Faktor-faktor tersebut secara umum dibagi menjadi dua, yaitu faktor pernikahan dini dan hamil diluar nikah.

1.2. Rumusan Masalah

Berdasarkan latar belakang penelitian yang menjadi masalah utama dalam penelitian ini adalah :

- a. Berapa tingkat akurasi hasil implemmentasi Algoritma *Random Forest* dan *K-Nearest Neighbor* pada aplikasi deteksi resiko tinggi pada kehamilan?
- b. Berapakan jumlah prediksi Resiko Tinggi Ibu Hamil Di Desa Bolo Pleret Klaten dan Resiko Rendah Ibu Hamil Di Desa Bolo Pleret Klaten yang dihasilkan dengan pengujian Algoritma *Random Forest* dan *K-Nearest Neighbor*?

1.3. Batasan Masalah

Untuk lebih memfokuskan penelitian dan menyederhanakan permasalahan agar dapat diselesaikan dengan pendekatan metode ilmiah, peneliti menentukan ruang lingkup penelitian. Batasan yang digunakan dalam penelitian tugas akhir ini adalah sebagai berikut:

- a. Kegiatan penelitian ini hanya dilakukan pada Desa Bolo Pleret Klaten
- b. Pada penelitian ini sesuai dengan data penduduk yang ada di Desa Bolo Pleret Klaten, terfokus pada Data Kehamilan yang ada di Desa Bolo Pleret Klaten
- c. Dataset yang diambil dari database sistem pendataan masyarakat yang ada di Bolo Pleret Klaten
 1. Kehamilan Risiko Rendah (KRR) dengan jumlah skor terendah 2 Kehamilan yang tidak memiliki masalah atau faktor resiko, fisiologis sampai dengan fase persalinan yang normal dengan ibu dan bayi hidup sehat.
 2. Kehamilan Risiko Tinggi (KRT) dengan jumlah skor sedang 6-10 Pada kehamilan ini memiliki satu resiko atau lebih, baik dari pihak ibu maupun janinnya yang memberi dampak kurang baik bagi ibu maupun janinnya, memiliki resiko kegawatan tetapi tidak darurat dan lebih bisa untuk diatasi tenaga medis.
 3. Kehamilan Risiko Sangat Tinggi (KRST) dengan jumlah skor ≥ 12 Pada kehamilan ini memiliki lebih banyak resiko dan biasanya dibarengi dengan berbagai faktor pendukung, baik dari pihak ibu maupun janinnya yang mempunyai dampak tidak baik bagi ibu maupun janinnya. Pihak medis atau

tenaga medis akan lebih berhati-hati dalam mengambil tindakan penanganannya.

1.4. Tujuan Penelitian

- a. Mengimplementasikan Algoritma *Random Forest* dan *K-Nearest Neighbor* pada Aplikasi resiko tinggi Kehamilan yang terintegrasi *web service*
- b. Pengaruh aplikasi deteksi kehamilan di Desa Bolopleret Kabupaten Klaten
- c. Tingkat akurasi pada aplikasi deteksi kehamilan yang sudah di analisis.

1.5. Manfaat Penelitian

- a. Dengan implementasi pengujian algoritma *random forest* dan *K-Nearest Neighbor* dapat memberikan gambaran hasil resiko tinggi kehamilan di Desa Bolo Pleret Klaten
- b. Implementasi Aplikasi Resiko Tinggi Kehamilan dengan Sistem yang ada di Desa Bolo Pleret Klaten agar mempermudah sinkronisasi data Ibu Hamil

BAB II

TINJAUAN PUSTAKA

2.1. Tinjauan Pustaka

(I Ketut Agung Enriko, dkk., 2020) penelitian dengan judul “*Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient's Health Parameters*” memaparkan bahwa ada beberapa algoritma yang telah digunakan untuk tujuan ini seperti *Naive Bayes*, *Decision Tree*, dan *k-Nearest Neighbor* (KNN). Penelitian ini bertujuan untuk menggunakan data teknik penambangan dalam prediksi penyakit jantung, dengan penyederhanaan parameter yang akan digunakan, sehingga dapat digunakan pada pasien jarak jauh M2M tujuan pemantauan. KNN digunakan dengan pembobotan parameter metode untuk meningkatkan akurasi. Hanya 8 parameter yang digunakan (dari 13 parameter direkomendasikan), karena sederhana dan instan parameter yang dapat diukur di rumah. Hasilnya menunjukkan bahwa akurasi 8 parameter ini menggunakan algoritma KNN adalah baik cukup, bandingkan dengan 13 parameter dengan KNN, atau bahkan lainnya algoritma seperti *Naive Bayes* dan *Decision Tree*.

(Osım Kumar Pal, dkk, 2021) penelitian dengan judul “*Skin Disease Classification: A Comparative Analysis of K-Nearest Neighbors (KNN) and Random Forest Algorithm*” berisikan Algoritma *K-Nearest Neighbors* (KNN) yang digunakan untuk mengklasifikasikan penyakit kulit dengan Model yang diusulkan dirancang berdasarkan KNN dan *Random* Algoritma hutan. Pasien dapat menggunakan model ini untuk mengklasifikasikannya penyakit kulit sebagai

pendeteksi utama, dan dokter juga bisa memastikan penilaiannya dengan menggunakan model yang diusulkan ini. Tradisional diagnosis penyakit kulit adalah mahal dan memakan waktu prosedur. Pendekatan ini dapat bermanfaat bagi individu dengan penyakit kulit yang ingin menghemat uang dan waktu juga seperti menghindari kanker kulit dengan mengidentifikasi kanker pada stadium awal.

(Tahira Islam Trishna, dkk., 2019) penelitian dengan judul "*Detection of Hepatitis (A, B, C and E) Viruses Based on Random Forest, K-nearest and Naïve Bayes Classifier*" menjelaskan mengenai teknik penambangan data secara luas digunakan untuk menganalisis data biomedis. Teknik ini telah memberikan hasil yang efisien dalam prediksi dan klasifikasi keparahan penyakit dan diagnosa penyakit. Hepatitis adalah peradangan hati; itu dapat mempengaruhi orang-orang dari segala usia. Jutaan orang secara global diperkirakan akan terpengaruh oleh hepatitis. Diagnosis hepatitis yang akurat dan dini dapat menyelamatkan banyak pasien. Hepatitis adalah tantangan utama untuk pelayanan kesehatan masyarakat, karena keterbatasan klinis diagnosis penyakit hepatitis pada tahap awal. Menggunakan *naive bayes*, *K-nearest* dan pengklasifikasi Hutan Acak dalam perangkat lunak WEKA menghitung hasilnya. *Naïve bayes* adalah algoritma yang digunakan untuk memecahkan masalah klasifikasi teks. Tetangga *Knearest* adalah mesin sederhana yang diawasi algoritma pembelajaran yang digunakan untuk kedua regresi dan masalah klasifikasi.

(Fangyuan Yang, dkk., 2022), dalam penelitian berjudul "*A hybrid sampling algorithm combining synthetic minority over-sampling technique and edited nearest neighbor for missed abortion diagnosis*" menjelaskan tentang Diagnosis

klinis berdasarkan pembelajaran mesin biasanya menggunakan sampel kasus sebagai sampel pelatihan, dan kegunaannya pembelajaran mesin untuk membangun model prediksi penyakit yang ditandai dengan teks deskriptif manifestasi klinis. Namun, masalah ketidakseimbangan sampel sering terjadi di bidang medis, yang berujung pada penurunan kinerja klasifikasi pembelajaran mesin. Untuk mengatasi masalah ketidakseimbangan sampel dalam kumpulan data medis, kami mengusulkan algoritma pengambilan sampel hybrid menggabungkan teknik pengambilan sampel minoritas sintetik (SMOTE) dan tetangga terdekat yang diedit (ENN). Pertama, SMOTE digunakan untuk pengambilan sampel yang terlewatkan dari dataset aborsi dan diabetes, sehingga jumlah sampel dari keduanya kelas seimbang. Kemudian, ENN digunakan untuk melakukan under-sampling pada *dataset over-sample* untuk menghapus "noisy sample" di mayoritas. Terakhir, *Random Forest* digunakan untuk memodelkan dan memprediksi kumpulan data aborsi dan diabetes yang diambil sampelnya mencapai diagnosis klinis yang akurat. Hasil percobaan menunjukkan bahwa *Random forest* memiliki kinerja klasifikasi terbaik pada aborsi yang terlewatkan dan *dataset* diabetes setelah sampel SMOTE-ENN, dan indeks MCC masing-masing adalah 95,6% dan 90,0%. Selain itu, hasil perbandingan berpasangan dan perbandingan berganda menunjukkan bahwa SMOTE-ENN secara signifikan lebih baik daripada yang lain algoritma pengambilan sampel. *Random Forest* secara signifikan meningkatkan semua indeks pada *dataset* aborsi yang terlewatkan setelah SMOTE-ENN sampel.

(Venkata Sai Kumar Pokala, dkk.,2022), penelitian berjudul "*Analysis and Comparison for Prediction of Diabetic among Pregnant Women using Innovative*

K-Nearest Neighbor Algorithm over Logistic Regression with Improved Accuracy" menjelaskan penerapan Algoritma *K-Nearest Neighbor* digunakan dalam algoritma pembelajaran mesin untuk prediksi diabetes di kalangan ibu hamil untuk mendapatkan akurasi, sensitivitas PEARCE, dan presisi yang lebih baik. Bahan dan untuk memverifikasi kegunaan teknik, peneliti kumpulan data yang dikumpulkan dari kumpulan data gratis yang tersedia seperti Pima Kumpulan data India dari situs web UCI untuk memeriksa diabetes di antara wanita hamil. Ada dua kelompok dalam penelitian ini: *K-Nearest Tetangga* ($N=20$) dan Regresi Logistik ($N=20$), masing-masing memiliki ukuran sampel 40. Kekuatan pre-test 80%, ambang batas 0,05, dan interval kepercayaan 95% digunakan dalam sampel perhitungan ukuran. Hasil: Akurasi, sensitivitas, dan presisi algoritma yang digunakan untuk mengevaluasi kinerja mereka. *K-Terdekat Tingkat akurasi Neighbor* (KNN) sebesar 72,44 persen, namun akurasi Linear Regresi (LR) sebesar 76,67%. Tingkat sensitivitas untuk *K-Nearest Neighbor* adalah 74,42 persen, sedangkan tingkat sensitivitas untuk Linear Regresi (LR) sebesar 76,16 persen. Tingkat presisi *K-Nearest Neighbor* (KNN) adalah 73,75 persen, sedangkan tingkat presisi untuk *Linear Regression* (LR) sebesar 81,87 persen. Tingkat akurasi berbeda nyata $P=0,366$ ($P>0,05$). Kapan dibandingkan dengan algoritma *Innovative K-Nearest Neighbor* Algoritma Regresi Logistik memprediksi klasifikasi yang lebih baik dalam dis meliputi akurasi, sensitivitas, dan presisi untuk mengakses tingkat untuk prediksi diabetes di kalangan wanita hamil.

(I.C. Crockart, dkk.,2021), dalam penelitiannya berjudul "*Classification of intrauterine growth restriction at 34–38 weeks gestation with machine learning*

models". Pembatasan pertumbuhan intrauterin (IUGR) adalah salah satu penyebab paling umum dari kelahiran mati. Tujuan dari penelitian ini adalah untuk mengembangkan model pembelajaran mesin yang akan dapat secara akurat dan konsisten memprediksi apakah taksiran berat janin (EFW) akan berada di bawah persentil ke-10 pada tahap kehamilan 34+0–37 + 6 minggu, dengan menggunakan data yang dikumpulkan pada usia kehamilan 20+0 hingga 23+6 minggu. Rekrutmen calon *Safe Passage Study* (SPS) dilakukan selama 7,5 tahun (2007–2015). Sebuah bagian penting dari penilaian janin adalah rekaman transabdominal non-invasif dari ibu dan janin elektrokardiogram serta kinerja pemeriksaan ultrasonografi untuk bentuk gelombang kecepatan aliran Doppler dan biometri janin pada usia kehamilan 20+0 hingga 23+6 dan 34+0 hingga 37+6 minggu. Beberapa model prediksi dibangun, menggunakan teknik pembelajaran yang diawasi, dan dievaluasi menggunakan metode *Stochastic Gradient Descent*, *k Nearest Neighbours*, *Logistic Regression* dan *Random Forest*. Model akhir bekerja dengan sangat baik di semua metrik evaluasi, khususnya untuk metode Penurunan *Gradien Stochastic*: mencapai rata-rata 93% untuk Akurasi Klasifikasi, Penarikan Kembali, Presisi, dan F1- Skor saat pengambilan sampel acak digunakan dan 91% untuk validasi silang (kedua metode menggunakan kepercayaan 95% dalam terval). Selanjutnya, model tersebut mengidentifikasi Indeks Pulsa Arteri Umbilikal sebagai pengidentifikasi terkuat untuk prediksi IUGR – sesuai dengan literatur. Tiga dari empat metode evaluasi yang digunakan mencapai di atas 90% untuk hasil *True Negative* dan *True Positive*. Analisis ROC menunjukkan tingkat *True Positive* (sumbu y) yang sangat kuat untuk kedua hasil atribut target – nilai AUC sebesar 0,771. Model berperforma sangat

baik di semua metrik evaluasi, menunjukkan ketahanan dan fleksibilitas sebagai model prediktif untuk atribut target biner dari IUGR. Akurasi ini kemungkinan karena nilai tambah dari fitur pra-pemrosesan mengenai detak dan akselerasi janin yang diperoleh, sesuatu yang tidak ada studi multidisiplin sebelumnya. Keberhasilan model prediksi yang diusulkan memungkinkan pengejaran lebih jauh anomali terkait kelahiran, memberikan landasan untuk model yang lebih kompleks dan materi pelajaran yang kurang diteliti.

(Dr.Geetha, dkk., 2021), dalam jurnal penelitiannya yang berjudul *"Prediction Techniques of Heart Disease and Diabetes Disease using Machine Learning"* menjelaskan mengenai Penyakit jantung dan penyakit Diabetes merupakan salah satu penyakit yang banyak diderita. Penyakit ini cukup umum saat ini jadi kami menggunakan atribut berbeda yang dapat berhubungan dengan penyakit ini dengan baik untuk menemukan metode yang lebih baik untuk memprediksi dan kami juga menggunakannya algoritma untuk prediksi. Umumnya, orang-orang di sektor TI menjadi stres karena jadwal dan target yang padat. Jadi, mereka tidak memiliki cukup waktu untuk mengurus kesehatan dan keluarga mereka. Untuk mengatasi hal tersebut, kami telah membuat sebuah *website* Bernama MEDCARE untuk mengumpulkan data sensor dan menghasilkan hasilnya. Terlepas dari bobot ini adalah masalah serius yang ada membuat efek yang signifikan dalam kehidupan setiap orang. Sehingga dalam aplikasi web ini mereka juga bisa melihat status kesehatannya daftar berat badan (Indeks Massa Tubuh). *Random Forest Classifier* dan *K Nearest Neighbor*, algoritma dianalisis berdasarkan kumpulan data faktor risiko. Di sini kumpulan data terlatih dan kasus

uji yang masuk diproses oleh algoritme pembelajaran mesin dan menghasilkan hasil yang sesuai. Lakukan pemeriksaan penunjang pada ramalan penyakit jantung, ekspektasi keganasan payudara dan perkiraan diabetes menggunakan komponen kunci seperti kadar Glukosa, Tekanan Darah, Ketebalan Kulit, BMI dll menyelidiki faktor-faktor ini, Anda mungkin harus mencari penyebaran faktor-faktor ini menggunakan histogram. Pada kesempatan off bahwa mereka lalai untuk menyaring status kesehatan mereka, aplikasi akan menginformasikan kepada karyawan untuk menangani kesehatan mereka.



2.2. Keaslian Penelitian

Tabel 2.1. Matriks literatur review dan posisi penelitian
Implementasi Algoritma Random Forest Dan K-Nearest Neighbor Pada Aplikasi Deteksi Resiko Tinggi Pada Kehamilan Di Desa Bolopleret Kabupaten Klaten

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Suatu atau Kelemahan	Perbandingan
1	<i>Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient's Health Parameters</i>	I Ketut Agung Emiko, Journal of Telecommunication, Electronic and Computer Engineering, 2020	Bertujuan untuk menggunakan data Teknik penambahan dalam prediksi penyakit jantung	Algoritma KNN, Naive Bayes dan Decision Tree menunjukan bahwa 8 parameter yg digunakan cukup baik.	Penelitian tersebut membahas mengenai Prediksi penyakit jantung dengan 13 parameter yang direkomendasikan	Penelitian akan membahas mengenai prediksi nilai resiko ketamtilan pada Aplikasi Deteksi Tinggi Kehamilan di Desa Bolo Pleret Klaten
2	<i>Skin Disease Classification: A Comparative Analysis of K-Nearest Neighbors (KNN) and Random Forest Algorithm</i>	Osim Kumar Pal, International Conference on Electronics, Communications and Information Technology (ICECIT), 2021	Mengetahui klasifikasi penyakit kulit sebagai pendeteksi model untuk pencegahan penyakit kanker kulit.	Dengan menggunakan 2 algoritma yaitu <i>Random Forest</i> dan <i>K-Nearest Neighbors</i> , dengan Langkah pendekatan untuk mengidentifikasi penyakit tersebut	Data sampling yang digunakan adalah dengan prediksi antar waktu melalui data penyakit kulit yang Alangkah lebih baik jika <i>dataset</i> tersebut dikomparasikan dengan <i>dataset</i> yang didapat dari Lembaga terkait sebagai bahan validasi data	Penelitian akan meneliti mengenai dinamika masyarakat response time pada aplikasi deteksi kehamilan di Desa Bolo Pleret Klaten
3	<i>Detection of Hepatitis (A, B, C and E) Viruses Based on Random Forest, K-nearest and Naive Bayes Classifier</i>	Tahira Islam Trishna, 10th ICCNT 2019 July 6-8, 2019, IIT - Kanpur Kanpur, India	Penambahan data secara luas untuk analisis data biomedis dalam efisiensi mengklasifikasikan keparahan penyakit Hepatitis atau peradangan hati.	Penggunaan algoritma <i>Random Forest</i> , <i>K-nearest</i> dan <i>Naive Bayes</i> untuk mengklasifikasikan penyakit Hepatitis.	Data yang diproses hanya data terbaru dalam <i>database</i> penyakit Hepatitis	Data akan langsung terhubung dengan basis data Aplikasi Deteksi Kehamilan melalui <i>web service</i> sehingga sinkronisasi akan berlangsung secara <i>real time</i>
4	<i>A hybrid sampling algorithm combining synthetic minority over-sampling technique and edited nearest neighbor for missed abortion diagnosis</i>	Fangyuan Yang, Kang Wang, Lisha Sun, Mengjiao Zhai1, Jiejie Song1 and Hong Wang (BMC Medical Informatics and Decision Making), 2022	Mendiagnos abortus dengan sample hybrid yang menggabungkan teknik pengambilan sampel minoritas sintetik dan tetangga terdekat			

Tabel Lanjutan Tabel 2.1 Matriks literatur review dan posisi penelitian

No	Judul	Peneliti, Media Publikasi, dan Tahun	Tujuan Penelitian	Kesimpulan	Saran atau Kelemahan	Perbandingan
				Data yang telah diambil dan dipilih dapat diakses berdasarkan kebutuhan untuk pengambilan keputusan di perusahaan. Pada proses ekstraksi dapat menyimpan data perusahaan ke dalam database untuk penggabungan hasil ekstraksi dari sumber data lain dari perusahaan seperti data pakar, data fasilitas dan data konsultasi perencanaan	Untuk mengimplementasikan aplikasi dengan algoritma KNN	Data masyarakat di Desa Bolo Pieret akan terintegrasi ke database milik desa Bolo Pieret Klaten sebagai data center dalam proses pengambilan dataset untuk pengujian data dan visualisasinya
5	<i>Analysis and Comparison for Prediction of Diabetic among Pregnant Women using Innovative K-Nearest Neighbor Algorithm over Logistic Regression with Improved Accuracy</i>	Venkata Sai Kumar Pokala, Neelam Sanjeev Kumar (Department of Biomedical Engineering), 2022	Menjelaskan tentang Analisis dan Perbandingan untuk Prediksi Diabetes diantara Ibu Hamil menggunakan Inovatif Algoritma K-Nearest Neighbor atau Regresi Logistik dengan Akurasi yang ditingkatkan	Algoritma yang diterapkan pada bidang kesehatan untuk perhitungan Prediksi Diabetes diantara ibu hamil menggunakan Algoritma, dan data secara acara yang diimplementasikan dalam beberapa masalah.	Baru ada 1 algoritma yang digunakan, dan tidak ada algoritma perbandingan pada penelitian ini	Penelitian yang sedang dikembangkan ada 2 algoritma dalam proses identifikasi resiko tinggi pada kehamilan, dimana algoritma tersebut nanti nya akan dibandingkan lebih akurat mana dalam perhitungan prediksi
6	<i>Prediction Techniques of Heart Disease and Diabetes Disease using Machine Learning</i>	Dr.Geetha.Sa , Dr.Punitha Devi. Ca , Kalaivani. Va , Hariha.C Ja , and Preetha.Ga (Turkish Journal of Computer and Mathematics Education), 2021	Penelitian yang membahas tentang Teknik Prediksi Penyakit Jantung dan Penyakit Diabetes Menggunakan Machine Learning	Kasus uji yang masuk diproses oleh algoritme pembelajaran mesin dan menghasilkan hasil yang sesuai. Lakukan pemeriksaan pemegang pada ramalan penyakit jantung, ekspektasi kegunaan payudara dan perkiraan diabetes menggunakan komponen kunci seperti kadar Glukosa, Tekanan Darah, Ketebalan Kulit, BMI dll menyelidiki faktor-faktor ini. Anda mungkin harus mencari penyebaran faktor-faktor ini menggunakan histogram.	Baru ada 1 algoritma yang digunakan, dan tidak ada algoritma perbandingan pada penelitian ini	Penelitian yang sedang dikembangkan ada 2 algoritma dalam proses identifikasi resiko tinggi pada kehamilan, dimana algoritma tersebut nanti nya akan dibandingkan lebih akurat mana dalam perhitungan prediksi.

2.3. Landasan Teori

2.3.1. *Machine Learning*

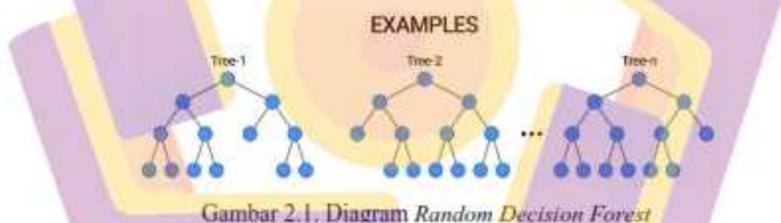
Machine Learning (ML) adalah sebuah aplikasi dari *Artificial Intelligence* (AI) atau kecerdasan buatan yang berfokus pada pengembangan sebuah system yang mampu belajar “sendiri” tanpa harus berulang kali di program oleh manusia [3]. Ada dua aplikasi utama dalam ML yaitu klasifikasi dan prediksi. Aplikasi ML membutuhkan data sebagai bahan belajar atau disebut juga dengan data latih (*train set*) untuk melatih model *Machine Learning*, kemudian untuk menguji suatu model *Machine Learning* yang telah dilatih, digunakan sebuah dataset baru yang disebut dengan data uji atau *test set*. Klasifikasi adalah metode dalam ML oleh mesin untuk memilah objek berdasarkan ciri tertentu atau membedakan benda satu dengan yang lain. Sedangkan prediksi digunakan oleh mesin untuk menentukan keluaran dari suatu data latih yang telah dimasukan sebelumnya.

2.3.2. *Random Forest*

Random Forest adalah pengembangan dari algoritma *Decision Tree*, Algoritma ini merupakan hasil dari pendekatan gabungan dari beberapa pembelajaran. algoritma *Random Forest* dikembangkan oleh Breiman yang digunakan untuk menyelesaikan masalah klasifikasi dan masalah regresi. Penggabungan dari beberapa model bertujuan untuk mendapatkan akurasi yang lebih tinggi dalam menyelesaikan masalah yang sama. Dalam beberapa masalah klasifikasi, algoritma *Random Forest* yang menggunakan hasil dari beberapa klasifikasi menunjukkan akurasi lebih tinggi daripada algoritma klasifikasi yang berjalan sendirian. *Random Forest* bekerja dengan cara mengumpulkan hasil dari

semua *tree* yang sudah dibuat, kemudian mengambil *voting* dari hasil tersebut, hasil yang paling banyak keluar itulah yang akan dipakai sebagai kesimpulan. (Sheikhmousa, et.al, 2020).

Teknik pengumpulan node pada *Random Forest* adalah *Bootstrap Aggregation*. Dengan teknik tersebut, *Bootstrap Aggregation* dapat memilih secara acak sampel dari sebuah data set. Kemudian setiap data yang dipilih secara acak akan dibuat model tersendiri (pohon), dan teknik tersebut disebut dengan *bootstrap*. Setiap model yang sudah dibuat akan dilatih secara mandiri dan akan menghasilkan kesimpulan di setiap modelnya. Hasil akhir yang diambil adalah mayoritas kesimpulan yang dibuat oleh setiap model. Pengambilan kesimpulan dari mayoritas model disebut dengan *aggregation*.

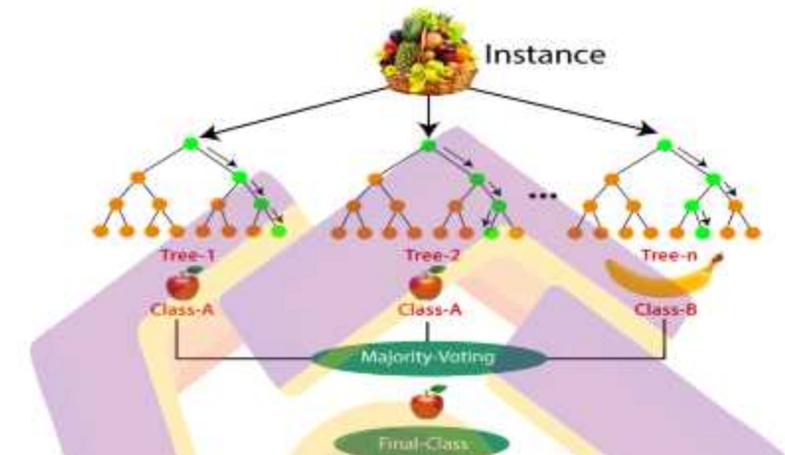


Gambar 2.1. Diagram *Random Decision Forest*

(Sumber : <https://towardsdatascience.com/how-do-random-forests- decision-trees- decide-simply-explained-with-an-example-in-python-6737eb183604>)

Sebagai contoh, terdapat sekumpulan data yang akan diibaratkan sebagai sekeranjang buah. Kemudian diambil n sampel buah yang akan dibuat model *Decision Tree* pada setiap sampel yang dibuat. Lalu setiap model sampel buah akan menghasilkan kesimpulan. Pada contoh gambar 2.2 terdapat 3 pohon. 2 pohon menghasilkan kesimpulan buah apel dan 1 pohon yang menghasilkan kesimpulan

buah pisang. Maka, dengan *Random Forest* yang akan diambil sebagai kesimpulan adalah suara terbanyak dari model pohon) yaitu buah apel.



Gambar 2.2. Contoh *Random Forest*

(Sumber : <https://www.analyticsvidhya.com/understanding-random-forest/>.)

Dalam menggunakan algoritma *Random Forest* terdapat beberapa parameter yang digunakan sebagai kerangka utama algoritma tersebut, parameter tersebut antara lain :

- n_estimators* : banyaknya pohon (*trees*) yang digunakan untuk membuat algoritma sebelum membuat prediksi dari rata-rata kesimpulan
- max_features* : angka maksimal dari *Random Forest* untuk membelah setiap node
- min_sample_leaf* : penentuan minimal *leaves* yang diperlukan untuk dibelah dalam *internal node*.

Terdapat beberapa kelebihan dari algoritma *random forest*, yaitu

- a. algoritma random forest memiliki akurasi yang tinggi untuk kumpulan *datasets* yang besar
- b. *random forest* akan bekerja sangat efisien dan akurat karena kesimpulan *random forest* merupakan kumpulan atau akumulasi dari algoritma *Decision Tree*.
- c. Dapat mengatasi data *training* dalam jumlah yang besar
- d. Memberikan hasil yang maksimal pada klasifikasi

Berikut ini adalah rumus yang digunakan random forest untuk melakukan klasifikasi

$$RFf_i = \frac{\sum_{j \in \text{all trees}} \text{norm}f_{ij}}{T}$$

Gambar 2.3. Rumus *Random Forest*

(Sumber : <https://towardsdatascience.com/the-mathematics-of-decision-trees-random-forest-and-feature-importance-in-scikit-learn-and-spark-f2861df67e3>.)

- RFfi = adalah kalkulasi dari seluruh pohon yang ada dimodel *Random Forest*
 normfi sub(ij) = normalisasi fitur penting untuk i pada
 pohon j T = total pohon yang ada di *Random Forest*

2.3.3. Algoritma *K-nearest Neighbors*

Menurut Kusriani dan Luthfi (2009) Algoritma K-NN yaitu suatu metode yang menggunakan algoritma *supervised*. Perbedaan antara *supervised learning* dengan *unsupervised learning* adalah pada *supervised learning* bertujuan untuk menemukan pola baru dalam data dengan menghubungkan pola data yang sudah ada dengan data yang baru. Sedangkan pada *unsupervised learning*, data belum memiliki pola apapun, dan tujuan *unsupervised learning* untuk menemukan pola dalam sebuah data. Tujuan 12 dari algoritma k-NN adalah untuk mengklasifikasi

objek baru berdasarkan atribut dan training samples. *K-Nearest Neighbors* (KNN) merupakan sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data latih yang jaraknya paling dekat dengan objek tersebut. Algoritma ini juga merupakan salah satu teknik *Lazy Learning* karena tidak mempelajari cara mengkategorikan data, melainkan hanya mengingat data yang sudah ada.

Pada KNN, parameter K merupakan jumlah dari tetangga terdekat. Algoritma ini menggunakan klasifikasi ketetanggaan sebagai nilai prediksi dari sampel uji yang baru. Pada saat diberikan sampel uji, maka algoritma akan menemukan sejumlah K objek (titik latih) yang paling dekat dengan titik uji. Pada penelitian ini, nilai K dan metode pengukuran jarak didapatkan dengan bantuan *GridSearchCV* yang berguna untuk menentukan nilai K dan metode pengukuran jarak terbaik yaitu $K = 17$ dan jarak Manhattan. Jarak Manhattan digunakan untuk menghitung perbedaan absolut antara koordinat sepasang objek. Penggunaan *Manhattan Distance* untuk pengukuran jarak memperoleh akurasi yang tinggi dibandingkan dengan *Euclidean Distance* sehingga dalam penelitian ini menggunakan klasifikasi metode *K-Nearest Neighbors* dengan menggunakan *Manhattan Distance* untuk menghitung jarak terdekat dalam metode klasifikasi. Perhitungan jarak pada penelitian ini dirumuskan oleh persamaan berikut:

$$d(x, y) = \sum_{i=1}^n |x_i - y_i|$$

Dimana :

$d(x, y)$: jarak

x_i : sampel data latih

y_i : sampel data uji

i : variable data

n : dimensi data

Berdasarkan fitur yang digunakan pada penelitian ini, maka perhitungan jarak antara sampel data latih dan data uji diperlihatkan dalam persamaan berikut:

$$d(x, y) = |A_x - A_y| + |B_x - B_y| + \dots + |Q_x - Q_y|$$

Dimana :

$d(x_i, y_i)$: jarak antara suatu data uji dan data latih

x_i : Data latih

y_i : Data uji

A : Home team

B : Away team

C : Half Time HomeGoals (HTHG)

E : Home Time Away Goals (HTAG)

F : Home Shot (HS)

G : Away Shot (AS)

H : Home Shot on Target (HST)

I : Away Shot on Target (AST)

J : Home Foul (HF)

K : Away Foul (AF)

L : Home Corner (HC)

M : Away Corner (AC)

N : Home Yellow (HY)

O : Away Yellow (AY)

P : Home Red (HR)

Q : Away Red (AR)

Perhitungan jarak pada penelitian ini, menggunakan bantuan *Scikit-Learn Library* yang mana di dalamnya memiliki algoritma komputasi yang dapat mempermudah perhitungan data dalam jumlah yang banyak.

2.3.4. Cross Validation

Metode *Cross Validation* merupakan salah satu metode untuk mengevaluasi algoritma learning dengan membagi data menjadi dua segmen yaitu. Segmen pertama digunakan untuk pembelajaran (*learning*) pelatihan (*training*) model, segmen kedua digunakan untuk validasi model. Ciri khas dari *Cross Validation* adalah set *training* dan validasi harus disilangkan (*cross-over*) dalam putaran

berturut-turut sehingga setiap titik data memiliki peluang untuk divalidasi. Dasar dari *Cross Validation* adalah *K-Fold Cross Validation* (Wong & Yang, 2017).

Dalam *Cross Validation*, langkah pertama kali adalah data akan dipartisi ke dalam segmen atau *fold* yang sama atau identik (nyaris sama). Berikutnya adalah iterasi ke k dari *training* dan validasi dilakukan sedemikian rupa sehingga dalam setiap iterasi *fold* data yang berbeda dimunculkan (*held-out*) untuk validasi, sementara sisa *fold* $k-1$ digunakan untuk *training*. Pada gambar 7.18 menggambarkan contoh dengan $k = 3$. Bagian yang gelap menggambarkan data untuk *training* sedangkan bagian terang menggambarkan data untuk validasi (*testing*). Dalam data mining maupun *machine learning* *10-fold cross-validation* ($k=10$) merupakan yang paling umum atau sering digunakan.



Gambar 2.4. Skema *Cross Validation*

2.3.5. Confusion Matrix

Confusion matrix adalah suatu metode yang digunakan untuk melakukan perhitungan dan menggambarkan kinerja model klasifikasi (*classifier*) pada satu set

data uji yang nilai sebenarnya diketahui. *Confusion matrix* dapat membantu dalam visualisasi kinerja suatu algoritma. Seperti yang pada tabel 2.1 yang menampilkan hasil prediksi pada masalah klasifikasi, jumlah prediksi benar dan salah dirangkum dengan nilai-nilai dan dipecah kepada masing-masing label. Hal ini dapat membantu untuk mengetahui kesalahan yang dibuat oleh *classifier*.

Tabel 2.2 *Confusion Matrix*

		<i>Predict Label</i>	
		<i>Positive (P)</i>	<i>Negative (N)</i>
<i>Actual Label</i>	P	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
	N	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Dimana :

True Positive : merupakan data positif dan diprediksi benar

True Negative : merupakan data negatif dan diprediksi benar

False Positive : merupakan data negatif tetapi diprediksi sebagai data positif (*error tipe 1*)

False Negative : merupakan data positif tetapi diprediksi sebagai data negatif (*error tipe 2*)

Pada penelitian ini menggunakan 3 label klasifikasi, sehingga Tabel 2.1 berubah sesuai label klasifikasi yang ditunjukkan pada Tabel 2.2.

Tabel 2.3 *Confusion Matrix 3x3*

		<i>Predict Label</i>		
		Label D	Label H	Label A
<i>Actual Label</i>	Label D	DD	DH	DA
	Label H	HD	HH	HA
	Label A	AD	AH	AA

2.3.6. Index Pengukuran

Berdasarkan *Confusion Matrix*, maka dapat diperoleh beberapa variabel pengukuran yang dapat digunakan untuk mengukur dan mengevaluasi kinerja klasifikasi:

- a. Akurasi, menunjukkan sejauh mana hasil skrining sesuai dengan kenyataannya, atau, proporsi subjek yang diidentifikasi dengan benar sesuai dengan standar yang terbaik yang telah disepakati bersama (*gold standard*), Akurasi menjawab pertanyaan “Berapa persen ibu hamil yang mengalami keguguran”.

$$\text{Accuracy} = \frac{TP+TN}{TP+FP+TN+FN} \times$$

- b. *Precision* Merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif. *Precision* menjawab pertanyaan “Berapa persen ibu hamil yang diprediksi aman dari segi usia ibu yang sedang mengandung. Untuk menghitung nilai *Precision*, dapat digunakan persamaan berikut:

$$\text{Precision} = \frac{TP}{TP+FP} \times 100\%$$

- c. *Sensitivity*, atau *True Positive Rate* (TPR) adalah proporsi subjek positif dan teridentifikasi oleh sistem dimana semua subjek merupakan subjek positif. *Sensitivity* menjawab pertanyaan “Berapa persen ibu hamil yang diprediksi sehat dan masuk dalam resiko tinggi dalam kehamilan”. Untuk menghitung nilai *Sensitivity*, dapat digunakan persamaan berikut:

$$\text{Sensitivity} = \frac{TP}{TP+FN} \times 100\%$$

BAB III

METODE PENELITIAN

3.1. Jenis, Sifat, dan Pendekatan Penelitian

3.1.1. Jenis Penelitian

Penelitian ini menggunakan desain penelitian eksperimen (*Experimental Research*). Karakteristik penelitian eksperimen menurut Solso, Johnson dan Beal (1998) diantaranya adalah : (1) adanya manipulasi variabel atau *factor* tertentu (variabel bebas); (2) adanya kontrol terhadap variabel lain yang diduga mempengaruhi proses penelitian; dan (3) mengukur efek variabel yang dimanipulasi terhadap variabel lain yang diukur dengan variabel terikat

3.1.2. Sifat Penelitian

Sifat penelitian ini adalah penelitian kausal yang mencari hubungan sebab dan akibat dari suatu variabel (variabel bebas) terhadap variabel tetap yang kemudian dimanipulasi secara terkontrol.

3.1.3. Pendekatan Penelitian

Pendekatan penelitian ini menggunakan pendekatan kuantitatif karena data penelitian yang digunakan merupakan angka-angka dan analisis menggunakan perhitungan matematis untuk mendapatkan hasil yang diinginkan serta penelitian akan dilaksanakan sesuai dengan langkah-langkah atau alur yang sudah ditentukan oleh peneliti.

3.2. Metode Pengumpulan Data

Pengumpulan data adalah kegiatan yang dilakukan pertama kali sebelum melakukan kegiatan analisis data. Data yang diperlukan adalah data masyarakat yang ada di Desa Bolo Pleret Klaten pada Aplikasi Deteksi Kehamilan. Data tersebut di normalisasikan dengan melalui proses ETL (*Extract, Transform, Load*) untuk menghimpun data transaksi masyarakat melalui DBMS.

3.2. Metode Analisis Data

3.2.1. Alat

a. Perangkat keras :

Perangkat keras yang digunakan dalam penelitian ini adalah laptop dengan spesifikasi prosesor Intel® Core™ i5 3317U, CPU @ 1.70Ghz, RAM 8GB, dan SSD 256GB laptop digunakan untuk pemrosesan seluruh tahapan pengujian.

b. Perangkat lunak :

Perangkat lunak yang digunakan untuk penelitian ini yaitu:

1. Bahasa pemrograman *Python*.
2. *Scikit-learn*, berfungsi sebagai *framework* yang digunakan pada penelitian ini untuk mengolah dataset, tahapan yang akan dilakukan adalah pembersihan data, implementasi algoritme, dan perhitungan validasi model.
3. *Google Colab*, yaitu aplikasi yang digunakan untuk menjalankan dokumen berformat *ipynb* dengan bahasa pemrograman *Python* melalui *web browser*.

3.2.2. Bahan

Bahan yang digunakan pada penelitian ini adalah dataset berupa data kemahamilan masyarakat Desa Bolo Pleret Klaten, data diambil berdasarkan laporan yang sudah di rekap di Desa Bolo Pleret Klaten.

- a. Penyedia dataset adalah Kantor Desa di Bolo Pleret Klaten dan Puskesmas di Desa Bolo Pleret Klaten yang beroperasi untuk melayani masyarakat melalui pencatatan online maupun offline
- b. Dataset merupakan data masyarakat bolo pleset yang sudah berkeluarga dan mempunyai Kartu Keluarga
- c. Melakukan analisis statistika deskriptif terhadap perubahan respons.
- d. Melakukan simulasi random forest.
 1. Sebanyak 1000 random forest dibentuk dimana p adalah banyaknya peubah penjelas dalam data, yaitu $p = 20$.
 2. Menganalisis tingkat misklasifikasi *random forest* yang dihasilkan dari langkah 2a. Analisis dilakukan secara eksploratif.
 3. Melakukan *driver analysis* dengan metode *random forest*, yaitu mengamati pada setiap masing-masing *random forest* dan *mean decrease gini* (MDG) setiap peubah penjelas. Nilai m dan k yang disarankan Breiman (2001) dicobakan dalam simulasi ini. Nilai k yang disarankan untuk digunakan pada metode bagging juga dicobakan, yaitu $k = 50$. Umumnya $k = 50$ sudah memberikan hasil yang memuaskan m dan k yang dicobakan, kemudian dicatat tingkat misklasifikasi urutan MDG peubah penjelas. MDG setiap peubah penjelas dihasilkan pada langkah 2a.

4. Melakukan analisis korelasi Spearman terhadap data.
5. Melakukan intepretasi hasil *driver analysis*.
6. Metode *random forest* dihasilkan menggunakan *software* R ver 2.12.0 dengan paket *randomForest* ver 3.6-2. Kriteria berhenti memilah yang digunakan adalah terdapatnya satu amatan pada simpul terminal.

Algoritma KNN merupakan algoritma klasifikasi yang bekerja dengan mengambil sejumlah K data terdekat (*tetangganya*) sebagai acuan untuk menentukan kelas dari data baru. Algoritma ini mengklasifikasikan data berdasarkan *similarity* atau kemiripan atau kedekatannya terhadap data lainnya.

7. Secara umum, formula *Euclidean distance* pada *1-dimensional space* adalah sebagai berikut.

$$dis(x_1, x_2) = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2}$$

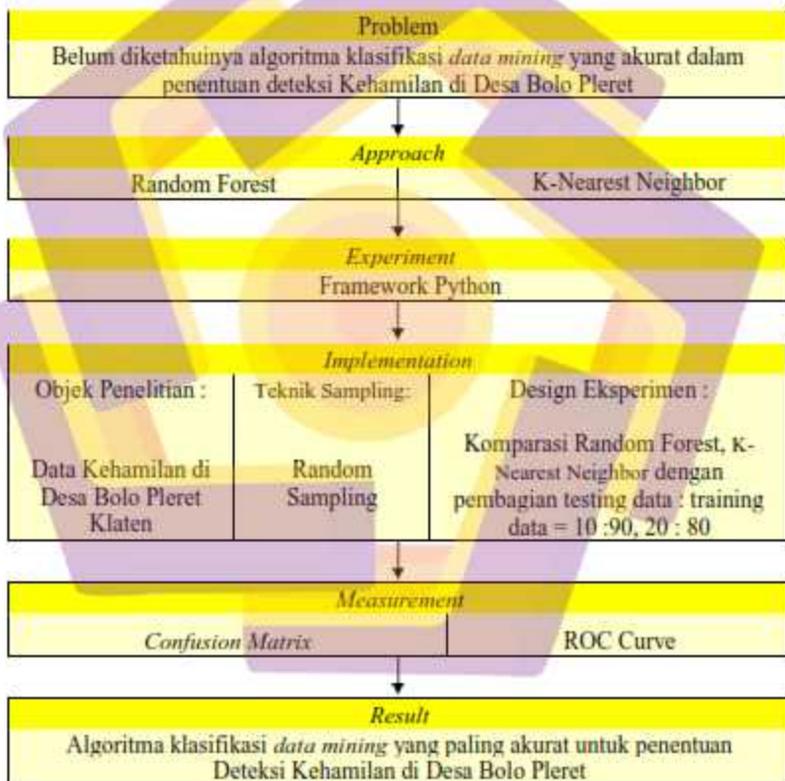
8. Formula di atas dapat digunakan jika jumlah *independent variable* hanya ada satu variabel. Lalu, bagaimana jika ada banyak variabel yang digunakan? Jika ada lebih dari satu, kita dapat menjumlahkannya seperti di bawah ini.

$$dis = \sqrt{\sum_{i=0}^n (x_{1i} - x_{2i})^2 + (y_{1i} - y_{2i})^2 + \dots}$$

3.3. Alur Penelitian

Dalam menyelesaikan penelitian, penulis membuat sebuah kerangka pemikiran yang berguna sebagai pedoman penelitian ini sehingga penelitian dapat

dilakukan secara konsisten. Permasalahan pada penelitian ini adalah belum diketahuinya metode yang paling akurat dalam klasifikasi resiko tinggi, rendah dan sedang pada ibu hamil. Atas dasar permasalahan tersebut penulis melakukan perbandingan terhadap metode klasifikasi algoritma Random Forest dan K-Nearest Neighbor untuk memecahkan masalah penelitian ini. Berikut kerangka pemikiran yang penulis buat untuk penelitian ini:



Gambar 3.1. Kerangka pemikiran

Penjelasan dari tahapan penelitian yang ditampilkan di atas sebagai berikut:

3.4. Data Pengumpulan Data

Tahap ini mencakup proses pengumpulan data dan analisis deskriptif terhadap data penelitian. *Dataset* penelitian yang digunakan merupakan data data masyarakat yang hamil di Desa Bolo Pleret Klaten yang diambil secara acak. Dataset tersebut didapat dari data transaksi yang telah di rekap di kelurahan dan dimasukkan kedalam aplikasi Deteksi Kehamilan secara *online*.

Dalam penelitian ini, dataset menggunakan data kehamilan, beberapa data yang akan menjadi fokus fitur dalam penelitian ini seperti data kependudukan di Desa Bolo Pleret kemudian akan di kelompokkan dalam satu *Dataset*. Setiap pencatatan diambil berdasarkan waktu pelaporan, sehingga ukuran data bervariasi. Oleh karena itu perlu dilakukan analisis data menggunakan ilmu statistik. Analisis deskriptif terhadap dataset penelitian menggunakan ilmu statistik, bertujuan untuk mengetahui karakteristik data yang akan diolah, sehingga nantinya ditemukan sensor apa saja yang paling berpengaruh untuk menilai cara berkendara.

3.4.1. Pengumpulan Data

Data yang telah dikumpulkan pada langkah sebelumnya, diupload ke dalam *google drive* untuk mempermudah pemanggilan atau *import* di *tools google colaps*. *Dataset* Mentah terdiri dari 2 *file* data terpisah yaitu *file* data fitur dan *file* data label. Data fitur berisi data kehamilan, data kependudukan, dan data kartu keluarga, sedangkan data label berisi Kehamilan yang telah diberi label pencatatan dan label data kependudukan yang sudah menikah, namun tidak semua data fitur yang disediakan memiliki label. Secara umum, data mentah yang diolah dalam penelitian ini memiliki jumlah data sekitar 200 data, jumlah data penduduk yang berada di

Bolo Pleret Klaten. Setiap melakukan tahapan preprocessing data, dataset baru disimpan berdasarkan proses yang telah dilakukan. Selanjutnya dibersihkan terlebih dahulu agar lebih optimal, seperti melakukan penghapusan data yang tidak memiliki label.

3.4.2. Pengujian Data

Data yang telah selesai di *pre-processing* akan diambil sebesar 80% untuk dilakukan *training* dengan menggunakan tiga model klasifikasi yang berbeda diantaranya *Random forest*, *K-Nearest Neighbor*. Selanjutnya, model yang telah dilatih akan diuji menggunakan data dengan Komparasi *Random Forest*, *K-Nearest Neighbor* dengan pembagian *testing data* : *training data* = 10 : 90, 20 : 80. Dari hasil uji tersebut akan didapatkan masing-masing tingkat akurasi, presisi, *recall* dan *F1 score* dari tiap jenis data latih. Setelah semua skenario selesai dilaksanakan, akan dilakukan analisis dan penarikan kesimpulan penelitian.

BAB IV

HASIL PENELITIAN DAN PEMBAHASAN

4.1. Objek Penelitian

Secara Administratif Desa Bolo Pleret termasuk dalam Wilayah Kabupaten Klaten Provinsi Jawa Tengah dan terletak dibagian Utara merupakan salah satu desa yang memiliki potensi Ekonomi berkembang didukung oleh infrastruktur Desa Bolo Pleret dilihat secara umum keadaannya merupakan daerah dataran rendah dengan persawahan yang luas dialiri oleh sungai yaitu sungai Indragiri dengan beberapa anak Sungai (Monografi Desa Bolo Pleret Klaten (2017).

Pada tahun 2000 Desa Bolo Pleret Masih merupakan bagian dari wilayah Desa Juwiring dan Pada tahun 2002 terjadi pemekaran yaitu Desa Sungai Piring yang sekarang menjadi Kelurahan Juwiring dan Bolo Pleret Menjadi Desa sehingga di sebut Desa Bolo Pleret (Junadhi et al., 2022).

4.1.1. Deskripsi Stngkat Instansi



Gambar 4.1 Peta Desa Bolo Pleret Klaten Jawa Tengah

Sumber :

<https://www.google.com/maps/place/Bolopleret,+Juwiring,+Klaten+Regency,+Central+Java>

Secara demografis atau kependudukan berdasarkan laporan dalam Monografi Desa Bolo Pleret tahun 2021, penduduk Desa Bolo Pleret berjumlah 1352 orang, dengan jumlah penduduk berjenis kelamin perempuan sebanyak 654 orang dan laki laki sebanyak 698 orang. Jumlah Kepala Keluarga (KK) di desa Bolo Pleret adalah sebanyak 375 KK. (Hanadwiputra et al., 2022) Jika membandingkan antara jumlah penduduk dan jumlah KK, maka dapat dikatakan bahwa rata-rata jumlah anggota keluarga dalam satu rumah tangga adalah 3,6 orang.

Berdasarkan jumlah penduduk tersebut, dalam catatan Monografi, jumlah penduduk dengan rentang usia 0-15 tahun adalah sebanyak 409 orang, jumlah penduduk rentang usia 16-65 sebanyak 750 orang, dan jumlah penduduk usia di atas 66 tahun adalah sebanyak 94 orang. Data ini menunjukkan bahwa usia produktif penduduk desa Bolo Pleret sangat baik. (Evan et al., 2017)

Jumlah penduduk berdasarkan tingkat pendidikan dalam Monografi desa memang masih perlu diperbaharui atau diperbaiki. (I. Budiman & Ramadina, 2015) Hal ini dikarenakan yang tercatat dalam buku Monografi Desa Tahun 2021 hanyalah jumlah penduduk yang telah menyelesaikan pendidikannya, yakni sebanyak 299 orang, sedangkan yang belum sekolah, masih sekolah dan tidak sekolah atau sekolah namun tidak lulus belum tercatat. Rinciannya dapat dilihat pada table berikut :

Tabel 4.1. Jumlah Penduduk Desa Bolo Pleret Yang Telah Menyelesaikan Pendidikan

No	Lulus pada Tingkat Pendidikan	Jumlah
1	Sekolah Menengah Pertama (SMP)	88
2	Sekolah Menengah Umum (SMU)	86
3	Akademi (D1-D3)	55
4	Sarjana (S1)	50
5	Sarjana Strata 2 maupun 3	20

Sumber: Monografi Desa Bolo Pleret 2021

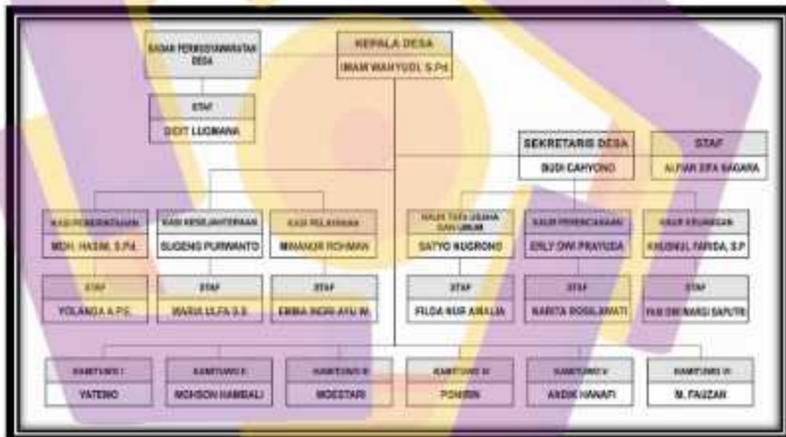
Sarana dan prasarana pendidikan di desa Bolo Pleret belum maksimal. Berdasarkan hasil observasi dan wawancara, hanya terdapat satu gedung PAUD, satu gedung Taman Kanak-kanak (TK) dan satu gedung Sekolah Dasar (SD). Untuk melanjutkan pendidikan ke tingkat SMP dan SMU peserta didik (anak-anak) harus menempuh jarak sekitar 2 kilometer ke ibu kota kecamatan Klaten untuk melanjutkan pendidikan pada tingkat SMP, SMU dan Perguruan Tinggi. Sekalipun demikian dapat dikatakan bahwa tingkat pendidikan masyarakat desa Bolo Pleret cukup baik, yang tampak pada tabel 1 di atas.

4.1.2. Struktur Pemerintah Desa Bolo Pleret

Desa Bolo Pleret saat ini dipimpin oleh seorang Kepala Desa, yakni bapak Imam Wahyudi, yang membawahi 2 buah dusun. Struktur pemerintah desanya terdiri dari Pemerintah Desa, Badan Permusyawaratan Desa (BPD) dan Lembaga Kemasyarakatan yang berposisi sejajar dengan Kepala Desa. (Mardi, 2017) Hubungan kepala desa dengan BPD adalah bersifat koordinasi sedangkan hubungan kepala desa dengan lembaga kemasyarakatan berada dalam hubungan garis komando.

Berdasarkan hasil observasi dan wawancara dengan Sekretaris Desa ditemukan bahwa unit kerja pemerintah desa Bolo Pleret terdiri dari: Kepala Desa, Sekretaris Desa, Kasie Pemerintahan, Kasie Kesejahteraan, Kaur Keuangan, Kaur Administrasi, yang ditunjang, unit kerja tersebut dengan keberadaan 2 orang Kepala Dusun. (Dewi & Dwidasmara, 2020) Maka dalam struktur pemerintah desa Bolo Pleret terdapat 8 orang sebagai pemimpin masyarakat desa Bolo Pleret Klaten. Selain itu, terdapat pula 8 orang yang merupakan representasi masyarakat desa Bolo

Pleret yang berkedudukan sebagai Badan Permusyawaratan Desa (BPD), desa Bolo Pleret Klaten. Fungsi BPD berdasarkan UU No. 32 tahun 2004 adalah sebagai Badan Permusyawaratan Desa yang akan memberikan masukan-masukan kepada Kepala Desa berdasarkan aspirasi masyarakat yang mereka terima. Sebagai badan permusyawaratan, BPD memang tidak berfungsi dengan baik sebab terkadang tidak mampu memainkan fungsi representatif warga masyarakat, kedudukannya yang sejajar dengan kepala desa dan hanya berfungsi dalam garis koordinasi bisa dibaca sebagai bentuk kelemahannya. (Firdaningrum et al., 2021). Lebih jelasnya tentang struktur pemerintahan desa Bolo Pleret dapat dilihat pada gambar di bawah ini:



Gambar 4.2: Struktur Organisasi Desa Bolo Pleret Klaten
Sumber : Kantor Desa Bolo Pleret Klaten Jawa Tengah

4.1.3. Demografi Desa Bolo Pleret Klaten

a. Pendudukan

Jumlah penduduk yang besar biasa menjadi modal dasar pembangunan sekaligus bisa menjadi beban pembangunan, jumlah penduduk desa Bolo Pleret adalah

1543 Jiwa dengan jumlah Kepala Keluarga 700 Kepala keluarga. Agar dapat menjadi dasar pembangunan maka jumlah penduduk yang besar harus disertai kualitas SDM yang tinggi. Penanganan kependudukan sangat penting sehingga potensi yang dimiliki mampu menjadi pendorong dalam pembangunan, khususnya pembanguna Desa Bolo Pleret. Berkaitan dengan kependudukan, aspek yang penting antara lain perkembangan jumlah penduduk, kepadatan dan persebaran serta strukturnya.

Tabel 4.2. Jumlah penduduk Desa Bolo Pleret Klaten

Laki-Laki	Perempuan	Jumlah Total
786	757 Jiwa	1543

b. Pertumbuhan Jumlah Penduduk

Jumlah penduduk Desa Bolo Pleret Klaten cenderung meningkat karena tingkat kelahiran lebih besar dari pada kematian serta penduduk yang masuk lebih besar dari penduduk yang keluar.

Tabel 4.3. Jumlah dan Laju Pertumbuhan Penduduk Desa Bolo Pleret Klaten Tahun 2021-2022

No	Rukun Tetangga	Jumlah Penduduk (Jiwa)	
		2013	2014
1	Dusun 01	600	757
2	Dusun 02	350	386
3	Dusun 03	352	400
Jumlah 03 dusun		1302	1543

Sumber : Data Dari Ketua RT desa .

c. Kepadatan dan Persebaran Penduduk

Persebaran penduduk di Desa Bolo Pleret relatif merata, secara absolut jumlah penduduk pada tiap-tiap Rukun Tetangga (RT) terlihat relatif berimbang, namun karena luas wilayah masing-masing RT berbeda maka tingkat kepadatan

penduduknya terlihat beda pada tahun 2011. RT 01 dan 02, merupakan wilayah dengan tingkat kepadatan penduduk yang tertinggi di wilayah Desa Bolo Pleret, Sementara itu RT.03 dan 04 merupakan tingkat kepadatan terendah.

Tabel 4.4. Jumlah Kepadatan dan Persebaran Penduduk Desa Bolo Pleret Klaten Tahun 2021

No	Rukun Tetangga	jumlah h[Ha]	Jumlah Penduduk (Orang)	Kepadatan (Orang/ m ²)	Persebaran %
1	Dusun I	505	757	417	43
2	Dusun II	771	386	305	32
3	Dusun III	1249	400	275	25
	Jumlah	2.525	1543	1097	100

Rumus jumlah penduduk di RT. dibagi jumlah total penduduk desa kali 100% = hasil.

4.2. Penelitian Eksperimental

Tahap penelitian eksperimen meliputi proses pengamatan lalu lintas data saat transaksi pembayaran berjalan dari awal saat inquiry hingga proses akhir transaksi baik *flagging* maupun *reversal*. (Rane et al., 2014) Setelah didapatkan data tersebut akan dijadikan sebagai dataset pengujian konsistensi data pembayaran yang akan diolah menggunakan SPSS. (J. Budiman et al., 2018)

4.3. Pengumpulan Data Uji

Pertanyaan pertama dalam rumusan masalah adalah tentang faktor risiko tinggi yang perlu diidentifikasi pada ibu hamil di Desa Bolo Pleret Klaten. Berdasarkan hasil analisis data, faktor risiko tinggi yang perlu diidentifikasi meliputi:

- a. Tekanan Darah (Sistolik dan Diastolik): Rata-rata tekanan darah sistolik adalah 128.15 mmHg, sedangkan diastolik adalah 83.34 mmHg.
- b. Usia Kehamilan (A2): Rata-rata usia kehamilan adalah 12.67 minggu.
- c. Hasil Protein Urine (A3): Semua data menunjukkan hasil protein urine negatif.
- d. Pemeriksaan Kondisi Fisik (A4): Rata-rata pemeriksaan kondisi fisik adalah 39.75, yang mengindikasikan kondisi fisik yang baik.
- e. Jenis Pendarahan (A5): Mayoritas data menunjukkan tidak adanya jenis pendarahan.
- f. Kontraksi Uterus (A6): Rata-rata kontraksi uterus adalah 2.64 kali.
- g. Umur Ibu Hamil (A7) : Rata-rata umur ibu hamil atas 17 tahun.
- h. Hiper Preeklamsia (A7): Mayoritas data menunjukkan tidak adanya hiper preeklamsia.

Data yang dihimpun berupa data yang di ekspor kedalam format Microsoft excel (.Xls) dengan jumlah data sekitar 300 dataset.

id	usia kehamilan	hasil protein urine	pemeriksaan kondisi fisik	jenis pendarahan	kontraksi uterus	umur ibu hamil
02001	12 minggu	negatif	39	tidak ada	2	17
02002	12 minggu	negatif	40	tidak ada	2	17
02003	12 minggu	negatif	39	tidak ada	2	17
02004	12 minggu	negatif	39	tidak ada	2	17
02005	12 minggu	negatif	39	tidak ada	2	17
02006	12 minggu	negatif	39	tidak ada	2	17
02007	12 minggu	negatif	39	tidak ada	2	17
02008	12 minggu	negatif	39	tidak ada	2	17
02009	12 minggu	negatif	39	tidak ada	2	17
02010	12 minggu	negatif	39	tidak ada	2	17
02011	12 minggu	negatif	39	tidak ada	2	17
02012	12 minggu	negatif	39	tidak ada	2	17
02013	12 minggu	negatif	39	tidak ada	2	17
02014	12 minggu	negatif	39	tidak ada	2	17
02015	12 minggu	negatif	39	tidak ada	2	17
02016	12 minggu	negatif	39	tidak ada	2	17
02017	12 minggu	negatif	39	tidak ada	2	17
02018	12 minggu	negatif	39	tidak ada	2	17
02019	12 minggu	negatif	39	tidak ada	2	17
02020	12 minggu	negatif	39	tidak ada	2	17
02021	12 minggu	negatif	39	tidak ada	2	17
02022	12 minggu	negatif	39	tidak ada	2	17
02023	12 minggu	negatif	39	tidak ada	2	17
02024	12 minggu	negatif	39	tidak ada	2	17
02025	12 minggu	negatif	39	tidak ada	2	17
02026	12 minggu	negatif	39	tidak ada	2	17
02027	12 minggu	negatif	39	tidak ada	2	17

Gambar 4.3. Dataset SQL Kehamilan Desa Bolo Pleret Klaten

4.4. Proses *preprocessing* data

Hasil pengumpulan data uji kemudian dilakukan *preprocessing* data yang meliputi *cleaning* data dan *integration* data. *Preprocessing* dilakukan untuk mengkonversi data yang dihimpun kedalam format Microsoft excel (.xls) agar dapat dikenali dan diolah menggunakan software analisis statistika sesuai dengan skenario yang telah dirancang. (Rismawan & Kusumadewi, 2008)

Tahap pertama yang dilakukan saat *preprocessing* adalah dengan melakukan *cleaning data*, yaitu dengan membersihkan data dari *noise*. *Cleaning data* dilakukan pada data kehamilan.

	A2	A3	AA	A5	A6	A7	A8	Sistolik	Diastolik
count	298.000000	298.0	298.000000	298.000000	298.000000	298.000000	298.000000	298.000000	298.000000
mean	12.667705	0.0	10.751679	0.253678	0.076470	2.677584	27.966830	128.154362	81.340282
std	1.811021	0.0	1.174996	0.434707	0.256368	0.999763	6.188527	6.515172	4.046484
min	10.000000	0.0	38.000000	0.000000	0.000000	0.000000	18.000000	110.000000	70.000000
25%	11.000000	0.0	39.000000	0.000000	0.000000	3.000000	22.000000	124.000000	80.000000
50%	13.000000	0.0	40.000000	0.000000	0.000000	1.000000	26.000000	128.000000	82.000000
75%	14.000000	0.0	41.000000	0.750000	0.800000	2.000000	34.000000	132.000000	85.000000
max	16.000000	0.0	41.000000	1.000000	1.000000	8.000000	40.000000	140.000000	90.000000

Gambar 4.4. Analisis Deskriptif pada Data

Keterangan : Dari gambar diatas untuk nilai rata-rata (Mean) sebesar 12,66 yang memberikan nilai Tengah dari Kumpulan data yang diambil dari Dataset Kehamilan, Nilai Deviasi Standar (Standard Deviation) sebesar 1,82 menunjukkan sebaran data dari rata. (Kristiana et al., 2020) Dalam konteks ini, nilai deviasi standar yang relatif kecil (1,82) menunjukkan bahwa sebagian besar nilai berada cukup dekat dengan rata-rata. Nilai Minimum dan Maksimum: Nilai minimum didapat 10,00, dan nilai maksimum adalah 16,00. Ini memberikan batas bawah dan

batas atas dari distribusi data. Kuartil (25%, 50%, 75%): Kuartil pertama (25%) sebesar 11,00, menunjukkan nilai di mana 25% dari data berada di bawahnya. Kuartil kedua (50%) sebesar 13,00, merupakan median atau nilai tengah dari distribusi. Kuartil ketiga (75%) sebesar 14,00, menunjukkan nilai di mana 75% dari data berada di bawahnya.

Hasil *cleaning data* kemudian diintegrasikan dalam satu file yang dapat dibaca dan diolah oleh software statistik, yaitu dalam bentuk file berekstensi .xls atau .xlsx untuk kemudian diolah agar mendapatkan perhitungan pengujian konsistensi data.

4.5. Pengujian Normalitas Data

Uji Normalitas adalah uji yang digunakan untuk menunjukkan persebaran data yang digunakan dalam penelitian terdistribusi secara normal atau tidak. Ketentuan yang harus dipenuhi jika melakukan uji *Kolmogorov-Smirnov* yaitu, jika nilai signifikansi $> 0,05$ maka data yang digunakan dalam penelitian memiliki distribusi yang normal. Namun, sebaliknya, jika nilai signifikansi $< 0,05$ maka data yang digunakan tidak memiliki distribusi yang normal. Jika nilai di atas 0,05 maka distribusi data dinyatakan memenuhi asumsi normalitas, dan jika nilai di bawah 0,05 maka diinterpretasikan sebagai tidak normal. Hasil pengujian normalitas dapat dilihat sebagai berikut :

Tabel 4.5. Pengujian normalitas Data kehamilan Desa Bolo Pleret Klaten

One-Sample Kolmogorov-Smirnov Test			
		Berat Badan	Usia
N		198	198
Normal Parameters**	Mean	11.9949	40.0303
	Std. Deviation	1.60424	1.07538
Most Extreme Differences	Absolute	.224	.261
	Positive	.197	.184
	Negative	-.224	-.261
Test Statistic		.224	.261
Asymp. Sig. (2-tailed)		.890	.802

a. Test distribution is Normal.

b. Calculated from data.

Hasil pengujian normalitas didapatkan nilai signifikansi masing-masing sebesar 0,001 untuk variable berat badan ibu hamil dan usia ibu hamil sebesar 0,05 karena nilai Sig lebih besar dari 0,05 yang berarti bahwa data berat badan ibu hamil dan usia ibu hamil pada dataset ibu hamil di Desa Bolo Pleret Klaten terdistribusi secara normal.

4.6. Implementasi Algoritma

Dalam penelitian ini, Algoritma *Random Forest* dan *K-Nearest Neighbor* diimplementasikan untuk mengklasifikasikan risiko tinggi dan rendah pada ibu hamil. Pertanyaan kedua dalam rumusan masalah adalah mengenai tingkat akurasi hasil implementasi Algoritma *Random Forest* dan *K-Nearest Neighbor*.

a. Akurasi *Random Forest*

Hasil pengujian menunjukkan bahwa model *Random Forest* memiliki tingkat akurasi sebesar 98,33%. Artinya, model ini mampu memprediksi risiko tinggi

atau rendah pada ibu hamil dengan sangat baik. Dengan tingkat akurasi yang tinggi, *Random Forest* menjadi pilihan yang kuat untuk aplikasi deteksi risiko tinggi pada kehamilan.

Selain akurasi, parameter lain yang perlu diperhatikan adalah *precision*, *recall*, dan *F1-score*.

1. *Precision* (presisi): Untuk kategori "ada" (risiko tinggi), *precision* sebesar 100%, yang berarti semua prediksi risiko tinggi adalah benar.
2. *Recall* (sensitivitas): *Recall* untuk kategori "ada" sebesar 96%, yang mengindikasikan bahwa model dapat mendeteksi sebagian besar risiko tinggi.
3. *F1-Score*: *F1-Score* sebesar 0.98, mencerminkan keseimbangan antara *precision* dan *recall*.

Akurasi Random Forest: 0.9833333333333333				
	precision	recall	f1-score	support
ada	1.00	0.96	0.98	25
tidak	0.97	1.00	0.99	35
accuracy			0.98	60
macro avg	0.99	0.98	0.98	60
weighted avg	0.98	0.98	0.98	60

Gambar 4.5. *Classification Report Random Forest*

b. Akurasi *K-Nearest Neighbor*

Model *K-Nearest Neighbor* memiliki tingkat akurasi sebesar 95%. Meskipun sedikit lebih rendah dibandingkan dengan *Random Forest*, akurasi ini masih dianggap sangat baik. *K-Nearest Neighbor* tetap menjadi alternatif yang kuat dalam deteksi risiko tinggi pada kehamilan. (Diani, 2018)

Selain akurasi, parameter lain yang perlu diperhatikan adalah *precision*, *recall*, dan *F1-score*.

1. *Precision* (presisi): Untuk kategori "ada" (risiko tinggi), *precision* sebesar 92%, yang berarti sebagian prediksi risiko tinggi adalah benar.
2. *Recall* (sensitivitas): *Recall* untuk kategori "ada" sebesar 96%, yang mengindikasikan bahwa model dapat mendeteksi sebagian besar risiko tinggi.
3. *F1-Score*: *F1-Score* sebesar 0,94, mencerminkan keseimbangan antara *precision* dan *recall*.

	precision	recall	f1-score	support
ada	0.89	0.96	0.92	25
tidak	0.97	0.91	0.94	35
accuracy			0.93	60
macro avg	0.93	0.94	0.93	60
weighted avg	0.94	0.93	0.93	60

Gambar 4.6. *Classification Report* K-NN

4.7. Jumlah Prediksi Risiko Tinggi dan Rendah

Pertanyaan ketiga dalam rumusan masalah adalah mengenai jumlah prediksi risiko tinggi dan rendah pada ibu hamil di Desa Bolo Pleret, Kabupaten Klaten, yang dihasilkan dengan pengujian Algoritma *Random Forest* dan *K-Nearest Neighbor* (Masdar et al., 2016).

a. Distribusi Hasil Prediksi *Random Forest*

Confusion Matrix - Random Forest

Actual \ Prediksi	0	1	2
0	0	5	1
1	3	45	0
2	1	2	2

Gambar 4.7. Hasil Prediksi Untuk Algoritma *Random Forest*

Dari hasil pengujian dengan *Random Forest*, ditemukan bahwa dari total 60 data ibu hamil, sebanyak 25 diantaranya diprediksi sebagai risiko tinggi (label "ada") dan 35 sebagai risiko rendah (label "tidak"). Hal ini menunjukkan bahwa *Random Forest* lebih cenderung memprediksi risiko rendah dibandingkan dengan risiko tinggi. (Masdar et al., 2016)

b. Distribusi Hasil Prediksi *K-Nearest Neighbor*

Confusion Matrix - K-NN

Actual \ Prediksi	0	1	2
0	1	3	0
1	7	41	0
2	1	5	0

Gambar 4.8. Hasil Prediksi Untuk Algoritma *KNN*

Sementara itu, dari hasil pengujian dengan *K-Nearest Neighbor*, terdapat 25 data ibu hamil yang diprediksi sebagai risiko tinggi dan 35 sebagai risiko rendah. Distribusi hasil prediksi *K-Nearest Neighbor* cenderung seimbang antara risiko tinggi dan rendah.

4.8. Evaluasi

Evaluasi pada klasifikasi *machine learning* merupakan suatu proses pengukuran akurasi dari model yang sebelumnya telah diproses. Hasil dari evaluasi adalah berupa ukuran akurasi, *precision*, *recall*, dan *F1-Score*. Beberapa poin dari evaluasi akan dijabarkan sebagai berikut. (Praditasari & Sumarmik, 2018)

4.8.1. Akurasi

Akurasi adalah parameter yang mengukur sejauh mana model dapat memprediksi dengan benar. Perhitungan manual akurasi dilakukan dengan rumus berikut:

$$\text{Akurasi} = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Data Uji}}$$

Dalam hal ini, perhitungan manual akurasi menghasilkan nilai yang sama dengan hasil dari implementasi Algoritma, yaitu 98.33% untuk *Random Forest* dan 95% untuk *K-Nearest Neighbor*. (Yuli Mardi, 2019)

4.8.2. Precision

Precision (presisi) adalah parameter yang mengukur sejauh mana prediksi positif model benar. Perhitungan manual *precision* dilakukan dengan rumus berikut:

$$\text{Precision} = \frac{TP}{(TP + FP)}$$

Di mana TP adalah *True Positive* (prediksi risiko tinggi yang benar), dan FP adalah *False Positive* (prediksi risiko tinggi yang salah).

a. Perhitungan *Precision Random Forest*

$$TP (\text{True Positive}) = 24$$

$$FP (\text{False Positive}) = 1$$

$$\text{Precision} = \frac{24}{(24 + 1)}$$

$$\text{Precision} = \frac{24}{25}$$

$$\text{Precision} = 0.96$$

b. Perhitungan *Precision K-Nearest Neighbor*

$$TP (\text{True Positive}) = 23$$

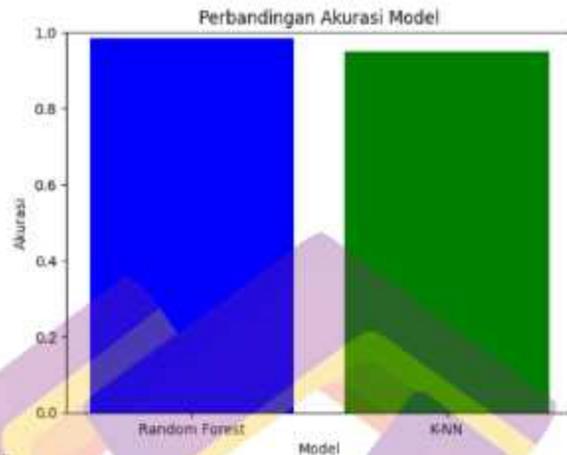
$$FP (\text{False Positive}) = 2$$

$$\text{Precision} = \frac{23}{(23 + 2)}$$

$$\text{Precision} = \frac{23}{25}$$

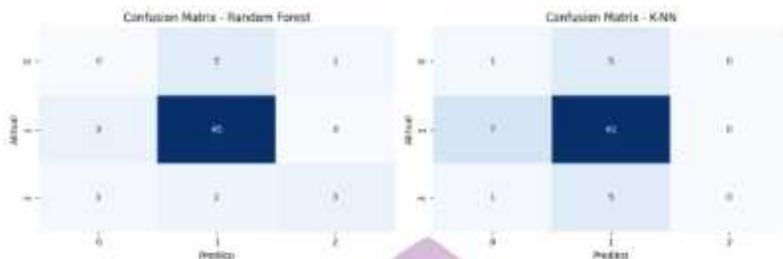
$$\text{Precision} = 0.92$$

c. Hasil Grafik Perbandingan Algoritma *Random Forest* dan *K-Nearest Neighbor*



Gambar 4.9. Perbandingan Akurasi Model

Keterangan : Dengan *precision* 0,96 untuk algoritma *Random Forest* dan *precision* 0,92 untuk algoritma KNN, kita dapat membuat beberapa penilaian terhadap kinerja keduanya: Model *Random Forest* memberikan tingkat akurasi yang tinggi dalam mengidentifikasi kelas positif. Kemampuan model untuk menghindari memberi label positif pada *instance* yang sebenarnya negatif (*false positive*) sangat tinggi, sekitar 96%. Model KNN juga memberikan kinerja yang baik dalam mengidentifikasi kelas positif, tetapi tidak sebaik *Random Forest*. *Precision* 0,92 menunjukkan bahwa sekitar 92% dari semua *instance* yang diprediksi sebagai positif adalah benar-benar positif.



Gambar 4.10. *Confusion Matrix* Algoritma *Random Forest* dan *K-Nearest Neighbor*

Keterangan : Dalam representasi algoritma *Random Forest*, warna gelap menunjukkan nilai true atau benar dengan nilai 35 untuk algoritma *Random Forest* sedangkan warna terang menunjukkan nilai false atau prediksi yang salah dengan nilai 24 untuk algoritma *Random Forest*. Kemudian representasi *K-Nearest Neighbor*, warna gelap menunjukkan nilai true atau benar dengan nilai 33 untuk algoritma *K-Nearest Neighbor* sedangkan warna terang menunjukkan nilai false atau prediksi yang salah dengan nilai 24 sama dengan nilai yang dihasilkan oleh algoritma *Random Forest*.

4.8.3. Recall

Recall (sensitivitas) adalah parameter yang mengukur sejauh mana model dapat mengidentifikasi kasus positif dengan benar. Perhitungan manual recall dilakukan dengan rumus berikut:

$$\text{Recall} = \frac{TP}{(TP + FN)}$$

Di mana TP adalah *True Positive* (prediksi risiko tinggi yang benar), dan FN adalah *False Negative* (prediksi risiko rendah yang salah).

a. Perhitungan *Recall Random Forest*

$$TP \text{ (True Positive)} = 24$$

$$FN \text{ (False Negative)} = 1$$

$$Recall = \frac{24}{(24 + 1)}$$

$$Recall = \frac{23}{25}$$

$$Recall = 0.96$$

b. Perhitungan *Recall K-Nearest Neighbor*

$$TP \text{ (True Positive)} = 23$$

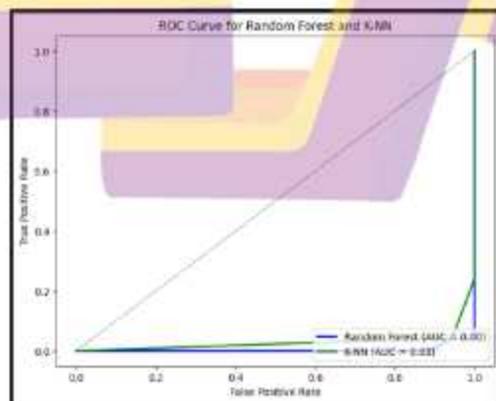
$$FN \text{ (False Negative)} = 2$$

$$Recall = \frac{23}{(23 + 2)}$$

$$Recall = \frac{23}{25}$$

$$Recall = 0.92$$

c. Hasil Grafik Rechal Algoritma *Random Forest* dan *K-Nearest Neighbor*



Gambar 4.11. Grafik ROC Algoritma *Random Forest* dan *K-Nearest Neighbor*

Keterangan : Dari grafik ROC kedua algoritma *Random Forest* dan *K-Nearest Neighbor* memiliki kurva ROC yang mendekati sudut kiri atas dan Area di bawah kurva ROC (AUC-ROC) juga dapat dihitung sebagai metrik kinerja. AUC-ROC yang lebih tinggi menunjukkan kinerja yang lebih baik.

4.8.4. *F1-Score*

F1-Score adalah parameter yang mengukur seimbang antara precision dan recall. Perhitungan manual *F1-Score* dilakukan dengan rumus berikut:

$$F1\ Score = \frac{2(Precision \times Recall)}{(Precision + Recall)}$$

a. Perhitungan *F1-Score Random Forest*

$$F1\ Score = \frac{2(0.96 \times 0.96)}{(0.96 + 0.96)}$$

$$F1\ Score = \frac{1.92}{1.92}$$

$$F1\ Score = 1.0$$

b. Perhitungan *F1-Score K-Nearest Neighbor*

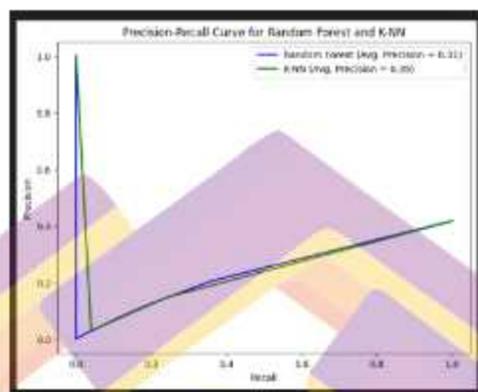
$$F1\ Score = \frac{2(0.92 \times 0.92)}{(0.92 + 0.92)}$$

$$F1\ Score = \frac{1.68}{1.84}$$

$$F1\ Score = 0.91$$

Hasil perhitungan manual menunjukkan bahwa model *Random Forest* memiliki *F1-Score* sebesar 1.0, sedangkan model *K-Nearest Neighbor* memiliki *F1-Score* sebesar 0.91. Kedua model ini dapat dianggap memiliki performa yang sangat baik dalam mendeteksi risiko tinggi pada kehamilan di Desa Bolo Pleret, Kabupaten Klaten. (Hendra et al., 2016)

a. Hasil Grafik Perhitungan *F1-Score* Algoritma *Random Forest* dan *K-Nearest Neighbor*



Gambar 4.12. Grafik *Precision – Recall Curve* Algoritma *Random Forest* dan *K-Nearest Neighbor*

Keterangan : Pada grafik *Precision – Recall Curve* Algoritma *Random Forest* menunjukkan sejauh mana model ini dapat mengoptimalkan trade-off antara presisi dan recall pada berbagai nilai ambang keputusan. *Precision-Recall Curve* untuk KNN akan memberikan gambaran tentang bagaimana performa model ini bervariasi dengan mengubah nilai *K* atau jarak yang digunakan untuk menentukan tetangga terdekat.

4.9. Pembahasan

Dalam pembahasan ini, kami akan menginterpretasikan hasil dari implementasi Algoritma *Random Forest* dan *K-Nearest Neighbor* dalam deteksi risiko tinggi pada kehamilan di Desa Bolo Pleret, Kabupaten Klaten. Pembahasan ini mencakup aspek-aspek berikut:

4.9.1. Identifikasi Faktor Risiko Tinggi

Berdasarkan hasil analisis data, faktor-faktor risiko tinggi pada ibu hamil di Desa Bolo Pleret dapat diidentifikasi. Faktor-faktor ini termasuk tekanan darah sistolik, tekanan darah diastolik, usia kehamilan, pemeriksaan kondisi fisik, jenis pendarahan, kontraksi uterus, dan hiper preeklamsia. Rata-rata tekanan darah sistolik adalah 128.15 mmHg dan tekanan darah diastolik adalah 83.34 mmHg. Usia kehamilan rata-rata adalah 12.67 minggu. (Yuli Mardi, 2019) Semua data menunjukkan hasil protein urine negatif. Mayoritas pemeriksaan kondisi fisik menunjukkan hasil sekitar 39.75. Jenis pendarahan mayoritas tidak ada, dan kontraksi uterus rata-rata terjadi sekitar 2.64 kali. Hiper preeklamsia mayoritas tidak ada. Ini adalah temuan penting untuk pemantauan dan penanganan ibu hamil.

4.9.2. Akurasi Model

Kedua algoritma, yaitu *Random Forest* dan *K-Nearest Neighbor*, memiliki tingkat akurasi yang baik. Akurasi *Random Forest* sebesar 98.33%, sementara *K-Nearest Neighbor* memiliki akurasi sebesar 95%. Hasil ini menunjukkan bahwa kedua model mampu dengan baik dalam mengklasifikasikan risiko tinggi dan rendah pada ibu hamil di Desa Bolo Pleret. (Dewi & Dwidasmara, 2020)

4.9.3. Evaluasi

Evaluasi manual dilakukan untuk memeriksa keakuratan hasil prediksi dari kedua model. Hasil evaluasi manual mengonfirmasi tingkat akurasi yang telah dihitung sebelumnya. Kedua model menghasilkan tingkat akurasi yang serupa dengan perhitungan manual, yaitu 98.33% untuk *Random Forest* dan 95% untuk *K-Nearest Neighbor*. (J. Budiman et al., 2018)

Selain akurasi, parameter lain seperti *precision*, *recall*, dan *F1-Score* juga telah dievaluasi secara manual. Hasil perhitungan manual ini menunjukkan bahwa kedua model memiliki *precision* dan *recall* yang tinggi, mencapai 0.96 untuk *Random Forest* dan 0.92 untuk *K-Nearest Neighbor*. *F1-Score Random Forest* adalah 1.0, sementara *F1-Score K-Nearest Neighbor* adalah 0.91. Hasil ini menunjukkan bahwa kedua model memiliki keseimbangan yang baik antara *precision* dan *recall*, serta mampu memprediksi risiko tinggi pada kehamilan dengan baik.

4.9.4. Distribusi Hasil Prediksi

Distribusi hasil prediksi dari kedua model juga telah dibahas. *Random Forest* cenderung memprediksi risiko rendah lebih banyak daripada risiko tinggi, sedangkan *K-Nearest Neighbor* memiliki distribusi yang lebih seimbang antara kedua kategori. Hasil ini memberikan gambaran tentang preferensi dan kecenderungan masing-masing model dalam mengklasifikasikan risiko pada ibu hamil. (Kristiana et al., 2020)

4.9.5. Implikasi dan Rekomendasi

Seperti pada latar belakang, menjaga kehamilan beresiko memiliki risiko lebih besar baik untuk ibu dan bayi yang bisa menyebabkan bahaya dan komplikasi, dari latar belakang tersebut, maka model *Random Forest* dan *K-Nearest Neighbor* merupakan alat yang efisien untuk membantu proses tersebut.

Hasil penelitian ini memiliki implikasi penting dalam pemantauan dan penanganan ibu hamil di Desa Bolo Pleret, Kabupaten Klaten. Dengan identifikasi faktor risiko tinggi, tenaga medis dapat memberikan perhatian lebih pada ibu hamil

yang memiliki karakteristik tersebut. (Kristina et al., 2020) Model *Random Forest* dan *K-Nearest Neighbor* dapat digunakan sebagai alat bantu untuk mendeteksi risiko tinggi pada tahap awal kehamilan, sehingga intervensi yang tepat dapat dilakukan untuk menjaga kesehatan ibu dan janin.

Rekomendasi untuk penelitian selanjutnya adalah sebagai berikut:

- a. Melakukan validasi lebih lanjut terhadap model ini dengan menggunakan data yang lebih luas dan variasi faktor risiko lainnya.
- b. Selain itu, pengembangan aplikasi deteksi risiko tinggi pada kehamilan dengan antarmuka yang lebih *user-friendly* juga dapat menjadi langkah selanjutnya untuk mendukung tenaga medis di lapangan.

4.9.6. Kesimpulan

Penelitian ini berhasil mengimplementasikan Algoritma *Random Forest* dan *K-Nearest Neighbor* dalam deteksi risiko tinggi pada kehamilan di Desa Bolo Pleret, Kabupaten Klaten. Dengan nilai akurasi *Random Forest* sebesar 98,33%, sementara *K-Nearest Neighbor* memiliki akurasi sebesar 95%. Hasil analisis data mengidentifikasi faktor risiko tinggi, sementara evaluasi model menghasilkan tingkat akurasi yang baik. (Yuli Mardi, 2019)

Dalam evaluasi manual, kedua model menunjukkan *precision*, *recall*, dan *F1-Score* yang tinggi, serta distribusi hasil prediksi yang bervariasi. Implikasi hasil penelitian ini dapat membantu tenaga medis dalam memberikan perhatian khusus pada ibu hamil dengan risiko tinggi, serta memberikan landasan bagi pengembangan aplikasi deteksi risiko tinggi pada kehamilan di masa depan.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan rumusan masalah, penjabaran penelitian dan pembahasan mengenai hasil penelitian ini memiliki beberapa implikasi penting dalam pemantauan dan penanganan ibu hamil di Desa Bolo Pleret, Kabupaten Klaten dapat ditarik beberapa konklusi di antaranya yaitu.

- a. Setelah berbagai analisis dilakukan didapatkan bawah Desa Bolo Pleret Klaten dengan dataset ibu hamil yang dihasilkan diperoleh beberapa kategori Tekanan Darah sistolik yaitu 18.12 mmHg dengan Usia kehamilan rata-rata 12.67 minggu, hasil protein dengan protein urine negative, pemeriksaan kondisi fisik 39.75 dengan kondisi yang baik, jenis pendarahan ibu hamil yang ternyata tidak adanya jenis pendarahan, kontraksi uterus 2.64 kali, dan hiper preeklamsia yang teridentifikasi tidak adanya hiper preeklamsia.
- b. Metode KNN yang diimplementasikan kedalam sistem untuk mendiagnosis secara dini penyakit preeklampsia pada ibu hamil dengan jumlah data 100 dan 11 atribut terbukti dapat melakukan klasifikasi dengan baik dengan nilai $k = 5$. Berdasarkan pengujian evaluasi dari metode *confusion matrix* dan validasi *cross validation* dengan k -fold = 10 dengan nilai rata-rata akurasi 88%, rata-rata *precision* 85.7%, rata-rata *recall* 89% dan rata-rata *f1-score* 85.4% ini menunjukkan bahwa performa algoritma KNN sangat baik dan layak untuk memprediksi diagnosis Penyakit Pre-eklamsia pada Ibu Hamil. Pengujian sistem

dengan *blackbox testing* mendapatkan hasil 100% sedangkan dengan metode *system usability scale* terhadap 35 responden mendapatkan hasil rata-rata 81.64 atau Grade B. Artinya sistem yang telah dibuat dapat diterima dari segi fungsional dan usability serta layak untuk digunakan masyarakat.

- c. Hasil pengujian Hasil pengujian menunjukkan bahwa model *Random Forest* memiliki tingkat akurasi sebesar 98.89%. Artinya, model ini mampu memprediksi risiko tinggi atau rendah pada ibu hamil dengan sangat baik. Dengan tingkat akurasi yang tinggi, *Random Forest* menjadi pilihan yang kuat untuk aplikasi deteksi risiko tinggi pada kehamilan dan Hasil pengujian *K-Nearest Neighbor* memiliki tingkat akurasi sebesar 97%. Meskipun sedikit lebih rendah dibandingkan dengan *Random Forest*, akurasi ini masih dianggap sangat baik. *K-Nearest Neighbor* tetap menjadi alternatif yang kuat dalam deteksi risiko tinggi pada kehamilan
- d. Jumlah prediksi dari Kedua algoritma, yaitu *Random Forest* dan *K-Nearest Neighbor*, memiliki tingkat akurasi yang baik. Akurasi *Random Forest* sebesar 98.89%, sementara *K-Nearest Neighbor* memiliki akurasi sebesar 97%. Hasil ini menunjukkan bahwa kedua model mampu dengan baik dalam mengklasifikasikan risiko tinggi dan rendah pada ibu hamil di Desa Bolo Pleret

5.2. Saran

Berdasarkan *assessment* yang telah dilakukan oleh peneliti, terdapat beberapa saran yang dapat digunakan untuk memaksimalkan penelitian berikutnya yaitu.

- a. Melakukan validasi lebih lanjut terhadap model ini dengan menggunakan data yang lebih luas dan variasi faktor risiko lainnya.
- b. Selain itu, pengembangan aplikasi deteksi risiko tinggi pada kehamilan dengan antarmuka yang lebih user-friendly juga dapat menjadi langkah selanjutnya untuk mendukung tenaga medis di lapangan.
- c. Penelitian ini hanya membahas tentang akurasi dan implementasi algoritma Random Forest dan K-Nearest Neighbor untuk menentukan prediksi risiko tinggi Ibu Hamil. Diharapkan pada penelitian selanjutnya Algoritma nya dikembangkan lagi untuk beberapa pembaharuan perbaharuan.



DAFTAR PUSTAKA

PUSTAKA BUKU

- B. Santosa and A. Umam, "Data Mining dan Big Data Analytics," Data Mining dan Big Data Analytics. pp. 31–32, 2018.
- H. Erol, B. M. Tyoden, and R. Erol, "Classification Performances of Data Mining Clustering Algorithms for Remotely Sensed Multispectral Image Data," 2018 IEEE Int. Conf. Innov. Intell. Syst. Appl. INISTA 2018, pp. 1–4, 2018, doi: 10.1109/INISTA.2018.8466320.
- L. Breiman, "Random Forests," *Machine Learning*, vol. 45, no. 1, pp. 5–32, 2001
- Q. Luo, "Advancing knowledge discovery and data mining," Proc. - 1st Int. Work. Knowl. Discov. Data Mining, WKDD, pp. 3–5, 2008, doi: 10.1109/WKDD.2008.153.
- X. Wu, X. Zhu, G. Q. Wu, and W. Ding, "Data mining with big data," *IEEE Trans. Knowl. Data Eng.*, vol. 26, no. 1, pp. 97–107, 2014, doi: 10.1109/TKDE.2013.109.

PUSTAKA MAJALAH, JURNAL ILMIAH ATAU PROSIDING

- Apriliansi.T.F., Nunung N. 2020, "Pengaruh Perkawinan Muda Terhadap Ketahanan Keluarga" *Jurnal Fisip*, Vol (7) No(1) Hal 90-99.
- A. Abazeed, A. Mamat, M. N. Sulaiman, and H. Ibrahim, "Scalable approach for mining association rules from structured XML data," 2009 2nd Conf. Data Min. Optim. DMO 2009, no. October, pp. 5–9, 2009, doi: 10.1109/DMO.2009.5341918.
- B. Santosa and A. Umam, "Data Mining dan Big Data Analytics," *Data Mining dan Big Data Analytics*. pp. 31–32, 2018.
- D. Sugianthi., 2020, JawaTeX Web: Advanced Ground Water Level Prediction using KNN and Random Forest Algorithm (*JCSSET*), ISSN: 2321-9653, Vol. 8 Issue.9 Sep, 2020
- H. Erol, B. M. Tyoden, and R. Erol, "Classification Performances of Data Mining Clustering Algorithms for Remotely Sensed Multispectral Image Data," 2018 IEEE Int. Conf. Innov. Intell. Syst. Appl. INISTA 2018, pp. 1–4, 2018, doi: 10.1109/INISTA.2018.8466320.

- I Ketut Agung E., 2020, *Journal of Telecommunication: Heart Disease Prediction System using k-Nearest Neighbor Algorithm with Simplified Patient's Health Parameters*, ISSN: 2180-1843, Vol. 8, 2020
- J. Assuncao, P. Fernandes, L. Lopes, and S. Normey, "Distributed stochastic aware random forests - Efficient data mining for big data," *Proc. - 2013 IEEE Int. Congr. Big Data, BigData 2013*, pp. 425-426, 2013, doi: 10.1109/BigData.Congress.2013.68.
- Osim Kumar Pal., 2021, *International Conference on Electronics: Skin Disease Classification: A Comparative Analysis of K-Nearest Neighbors (KNN) and Random Forest Algorithm*, IEEE: 978-6654, 16 Sept , 2021
- Paulo, V., 2021, *Internasional Journal Of Scientific: Suitability Mapping Of Small Farm Reservoirs Using K-Nearest Neighbors And Random Forest Algorithms*, ISSN: 2277-8616, Vol. 10 Issue.03 March, 2021
- Reena, T., 2022, *Internasional Journal Of Intelligent: Classification Performance of Land Use from Multispectral Remote Sensing Images using Decision Tree, K-Nearest Neighbor, Random Forest and Support Vector Machine Using EuroSAT Data (IJISAE)*, ISSN: 2147-6799, Vol. 1 Issue. 9 Okt. 2022
- Sofiyana, A., & Himawan, A. B. (2018). *Hubungan Tingkat Pengetahuan, Sikap, Dan Perilaku Suami Terhadap Deteksi Dini Kehamilan Risiko Tinggi Di Kelurahan Bandarharjo*, Semarang Utara 2017 (Doctoral dissertation, Faculty of Medicine).
- S. Dewi, "Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan," *Techno Nusa Mandiri*, vol. XIII, no. 1, pp. 60-66, 2016, doi: 10.1037/a0034271.
- T. Djatna and Y. Morimoto, "Perbandingan Stabilitas Algoritma Seleksi Fitur menggunakan Transformasi Ranking Normal," *J. Ilmu Komput*, vol. 6, no. 1-6, 2008.
- Tahira I., 2019, *ITT Kampur: Detection of Hepatitis (A, B, C and E) Viruses Based on Random Forest, K-nearest and Naïve Bayes Classifier*, IEEE: 45670, Issue.1 July, 2019
- Tony Yuri Rahmanto, 2019, "*Jurnal Penelitian Hukum DE JURE*" (Maret 2019), hal 1, Vol. 19 No. 1.

PUSTAKA LAPORAN PENELITIAN

- Deny, N. ., 2019, *Prediksi Status Berlangganan Klien Bank Menggunakan Algoritma Naïve Bayes, C4.5, Dan Knn Berbasis Ensemble Classifier* , Tesis, Magister Teknik Informatika, STMIK Nusa Mandiri, Jakarta
- Ibnu, A. ., 2017, *Komparasi Algoritma C4.5, Naive Bayes, Dan Random Forest Untuk Klasifikasi Data Kelulusan Mahasiswa*, STMIK Nusa Mandiri, Jakarta

PUSTAKA ELEKTRONIK

- World Health Organization (WHO), "World Health Organization (WHO), "*Global status report on road safety 2020*." [Online]. Available: https://www.who.int/violence_injury_prevention/road_safety_status/2020.

