

**IMPLEMENTASI METODE K-NN UNTUK PENENTUAN STATUS
KESEHATAN JANIN**

SKRIPSI

Diajukan untuk Memenuhi Salah Satu Persyaratan dalam Menyelesaikan Program
Sarjana (S-1) pada Program Studi Teknologi Informasi
Fakultas Sains dan Teknologi Universitas Ibrahimy

Oleh:

MAULIDATUL MAWADDAH

2020503050

PROGRAM STUDI TEKNOLOGI INFORMASI

FAKULTAS SAINS DAN TEKNOLOGI UNIVERSITAS IBRAHIMY

2024

PERNYATAAN KEASLIAN TULISAN

Yang bertanda tangan di bawah ini :

Nama : Maulidatul Mawaddah

NPM : 2020503050

Prodi : Teknologi Informasi

Menyatakan dengan sebenarnya, bahwa tugas akhir/skripsi ini secara keseluruhan adalah hasil penelitian atau karya saya sendiri, kecuali pada bagian-bagian yang dirujuk sebagai sumber referensi dan disebutkan dalam duftar pustaka. Apabila di kemudian hari terbukti atau dapat dibuktikan bahwa tugas akbir/skripsi ini hasil plagiasi, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Situbondo, 31 Agustus 2024



Maulidatul Mawaddah

LEMBAR PERNYATAAN**KESEDIAAN PUBLIKASI KARYA ILMIAH**

Saya yang bertanda tangan di bawah ini :

Nama : Maulidatul Mawaddah
NPM : 2020502046
Program Studi : S-1 Teknologi Informasi
Fakultas : Fakultas Sains dan Teknologi
Jenis Karya Ilmiah : Hasil Penelitian

Demikian pengembangan ilmu pengetahuan, menyetujui untuk memberikan Hak Bebas Royalti Non-eksklusif (*Non-exclusive Royalty-Free Right*) kepada Perpustakaan Universitas Ibrahimi atas karya ilmiah saya berupa hasil penelitian yang berjudul :

IMPLEMENTASI METODE K-NN UNTUK PENENTUAN STATUS KESEHATAN JANIN

Dengan Hak Bebas Royalti Non-eksklusif ini Pusat Perpustakaan Universitas Ibrahimi berhak menyimpan, alih media/format, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat untuk dapat dipergunakan sebagaimana mestinya.

Situbondo, 31 Agustus 2024

Saya Menyatakan,


Maulidatul Mawaddah

PERSETUJUAN PEMBIMBING

Nama : **Maulidatul Mawaddah**
NPM/NIM : **2020503050**
Judul : **IMPLEMENTASI METODE K-NN UNTUK PENENTUAN
STATUS KESEHATAN JANIN**

Telah ditelaah dan disetujui oleh pembimbing untuk diujikan pada sidang
skripsi/munaqosah




Disetujui oleh :

Pembimbing 1,



Ahmad Homaidi, M.Kom
NIDN : 705078901

Pembimbing 2,



Lukman Fakhid Lidimillah, M.Kom
NIDN : 715099001

HALAMAN PENGESAHAN

SKRIPSI

IMPLEMENTASI METODE K-NN UNTUK PENENTUAN STATUS KESEHATAN JANIN

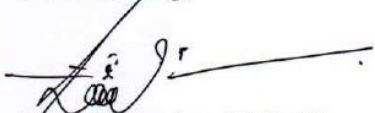
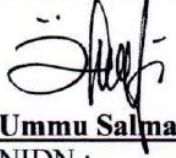
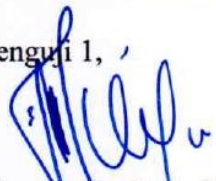
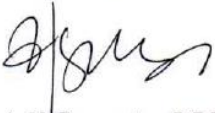
MAULIDATUL MAWADDAH

2020503060

telah dipertahankan di depan dewan penguji Sidang/Munaqosah Skripsi pada hari Selasa, 06 Agustus 2024 sebagai salah satu syarat memperoleh gelar Sarjana (S.Kom) pada Fakultas Sains dan Teknologi Universitas Ibrahimiy



Tim Penguji,

<p>Ketua Sidang,</p>  <p><u>Farihin Lazim, M.Tr.T</u> NIDN : 0711099201</p>	<p>Sekretaris Sidang,</p>  <p><u>Ummu Salmah, Amd.Pi</u> NIDN :</p>
<p>Penguji 1,</p>  <p><u>Firman Santoso, M.Kom</u> NIDN : 722129201</p>	<p>Penguji 2,</p>  <p><u>Adi Susanto, M.Kom</u> NIDN: 708079104</p>

Mengetahui
Dekan,



Abd. Ghofur, M.Kom
NIDN: 0711088303

MOTTO

بَلِ الْإِنْسَانُ عَلَىٰ نَفْسِهِ بَصِيرَةٌ

“Jangan pedulikan apa yang dikatakan orang lain mengenaimu, engkau tau siapa dirimu dan lebih tau keadaan dirimu dan niat yang ada dalam hatimu” (Q.S Al-Qiyamah : 14)



PERSEMBAHAN

Dengan segenap rasa syukur yang tiada terkira kepada Allah yang Maha Segalanya, Laporan ini saya persembahkan kepada :

1. KHR. Ach. Azaim Ibrahimi, S, Sy M. HI Selaku Pengasuh Pondok Pesantren Salafiyah Syafi'iyah Sukorejo Situbondo beserta sekeluarga besar.
2. Kedua orang tua beserta seluruh keluarga yang telah berjuang demi masa depan saya, dan terima kasih atas jerih payah beserta doa dan dukungan kalian yang tiada batas.
3. Seluruh Dosen Fakultas Sains dan Teknologi yang selalu memberi support dan motivasi.
4. Keluarga kedua saya Nurul, Sopi, Lafa, Sila, Ulfa, Huril, Linda, Khofifah, Farida dan seluruh adik – adik dikamar D.07 (yang tidak bisa saya sebutkan namanya satu persatu) yang selalu menjadi tempat keluh kesah saya, terima kasih untuk dukungan dan motivasi kalian.
5. Sahabat terbaik dan seperjuangan ANTM dan Bani Julaiman, yang selalu menjadi teman sekaligus tempat berbagi selama ini.
6. Teman-teman saya di Universitas Ibrahimi khususnya prodi Teknologi Informasi 2020 yang menegur disaat saya salah dan membantu saya ketika ada kesusahan dalam mata kuliah.
7. Terakhir, terima kasih kepada diri sendiri, karena telah mampu berusaha keras berjuang sejauh ini. Mampu mengendalikan diri dari berbagai tekanan diluar keadaan dan tak pernah memutuskan untuk menyerah sesulit apapun proses penyusunan skripsi dari awal hingga akhir dengan menyelesaikan skripsi ini dengan sebaik dan semaksimal mungkin, ini merupakan pencapaian yang patut dibanggakan untuk diri sendiri.

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT. Karena atas Rahman dan Rahim-Nya lah penulis dapat menyelesaikan Laporan Praktek Kerja Lapangan (PKL) ini. Laporan PKL ini dilakukan untuk memenuhi syarat ujian Praktek Kerja Lapangan. Penyusunan Laporan ini tidaklah terlepas dari pihak-pihak yang telah banyak membantu dalam hal segala apapun. Pada kesempatan kali ini, penulis ingin mengucapkan terima kasih kepada :

1. KHR. Ach. Azaim Ibrahimi, S, Sy M. HI selaku Pengasuh Pondok Pesantren Salafiyah Syafi'iyah Sukorejo Situbondo.
2. Bapak KH. Ach. Fadlail, S.H, M.H selaku Rektor Universitas Ibrahimi Situbondo.
3. Bapak Abdul Ghofur, M.Kom selaku Dekan Fakultas Sains dan Teknologi Universitas Ibrahimi Situbondo.
4. Bapak Firman Santoso, M.Kom selaku Ka.Prodi Teknologi Informasi Universitas Ibrahimi Situbondo.
5. Bapak Ahmad Homaidi, M.Kom selaku pembimbing I dan Bapak Lukman Fakhid Lidimillah, M.Kom selaku pembimbing II yang telah banyak memberikan bimbingan, masukan, dan arahan selama penyusunan skripsi ini.
6. Seluruh dosen dan civitas akademika Universitas Ibrahimi Sukorejo Situbondo yang dengan tulus ikhlas banyak memberi bekal ilmu kepada peneliti selama menempuh pendidikan dari awal hingga akhir semester.

Situbondo, 30 Agustus 2024

Penulis,

Maulidatul Mawaddah

DAFTAR ISI

	Halaman
PERNYATAAN KEASLIAN TULISAN	i
LEMBAR PERNYATAAN	ii
KESEDIAAN PUBLIKASI KARYA ILMIAH	ii
PERSETUJUAN PEMBIMBING	iii
HALAMAN PENGESAHAN	iii
MOTTO	iv
PERSEMBAHAN	v
KATA PENGANTAR	vi
DAFTAR ISI	vii
DAFTAR TABEL	x
DAFTAR GAMBAR	xi
ABSTRAK	xii
ABSTRACT	xiii
BAB I	1
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Identifikasi Masalah	4
1.3 Rumusan Masalah	4
1.4 Batasan Masalah	5
1.5 Tujuan Penelitian	5
1.6 Manfaat Penelitian	5
1.7 Metode Penelitian	6
1.7.1 Jenis penelitian	6
1.7.2 Tahapan pengumpulan data	6
1.8 Sistematika Pembahasan	7
BAB II	9
TINJAUAN PUSTAKA	9
2.1 Penelitian Terdahulu	9
2.2 Landasan Teori	12
2.2.1 Data Mining	12

2.2.2 Algoritma K-Nearest Neighbor	13
2.2.3 Dataset	14
2.2.4 Atribut	15
2.3 Perangkat Lunak yang Digunakan.....	16
BAB III.....	18
METODOLOGI PENELITIAN	18
3.1 Metode Penelitian	18
3.2 Tahapan Penelitian.....	18
3.2.1 Pengumpulan Data.....	19
3.2.2 Penentuan Atribut.....	20
3.2.3 Pemisahan Data.....	34
3.3 Perancangan Sistem	35
3.3.1 Metode algoritma K-NN.....	35
3.4 Implementasi dan pengujian Metode.....	36
BAB IV.....	39
HASIL DAN PEMBAHASAN	39
4.1 Hasil Penelitian	39
4.2 Perhitungan Manual	39
4.2.1 Menentukan atribut dan label.....	39
4.2.3 Menentukan nilai K terbaik	40
4.2.4 Menghitung jarak antara data training dan data testing	40
4.2.5 Identifikasi tetangga terdekat.....	44
4.2.6 Menghitung nilai Accuracy.....	45
4.3 Perhitungan RapidMiner.....	46
4.3.1 Pemrosesan Data.....	46
4.3.2 Perhitungan dan validasi	47
4.3.3 Hasil Performance Vector.....	49
BAB V.....	52
PENUTUP.....	52
5.1 Kesimpulan	52
5.2 Saran.....	53
DAFTAR PUSTAKA	54

LAMPIRAN.....	57
CURICULUM VITAE.....	62



DAFTAR TABEL

Tabel 3. 1	21
Tabel 3. 2	21
Tabel 3. 3	22
Tabel 3. 4	23
Tabel 3. 5	23
Tabel 3. 6	24
Tabel 3. 7	24
Tabel 3. 8	25
Tabel 3. 9	26
Tabel 3. 10	27
Tabel 3. 11	28
Tabel 3. 12	34
Tabel 4. 1	45



DAFTAR GAMBAR

Gambar 3. 1	19
Gambar 3. 2	28
Gambar 3. 3	29
Gambar 3. 4	29
Gambar 3. 5	30
Gambar 3. 6	31
Gambar 3. 7	31
Gambar 3. 8	32
Gambar 3. 9	32
Gambar 3. 10	33
Gambar 3. 11	33
Gambar 3. 12	35
Gambar 3. 13	37
Gambar 3. 14	37
Gambar 3.15	36
Gambar 4. 1	39
Gambar 4. 2	40
Gambar 4. 3	40
Gambar 4. 4	47
Gambar 4. 5	48
Gambar 4. 6	49
Gambar 4. 7	50
Gambar 4. 8	50
Gambar 4. 9	51

ABSTRAK

Maulidatul Mawaddah, 2024. **Implementasi Metode KNN Untuk Penentuan Status Kesehatan Janin.** Skripsi, Program Studi Teknologi Informasi, Universitas Ibrahimi. Pembimbing (I) Ahmad Homaidi, M.Kom., Pembimbing (II) Lukman Fakhri Lidimillah, M.Kom.

Penentuan status kesehatan janin merupakan aspek krusial dalam pemantauan kehamilan untuk mengurangi resiko komplikasi dan meningkatkan keselamatan ibu dan bayi. Metode K-Nearest Neighbors (KNN) telah diimplementasikan sebagai teknik pengklasifikasian dalam penentuan status kesehatan janin berdasarkan data cardiotocography (CTG). Studi ini memaparkan penggunaan algoritma KNN untuk menganalisis berbagai parameter CTG, termasuk detak jantung janin dan frekuensi kontraksi uterus, untuk mengklasifikasikan status kesehatan janin menjadi tiga kategori: normal, suspect, dan pathologic. Proses implementasi melibatkan pengumpulan data normalisasi, pemilihan fitur relevan, dan penggunaan algoritma KNN dengan variasi nilai K untuk menentukan nilai yang paling optimal. Hasil penelitian menunjukkan bahwa metode KNN dengan nilai K yang tepat dapat mencapai akurasi tinggi dalam klasifikasi status kesehatan janin, dengan akurasi mencapai hingga 89%. Temuan ini menunjukkan bahwa KNN merupakan metode yang efektif dan dapat diandalkan dalam mendukung tenaga medis untuk mengambil keputusan berdasarkan CTG, yang pada akhirnya dapat meningkatkan kualitas perawatan kesehatan ibu dan bayi. Selain itu, implementasi metode ini relatif sederhana dan dapat diintegrasikan kedalam sistem kesehatan yang sudah ada tanpa memerlukan sumber daya komputasi yang besar. Penelitian lebih lanjut disarankan untuk membandingkan performa KNN dengan metode machine learning lainnya seperti Support Vector Machine (SVM) dan Random Forest untuk mengidentifikasi metode terbaik dalam konteks ini. Penggunaan data yang lebih besar dan beragam juga diharapkan dapat meningkatkan akurasi dan generalisasi model dalam berbagai kondisi klinis.

Kata kunci: *Cardiotocography, Data mining, K-Nearest Neighbors, klasifikasi, status kesehatan janin*

ABSTRACT

Maulidatul Mawaddah, 2024. **Implementation of the KNN Method for Determining Fetal Health Status**. Thesis, Information Technology Study Program, Ibrahimi University. Supervisor (I) Ahmad Homaidi, M.Kom., Supervisor (II) Lukman Fakhri Lidimillah, M.Kom.

Determining the health status of the fetus is a crucial aspect of pregnancy monitoring to reduce the risk of complications and increase the safety of the mother and baby. The K-Nearest Neighbors (KNN) method has been implemented as a classification technique in determining fetal health status based on cardiotocography (CTG) data. This study describes the use of the KNN algorithm to analyze various CTG parameters, including fetal heart rate and uterine contraction frequency, to classify fetal health status into three categories: normal, suspect, and pathologic. The implementation process involves collecting normalized data, selecting relevant features, and using the KNN algorithm with varying K values to determine the most optimal value. The research results show that the KNN method with the right K value can achieve high accuracy in classifying fetal health status, with accuracy reaching up to 89%. These findings indicate that KNN is an effective and reliable method in supporting medical personnel to make decisions based on CTG, which can ultimately improve the quality of maternal and infant health care. In addition, the implementation of this method is relatively simple and can be integrated into existing health systems without requiring large computing resources. Further research is recommended to compare the performance of KNN with other machine learning methods such as Support Vector Machine (SVM) and Random Forest to identify the best method in this context. The use of larger and more diverse data is also expected to increase the accuracy and generalization of the model in various clinical conditions.

Keywords: *Cardiotocography, classification, data mining, fetal health status, K-Nearest Neighbors*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Penentuan status kesehatan janin merupakan aspek penting dalam pemantauan kehamilan untuk memastikan kesejahteraan ibu dan bayi. Metode tradisional untuk pemantauan janin seringkali bergantung pada keahlian dan pengalaman klinis dokter, yang dapat bervariasi antara satu individu dengan individu lainnya. Dalam beberapa dekade terakhir, kemajuan teknologi telah memungkinkan penggunaan alat diagnostik yang lebih canggih, seperti cardiotocography (CTG), untuk memantau kondisi janin. CTG adalah metode non-invasif yang merekam detak jantung janin dan aktivitas uterus, memberikan data yang kaya untuk analisis lebih lanjut .[1]

Namun, interpretasi data CTG secara manual dapat menjadi proses yang rumit dan rentan terhadap kesalahan. Untuk mengatasi tantangan ini, metode komputasional dan algoritma pembelajaran mesin telah diusulkan untuk membantu dalam interpretasi data CTG. Salah satu metode yang menonjol adalah K-Nearest Neighbors (KNN), sebuah algoritma klasifikasi yang sederhana namun efektif. KNN bekerja dengan mengklasifikasikan data berdasarkan kedekatan dengan data yang telah dilabeli sebelumnya. Algoritma ini memiliki kelebihan dalam hal kesederhanaan dan kemampuan untuk menghasilkan hasil yang akurat dengan jumlah data pelatihan yang cukup .[2]

Pengurangan angka kematian anak tercermin dalam beberapa Tujuan Pembangunan Berkelanjutan PBB dan merupakan indikator utama kemajuan manusia. PBB memperkirakan bahwa pada tahun 2030, negara-negara akan mengakhiri kematian bayi baru lahir dan anak di bawah usia 5 tahun yang dapat dicegah, dan semua negara bertujuan untuk mengurangi angka kematian di bawah 5 tahun setidaknya hingga 25 per 1.000 kelahiran hidup.[3]

Perawatan prenatal mengurangi risiko selama kehamilan dan meningkatkan peluang kelahiran yang aman dan sehat. Bayi yang lahir dari ibu yang tidak mendapatkan pelayanan prenatal mempunyai kemungkinan tiga kali lebih besar untuk dilahirkan dengan berat badan lahir rendah. Bayi dengan berat badan lahir rendah lima kali lebih mungkin meninggal dibandingkan bayi yang ibunya menerima perawatan prenatal. Perawatan prenatal idealnya dimulai setidaknya tiga bulan sebelum upaya untuk hamil. Sebagai salah satu alternatif solusi yang memanfaatkan kemajuan ilmu pengetahuan dan teknologi informasi saat ini, sistem yang dikembangkan dapat digunakan sebagai pembawa informasi untuk mengetahui apakah janin mendapatkan status normal, suspect, atau patologis. Salah satu teknologi informasi yang dapat dimanfaatkan adalah pengembangan sistem pakar yang menentukan status kesehatan janin.[4]

Klasifikasi adalah proses mengidentifikasi objek dalam suatu kategori, kelas, atau kelompok berdasarkan prosedur, definisi, dan karakteristik tertentu. Klasifikasi melibatkan pengklasifikasian dua objek yang hanya termasuk dalam salah satu kategori yang disebut kelas. Klasifikasi dalam data mining adalah teknik umum untuk membagi titik data ke dalam kelas yang berbeda. Hal ini

memungkinkan untuk mengatur semua jenis kumpulan data, dari kumpulan data yang kompleks dan besar hingga kumpulan data yang kecil dan sederhana. terutama tentang penggunaan algoritma yang dapat dengan mudah dimodifikasi untuk meningkatkan kualitas data. Tujuan utama klasifikasi adalah untuk mengasosiasikan variabel yang diinginkan dengan variabel yang diinginkan. Variabel yang diminati harus bertipe kualitatif.[5]

Metode klasifikasi yang umum digunakan adalah K-Nearest Neighbor. Teknik yang digunakan adalah metode KNN yang merupakan salah satu teknik klasifikasi dalam data mining. Penelitian sebelumnya juga telah menggunakan metode KNN untuk mendeteksi pola data dan mengungkap penyakit dermatologis yang mendasarinya untuk memprediksi penyakit dermatologis yang terabaikan namun mematikan. Metode K- Nearest Neighbor juga diyakini berpotensi unggul untuk klasifikasi dokumen dibandingkan metode klasifikasi lainnya dalam hal akurasi dan efisiensi komputasi.[6]

Penelitian sebelumnya telah menunjukkan bahwa KNN dapat diterapkan dengan sukses dalam berbagai bidang medis, termasuk dalam deteksi penyakit jantung, diabetes, dan kondisi medis lainnya . Implementasi KNN dalam penentuan status kesehatan janin bertujuan untuk meningkatkan akurasi diagnosis dan mengurangi beban kerja tenaga medis. Dengan menggunakan parameter CTG seperti detak jantung janin dan frekuensi kontraksi uterus, KNN dapat mengklasifikasikan status kesehatan janin menjadi kategori normal, suspek, atau patologis. Studi ini bertujuan untuk mengeksplorasi efektivitas

KNN dalam klasifikasi status kesehatan janin, mengidentifikasi nilai K yang optimal, dan mengevaluasi performa algoritma dalam konteks klinis.[7]

Melalui implementasi KNN, diharapkan dapat dihasilkan alat bantu yang dapat diandalkan untuk mendukung tenaga medis dalam mengambil keputusan yang tepat, sehingga meningkatkan kualitas perawatan kesehatan ibu dan bayi serta mengurangi risiko komplikasi selama kehamilan dan persalinan .

Oleh karena hal tersebut, maka penulis bermaksud untuk melakukan sebuah penelitian tentang **“Implementasi Metode K-NN Untuk Penentuan Status Kesehatan Janin”**. untuk membantu profesional medis memprediksi status kesehatan janin.

1.2 Identifikasi Masalah

Berdasarkan latar belakang diatas, maka didapatkan suatu identifikasi masalah yaitu:

- a. Diperlukan suatu metode yang dapat menganalisis data kesehatan janin dengan cepat dan akurat untuk membantu tenaga medis dalam penentuan status kesehatan janin.

1.3 Rumusan Masalah

Rumusan masalah yang dapat diambil dari latar belakang diatas yaitu :

- a. Bagaimana cara menentukan status kesehatan janin menggunakan metode KNN?

1.4 Batasan Masalah

Untuk memudahkan penulisan dalam penyusunan penelitian ini, dan agar pembahasan penelitian ini tidak menyimpang dari apa yang telah dirumuskan, maka penulis membatasi permasalahan sebagai berikut:

- a. Penggunaan data yang diteliti didapat dari website kaggle.
- b. Data yang akan diuji dan diolah dalam data mining berupa data kesehatan janin menggunakan metode KNN

1.5 Tujuan Penelitian

Pada dasarnya segala kegiatan yang dilakukan manusia selalu tertuju pada tujuan dan sasaran yang ingin dicapai. Tujuan dari penyusunan proposal ini sebagai berikut :

- a. Untuk menerapkan metode KNN dan mengklasifikasi status kesehatan janin
- b. Untuk memahami faktor kesehatan janin

1.6 Manfaat Penelitian

Penulis berharap analisis dan penelitian ini dapat memberikan kontribusi yang dapat diketahui oleh berbagai pihak khususnya penulis, dan pada umumnya oleh semua pihak yang terlibat dalam penyusunan proposal ini, antara lain:

- a. **Mempermudah pemantauan kesehatan janin:** Metode KNN dapat digunakan untuk memantau kesehatan janin secara berkala, sehingga membantu mengidentifikasi potensi masalah kesehatan sejak dini.

- b. **Membantu mengurangi angka kematian bayi:** Dengan meningkatkan akurasi diagnosis dan membantu pengambilan keputusan medis yang tepat, penelitian ini dapat membantu mengurangi angka kematian bayi.

1.7 Metode Penelitian

1.7.1 Jenis penelitian

- a. Library research

Peneliti mengumpulkan, mengolah, menganalisis data-data terkait kesehatan janin pada ibu hamil dengan algoritma K-Nearest Neighbor (KNN) yang bersumber dari perpustakaan online berupa buku, jurnal maupun artikel.

- b. Deskriptif kuantitatif

Metode penelitian deskriptif kuantitatif dipilih untuk menggambarkan, atau menjelaskan sesuatu apa adanya dan untuk menarik kesimpulan dari data yang disajikan dalam bentuk numerik.

1.7.2 Tahapan pengumpulan data

Penelitian ini menggunakan teknik studi Pustaka yaitu mengumpulkan data dari berbagai sumber seperti buku, jurnal, artikel dan situs-situs yang dapat dipercaya sesuai dengan topik yang digunakan. Teknik pengumpulan data yang digunakan pada penelitian ini sebagai berikut:

a. Teknik Pengumpulan Data Sekunder

Teknik pengumpulan data sekunder yang digunakan oleh penulis adalah data yang disediakan oleh penyedia dataset online melalui website <https://www.kaggle.com/>

b. Studi Pustaka

Teknik pengumpulan data yang kedua adalah studi pustaka. Studi pustaka bertujuan mencari informasi yang mendukung penelitian ini, baik melalui media seperti buku, internet, maupun media informasi lainnya.

1.8 Sistematika Pembahasan

Sistematika pembahasan yang akan dicantumkan dalam karya tulis ilmiah ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab ini berisi tentang latar belakang penelitian, identifikasi, rumusan, dan Batasan masalah, tujuan dan manfaat penelitian, metode dan jenis penelitian, Teknik pengumpulan data, metode pengembangan data, metode pengembangan sistem, dan sistematika pembahasan.

BAB II TINJAUAN PUSTAKA

Bab ini berisi tentang beberapa literatur yang menjadi dasar pemikiran dalam penelitian tersebut, diantaranya adalah penelitian terdahulu, landasan teori, pemodelan, dan perangkat lunak yang digunakan.

BAB III METODELOGI PENELITIAN

Bab ini membahas tentang bagaimana peneliti mendapatkan data yang digunakannya dalam penelitian ini dan bagaimana cara mendapatkannya berikut memprosesnya sebelum melakukan Analisa lebih lanjut menggunakan algoritma yang terdapat pada Data Mining.

BAB IV HASIL DAN PEMBAHASAN

Bab ini adalah hasil dari Analisa yang telah diputuskan sebelumnya. Berisi tentang beberapa penjelasan maupun perhitungan dari data yang telah diolah sehingga menghasilkan sebuah prediksi berdasarkan kemungkinan yang telah didapatkan sebelumnya dan juga tingkat akurasi dari prediksi yang telah dihasilkan.

BAB V PENUTUP

Bab terakhir ini berisi tentang kesimpulan maupun saran yang diberikan dari sistem yang telah diteliti, dirancang, dan diimplementasikan.

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Penelitian ini meninjau beberapa penelitian sebelumnya meliputi:

a. PENERAPAN DATA MINING UNTUK MEMPREDIKSI HARGA BAHAN PANGAN DI INDONESIA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR

Penelitian ini dilakukan oleh Rahmadini Enjel Erika LorencisLubisb , Aji Priansyahc , Yolanda R.W.N.d , dan Tuti Meutia, yang di publish di JURNAL MAHASISWA AKUNTANSI SAMUDRA (JMAS) Volume 4 Nomor 4, pada bulan Agustus 2023 halaman 223 – 235.

Permasalahan yang dibahas dalam penelitian ini adalah bagaimana penerapan algoritma regresi K-Nearest Neighbor (KNN) untuk memprediksi harga produk beras di Indonesia. Batasan masalah penelitian ini adalah . Data yang digunakan merupakan dataset yang diperoleh dari Algoritma K-Regresi - Proses -Nearest Neighbor (KNN) dari website PIHPS Nasional periode Januari 2019 s/d Desember 2021. dan menghitung menggunakan RMSE (root mean square error) dan MAE (mean absolute error). Tujuannya adalah untuk membuat prediksi.

Algoritma regresi diproses dan dibandingkan dengan data tetangga terdekat K sehingga diperoleh nilai RMSE (root mean square error) yang kecil.

Semakin rendah nilai RMSE maka semakin tinggi akurasi. Tujuan dari penelitian ini adalah mengimplementasikan algoritma regresi K-Nearest Neighbor (KNN) untuk memprediksi harga bahan pangan dan menganalisis algoritma regresi K-Nearest Neighbor dengan cara lain untuk memprediksi harga bahan pangan.[8]

b. Penerapan Data Mining Terhadap Klasifikasi Pasien Penderita

Penyakit Liver Menggunakan Metode K-Nearest Neighbor

Penelitian ini dilakukan oleh mahasiswa Teknik Informatika, Universitas Stikubank (Unisbank) Semarang. Mahasiswa tersebut adalah Sabilla Aldana, dan Jati Sasongko Wibowo, yang publish di Jurnal Ilmiah Komputer Vol. 20, No. 1, pada bulan Februari 2024.

Metode klasifikasi dapat melakukan pembelajaran dengan memetakan suatu item kedalam kelas berdasarkan kelas data yang telah didefinisikan sebelumnya. Untuk memecahkan masalah tersebut, pada penelitian ini menggunakan metode yang digunakan untuk membangun model klasifikasi dalam penyakit liver yaitu algoritma K-Nearest Neighbor (KNN). K-Nearest Neighbor pada umumnya sudah banyak digunakan peneliti karena mempunyai kelebihan, yaitu nilai akurasi yang tinggi dan tidak ada asumsi pada data. Setelah melakukan 4 kali percobaan menggunakan metode K-Nearest Neighbor berdasarkan nilai K yaitu 3,5,7, dan 9 mendapatkan hasil keakuratan data yang terbaik adalah pada nilai K=7 yaitu nilai akurasi sebesar 70%, presisi 66,5%, recall 59,5%, dan f1-score 59,5%. Jadi algoritma K-Nearest Neighbor dinilai

cukup akurat untuk mengklasifikasi data pasien penderita penyakit liver karena mendapat nilai akurasi diatas 50%. Hasil pengujian keakuratan data tersebut menunjukkan bahwa kinerja pada nilai ketetanggaan atau $K=7$ lebih tinggi dari nilai K yang lainnya, dengan demikian dapat dikatakan bahwa kinerja algoritma K -Nearest Neighbor pada nilai $K=7$ adalah yang paling baik dari nilai K yang lainnya. Tujuan penelitian ini adalah menerapkan data mining terhadap klasifikasi pasien penderita penyakit liver menggunakan metode K -Nearest Neighbor. Kemudian Manfaat yang diharapkan dapat diperoleh dari penelitian ini adalah mengetahui kinerja terbaik berdasarkan performa tingkat akurasi dari metode K -Nearest Neighbor dalam melakukan klasifikasi pasien penderita penyakit liver.[9]

c. **Deteksi Kesehatan Janin Menggunakan Decision Tree dan Feature**

Forward Selection

Penelitian ini dihimpun dan dilakukan oleh Indah Sulihati, Abdul Syukur, Aris Marjuni yang merupakan mahasiswa ilmu komputer, teknik informatika, universitas Dian Nuswantoro, Semarang. Penelitian ini dipublikasikan pada jurnal Building of Informatics, Technology and Science (BITS), pada tanggal 30 Desember 2022.

Yang dibahas dalam penelitian ini adalah deteksi kesehatan janin yang menggunakan dataset public kesehatan janin dan menggunakan algoritma klasifikasi decision tree dengan feature forward selection untuk mendeteksi kesehatan janin, peneliti menggunakan data sebanyak 2.126 data record yang

diperoleh dari dataset public, maka dalam penelitian ini dapat di analisa dari hasil penelitian bahwa dengan menggunakan 2.126 data dengan penambahan feature forward selection dalam deteksi kesehatan janin ini ternyata dapat meningkatkan hasil class recall, class precision dan class accuracy. Sehingga dalam penelitian ini dapat disimpulkan dari hasil accuracy yang sebesar 91.06% setelah adanya penambahan feature forward selection dalam deteksi kesehatan janin dibandingkan prediksi yang hanya menggunakan algoritma decision tree saja yang hasilnya 89.84% dan ternyata lebih rendah hasilnya. Pengujian ini menggunakan model klasifikasi rekam kesehatan janin yang dilakukan dengan 2 tahap yaitu tahap pelatihan dan pengujian. Data pelatihan atau training merupakan tahapan untuk melatih suatu algoritma sehingga dapat membentuk model prediktif, dan data testing merupakan tahapan untuk menentukan hasil performansi suatu algoritma klasifikasi. Pengujian sebesar 30% dan 70 % dari dataset. Tujuan penelitian ini adalah untuk mendeteksi kesehatan janin menggunakan algoritma klasifikasi decision tree dengan forward feature selection. Dan bereksperimen dengan data public tersebut.[4]

2.2 Landasan Teori

Landasan teori ini adalah beberapa referensi atau pengertian dari beberapa kata kunci yang akan dibahas dalam penelitian ini. Berikut adalah beberapa pengertian tersebut.

2.2.1 Data Mining

Pengertian data mining adalah proses penggalian informasi untuk mengidentifikasi pola, tren, dan data yang dapat membantu pengambilan

keputusan bisnis. Proses penambangan data sering kali menggunakan metode statistik, matematika, dan bahkan teknologi kecerdasan buatan. Data mining dapat digambarkan sebagai proses yang digunakan oleh perusahaan untuk mengekstrak data dalam jumlah besar menjadi data tertentu. Hal ini melibatkan transformasi data mentah menjadi informasi berguna yang dapat memecahkan masalah bisnis. Penambangan data adalah bagian penting dari analisis data dan termasuk dalam bidang ilmu data. Dalam teknik yang lebih maju, data mining melibatkan proses penemuan pengetahuan dalam database, atau dikenal juga dengan Knowledge Discovery in Database (KDD), yang merupakan metodologi ilmu data untuk mengumpulkan, memproses, dan menganalisis data.

Menurut Gartner Group, data mining adalah penggunaan teknologi pengenalan pola, seperti metode statistik dan matematika, untuk mengklasifikasikan sejumlah besar data yang disimpan pada media penyimpanan untuk menemukan hubungan baru dengan makna, pola, dan kebiasaan menemukan. Data mining adalah kombinasi berbagai disiplin ilmu yang menggabungkan teknik seperti pembelajaran mesin, pengenalan pola, statistik, database, dan visualisasi untuk mengatasi masalah pengambilan informasi dari database besar.[10]

2.2.2 Algoritma K-Nearest Neighbor

K-Nearest Neighbors (knn) merupakan algoritma yang mengklasifikasikan data berdasarkan data pembelajaran (training data set) yang diambil dari k-tetangga terdekat (nearest neighbours). dimana k adalah banyaknya tetangga terdekat. Metode K-Nearest Neighbors melakukan

klasifikasi dengan memproyeksikan data latih ke dalam ruang multidimensi. Area ini dibagi menjadi beberapa bagian yang mewakili data dasar pelatihan. Semua data pelatihan direpresentasikan sebagai titik c dalam ruang multidimensi.

K-Nearest Neighbors (KNN) merupakan algoritma klasifikasi yang menggunakan himpunan nilai K dari data terdekat (tetangganya) sebagai acuan untuk menentukan kelas data baru. KNN mengklasifikasikan data berdasarkan kemiripan atau kedekatannya dengan data lain. Algoritma KNN ini adalah pembelajaran yang malas. Artinya, tidak menggunakan titik data pelatihan untuk membangun model. Dengan kata lain algoritma KNN mempunyai fase pelatihan yang sangat minim. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel pada data pelatihan.

2.2.3 Dataset

Data mining tidak bisa dipisahkan dari data set karena proses data mining sebenarnya membutuhkan data set sebagai objek ekstraksi pengetahuan. Dalam terminologi statistik, kumpulan data adalah kumpulan objek dengan atribut atau variabel tertentu, dan setiap objek merupakan bagian data dengan sekumpulan atribut atau variabel. Nama lain yang umum digunakan untuk objek termasuk catatan, titik, vektor, pola, peristiwa, observasi, dan kasus. Sebaliknya, baris yang mewakili objek data atau kolom disebut atribut. Atribut terkadang disebut variabel, bidang, fitur, atau dimensi.

Karakteristik umum kumpulan data yang dapat memengaruhi proses penambahan data adalah dimensi, ketersebaran, dan resolusi. Saat ini, catatan diklasifikasikan menjadi tiga jenis tergantung pada jenisnya: Yang pertama adalah record yang berupa record data. Artinya, data adalah kumpulan kumpulan data yang masing-masing kumpulan atributnya tetap. Yang kedua adalah data grafik, Dengan kata lain merupakan data yang berbentuk grafik yang terdiri dari node dan edge. Misalnya link HTML (di WWW), struktur molekul, dll. Yang terakhir adalah data terurut, yaitu data yang berisi sekumpulan nilai, Atau menampilkan data yang diurutkan menurut pola tertentu. Contohnya termasuk data genomic sequence atau spatio-temporal.[11]

2.2.4 Atribut

Atribut adalah simbol yang menggambarkan identitas atau sifat suatu objek. Atribut yang menggambarkan suatu objek rawat inap antara lain nama, umur, golongan darah, dan tekanan darah. Atribut diklasifikasikan menjadi empat jenis: Atribut nominal, yaitu atribut yang diperoleh melalui klasifikasi untuk menggambarkan kategori, kode, atau status yang tidak berurutan. Yang kedua adalah atribut ordinal, yaitu atribut yang menggambarkan urutan atau peringkat. Namun besarnya selisih antara dua nilai yang berurutan tidak diketahui. Atribut spasi (jarak) merupakan atribut numerik yang ditentukan oleh pengukuran. Oleh karena itu, jarak antara dua titik pada skala diketahui dan tidak ada nol mutlak. Yang terakhir adalah atribut rasio (mutlak), yaitu atribut

numerik yang mempunyai titik nol mutlak. Artinya, Anda dapat menghitung perkalian atau perbandingan antara satu nilai dengan nilai lainnya.[10]

2.3 Perangkat Lunak yang Digunakan

a. RapidMiner Studio

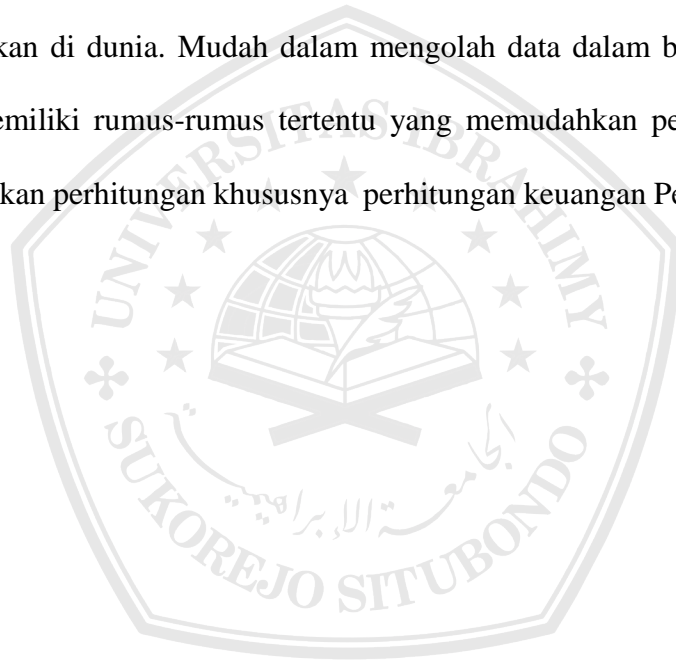
RapidMiner pertama kali dikembangkan dengan nama YALE (Yet Another Learning Environment) pada tahun 2001 oleh Ralf Klinkenberg, Ingo Mierswa, dan Simon Fischer di Artificial Intelligence Unit di Universitas Dortmund. RapidMiner sendiri merupakan software open source, sehingga setiap pengguna dapat mengembangkannya tanpa harus membayar royalti kepada pembuatnya. RapidMiner mencakup sekitar 500 operator prapemrosesan dan visualisasi data. RapidMiner ditulis dalam bahasa pemrograman Java, sehingga dapat digunakan di semua sistem operasi. RapidMiner menyediakan GUI (Graphic User Interface) untuk merancang jalur pipa analisis.

GUI ini menghasilkan file Extensible Markup Language (XML) yang mendefinisikan proses analisis yang Anda terapkan pada data Anda. File tersebut kemudian dibaca oleh RapidMiner dan dianalisis secara otomatis. RapidMiner menduduki peringkat pertama perangkat lunak data mining dalam studi yang dilakukan oleh portal data mining KDnuggets pada tahun 2010-2011.[12]

b. Microsoft Excel

Microsoft Excel adalah bagian dari rangkaian produk perangkat lunak Microsoft Office, bersama dengan Microsoft Word dan Microsoft

PowerPoint. Perkembangan terbaru pada Microsoft Office adalah Microsoft Office 365. Saat data mining diciptakan, aplikasi pertama yang digunakan untuk mencatat data adalah Microsoft Excel. Peneliti menggunakan Microsoft Excel 2019 karena Microsoft Excel memiliki kemampuan pemrosesan angka berbasis tabel yang menyertakan baris dan kolom untuk membuat data yang Anda perlukan. Microsoft Excel merupakan salah satu aplikasi pembuatan data yang paling banyak digunakan di dunia. Mudah dalam mengolah data dalam bentuk numerik dan memiliki rumus-rumus tertentu yang memudahkan pengguna dalam melakukan perhitungan khususnya perhitungan keuangan Perusahaan. [13]



BAB III

METODOLOGI PENELITIAN

3.1 Metode Penelitian

Pada dasarnya metode penelitian adalah cara sistematis mengumpulkan data untuk tujuan tertentu. Metode penelitian adalah suatu usaha atau prosedur untuk mencari pemecahan suatu topik atau permasalahan secara cermat, terencana, sistematis atau ilmiah dengan tujuan untuk menemukan fakta, asas, asas atau prosedur. Untuk menentukan dan menguji kebenaran ilmiah suatu informasi, proses ilmiah berupaya mengungkap fakta atau prinsip yang dimaksud.

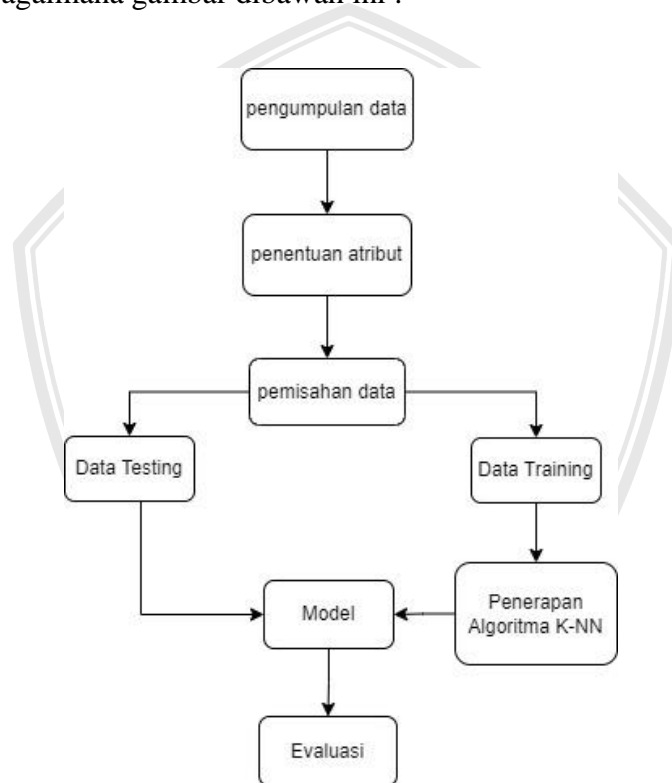
Fase ini penting untuk dimanfaatkan, karena keberhasilan atau kegagalan suatu proyek penelitian bergantung pada pilihan metode penelitian. Jika peneliti ingin memecahkan misteri dan mengungkap kebenaran melalui penelitian ilmiah, mereka harus menerapkan metodologi penelitian secara akurat. Dataset yang digunakan dalam penelitian kuantitatif ini dikumpulkan, diolah, dan dianalisis untuk mengetahui hubungan antar variabel yang diteliti. Penelitian ini menggunakan data 2.126 record dengan menggunakan pendekatan kuantitatif dan tinjauan literatur online. Data yang digunakan diperoleh dari aplikasi Kaggle. Pendekatan klasifikasi menggunakan algoritma K-Nearest Neighbors digunakan untuk mengolah data ini.

3.2 Tahapan Penelitian

Bagian ini menjelaskan proses bisnis sebelum dilakukannya pemodelan data. Tahap persiapan data (pengumpulan data) meliputi kegiatan seperti pemilihan

data, transformasi data, pembersihan data, dan validasi data. Tugas persiapan data dapat dijalankan beberapa kali dalam urutan apa pun. Hal ini mencakup pemilihan data yang relevan, pembersihan data, membuang data yang tidak terpakai atau tidak sesuai, dan membuat data yang sesuai untuk memfasilitasi proses data mining selanjutnya.

Dalam melakukan penelitian ini dilakukan beberapa tahapan atau langkah penelitian. Sebagaimana gambar dibawah ini :



Gambar 3. 1
Tahapan Penelitian

3.2.1 Pengumpulan Data

Data diperoleh dari repositori online aplikasi Kaggle yaitu *Classify The Health Of A Fetus As Normal, Suspect Or Pathological Using Ctg Data*.

Di bawah ini adalah tautan yang dapat Anda akses [Fetal Health](#)

Classification (kaggle.com) Kumpulan dataset ini diambil dari pemeriksaan CTG (*Cardiotogram*) yang kemudian diklasifikasikan oleh beberapa ahli obstetric ke dalam tiga kelas diantaranya: Normal, Suspect, Phatological.

3.2.2 Penentuan Atribut

Dalam penelitian ini, data kesehatan janin yang disebutkan di atas berisi 2.126 catatan *record*. Dan terdapat 22 atribut didalam dataset ini ,atribut '*fetal_health*' adalah label dari seluruh atribut yang ada. Penentuan atribut pada dataset "*fetal health*" melibatkan memilih berbagai fitur yang dapat digunakan untuk menganalisis kesehatan janin berdasarkan data medis yang dikumpulkan. Dataset ini berisi berbagai atribut yang berkaitan dengan rekam medis dan kondisi janin selama kehamilan. Berikut adalah beberapa atribut umum dan format tipe data yang ditemukan dalam dataset fetal health yang digunakan untuk menjalankan pengklasifikasian dengan mudah.

Atribut yang digunakan dalam memrprediksi kesehatan janin adalah sebagai berikut:

a. *Baseline_Value*

Baseline Value adalah tingkat detak jantung dasar janin dalam kondisi normal. Ini adalah indikator penting dari kesehatan janin.

No	Nilai Nominal	Jumlah	No	Nilai Nominal	Jumlah
1	133.0	117	25	139.0	28
2	122.0	86	26	126.0	27

Tabel Lanjutan

3	130.0	86	27	145.0	26
4	138.0	80	28	115.0	22
5	128.0	72	29	150.0	21
6	125.0	69	30	110.0	17
7	132.0	65	31	112.0	15
8	136.0	61	32	152.0	15
9	120.0	60	33	147.0	13
10	144.0	59	34	149.0	13
11	142.0	58	35	151.0	12
12	134.0	55	36	119.0	10
13	140.0	55	37	159.0	10
14	146.0	52	38	124.0	9
15	127.0	51	39	118.0	8
16	135.0	48	40	158.0	8
17	137.0	48	41	106.0	7
18	143.0	48	42	114.0	7
19	148.0	47	43	154.0	6
20	131.0	45	44	157.0	4
21	123.0	44	45	116.0	3
22	129.0	43	46	156.0	3
23	121.0	34	47	117.0	2
24	141.0	31	48	160.0	1

Tabel 3. 1

Atribut detak jantung dasar janin*b. Accelerations*

Accelerations adalah Peningkatan sementara dalam detak jantung janin yang bisa menunjukkan kesehatan yang baik.

No	Nilai Nominal	Jumlah	No	Nilai Nominal	Jumlah
1	0.0	725	11	00.01	36
2	0.002	131	12	0.011	29
3	0.003	127	13	0.012	21
4	0.001	113	14	0.013	20
5	0.004	88	15	0.014	16
6	0.005	88	16	0.015	9
7	0.006	88	17	0.016	3
8	0.008	84	18	0.017	3
9	0.007	71	19	0.018	2
10	0.009	46	20	0.019	1

Tabel 3. 2

Atribut Accelerations

c. *Fetal_movement*

Fetal movement adalah Gerakan janin yang direkam. Aktivitas janin yang baik umumnya merupakan tanda positif.

No	Nilai Nominal	Jumlah	No	Nilai Nominal	Jumlah
1	00.00	1063	29	0.028	3
2	0.001	122	30	0.043	3
3	0.002	77	31	0.048	3
4	0.003	70	32	00.05	3
5	0.004	41	33	0.054	3
6	0.005	31	34	0.085	3
7	0.006	26	35	0.027	2
8	0.007	24	36	00.03	2
9	0.009	24	37	00.04	2
10	00.01	20	38	0.049	2
11	0.008	18	39	0.052	2
12	0.013	13	40	0.058	2
13	0.011	10	41	00.06	2
14	0.012	10	42	0,283333	2
15	0.016	9	43	0.023	1
16	0.017	9	44	0.031	1
17	0.019	8	45	0.032	1
18	0.014	6	46	0.033	1
19	0.015	6	47	0.036	1
20	00.02	4	48	0.037	1
21	0.021	4	49	0.038	1
22	0.022	4	50	0.041	1
23	0.025	4	51	0.051	1
24	0.029	4	52	0.055	1
25	0.035	4	53	0.063	1
26	0.018	3	54	0.065	1
27	0.024	3	55	0.072	1
28	0.026	3	56	0.079	1
			57	0.088 – 0.33	34

Tabel 3. 3
Atribut *fetal_movement*

d. *Uterine_Contraction*

Atribut ini menjelaskan jumlah kontraksi rahim yang direkam selama pemeriksaan. Ini bisa mempengaruhi kesehatan janin dan proses persalinan.

No	Nilai Nominal	Jumlah	No	Nilai Nominal	Jumlah
1	00.00	261	9	0.001	87
2	0.005	240	10	0.009	66
3	0.004	201	11	00.01	41
4	0.006	184	12	0.011	14
5	0.007	176	13	0.012	9
6	0.003	167	14	0.013	2
7	0.002	128	15	0.014	1
8	0.008	123	16	0.015	1

Tabel 3. 4
Atribut Kontraksi Rahim

e. *Light_decelaration*

Atribut ini menjelaskan penurunan sementara dalam detak jantung janin yang biasanya tidak berbahaya.

No	Nilai Nominal	Jumlah	No	Nilai Nominal	Jumlah
1	00.00	992	9	0.008	41
2	0.001	128	10	0.009	33
3	0.004	98	11	00.01	11
4	0.003	94	12	0.011	10
5	0.002	89	13	0.012	7
6	0.005	82	14	0.013	7
7	0.006	56	15	0.014	6
8	0.007	46			

Tabel 3. 5
Penurunan sementara detak jantung janin

f. *Severe_deceleration*

Atribut ini menjelaskan tentang penurunan signifikan dalam detak jantung janin yang bisa menjadi tanda masalah.

No	Nilai Nominal	Jumlah
1	0.0	1694
2	0.001	7

Tabel 3.6

Atribut *severe deceleration*g. *Prolongued deceleration*

Atribut ini menjelaskan penurunan dalam detak jantung janin yang berlangsung lebih lama dari normal dan bisa menunjukkan stres janin.

No	Nilai Nominal	Jumlah
1	0.0	1559
2	0.001	60
3	0.002	55
4	0.003	18
5	0.004	7
6	0.005	2

Tabel 3.7

Atribut *Prolongued deceleration*h. *Abnormal short term variability*

Atribut ini menjelaskan variabilitas jangka pendek dalam detak jantung janin yang abnormal. Kurangnya variabilitas bisa menunjukkan masalah pada sistem saraf janin.

No	Nilai Nominal	Jumlah	No	Nilai Nominal	Jumlah
1	65.0	55	39	67.0	23
2	58.0	49	40	52.0	22
3	60.0	49	41	20.0	21
4	61.0	49	42	27.00	21
5	63.0	48	43	43.0	20
6	64.0	48	44	45.0	20
7	62.0	40	45	19.0	18
8	51.0	39	46	46.0	18
9	22.0	35	47	48.0	18
10	24.0	34	48	47.0	17
11	25.0	34	49	49.0	17

Tabel Lanjutan

12	28.0	34	50	70.0	17
13	32.0	34	51	78.0	16
14	36.0	34	52	68.0	15
15	57.0	34	53	73.0	14
16	41.0	33	54	74.0	14
17	50.0	33	55	77.0	14
18	56.0	32	56	75.0	13
19	37.0	31	57	79.0	13
20	26.0	30	58	16.0	11
21	38.0	30	59	69.0	11
22	59.0	30	60	17.0	10
23	23.0	29	61	71.0	9
24	29.0	29	62	72.0	9
25	34.0	29	63	76.0	9
26	55.0	29	64	18.0	8
27	35.0	28	65	13.0	7
28	39.0	28	66	14.0	4
29	42.0	28	67	80.0	4
30	40.0	27	68	81.0	4
31	44.0	27	69	84.0	4
32	53.0	27	70	86.0	4
33	21.0	26	71	15.0	3
34	33.0	26	72	83.0	3
35	66.0	26	73	12.0	2
36	54.0	25	74	82.0	2
37	30.0	23	75	87.0	1
38	31.0	23			

Tabel 3. 8

Atribut *Abnormal_short_therm_variability**i. Mean_value_of_short_therm_variability*

Atribut ini menjelaskan Rata-rata nilai variabilitas jangka pendek dalam detak jantung janin.

No	Nilai Nominal	Jumlah
1	01.03	98
2	00.09	97
3	00.04	96
4	00.06	94

No	Nilai Nominal	Jumlah
24	02.04	21
25	02.05	21
26	02.08	19
27	02.06	16

5	00.07	94	28	03.00	13
---	-------	----	----	-------	----

Tabel Lanjutan

6	00.08	93	29	03.02	11
7	00.05	91	30	03.04	11
8	01.02	87	31	02.09	8
9	01.00	84	32	03.01	6
10	01.05	80	33	03.08	6
11	01.01	76	34	03.03	5
12	01.04	76	35	03.06	4
13	00.03	69	36	03.05	3
14	01.06	61	37	04.02	3
15	01.07	56	38	04.09	3
16	01.09	48	39	03.07	2
17	00.02	41	40	04.00	2
18	02.01	38	41	04.03	2
19	01.08	37	42	04.05	2
20	02.02	35	43	05.00	2
21	02.00	29	44	05.04	2
22	02.03	26	45	06.03	2
23	02.07	23	46	3.9 – 6.9	7

Tabel 3. 9

Atribut Mean_value_of_short_therm_variability

j. Percentage_of_Time_with_Abnormal_Long_Term_Variability

Atribut ini menjelaskan persentase waktu di mana variabilitas jangka panjang dalam detak jantung janin tidak normal.

No	Nilai Nominal	Jumlah	No	Nilai Nominal	Jumlah
1	0.0	382	39	35.0	2
2	1.0	17	40	40.0	2
3	2.0	17	41	44.0	2
4	3.0	12	42	47.0	2
5	4.0	11	43	49.0	2
6	5.0	9	44	52.0	2
7	6.0	9	45	54.0	2
8	12.0	8	46	58.0	2
9	13.0	8	47	61.0	2
10	32.0	8	48	84.0	2
11	9.0	8	49	18.0	1
12	11.0	7	50	23.0	1

13	31.0	6	51	26.0	1
14	8.0	6	52	36.0	1

Tabel Lanjutan

15	10.0	5	53	37.0	1
16	17.0	5	54	41.0	1
17	22.0	5	55	45.0	1
18	7.0	5	56	46.0	1
19	15.0	4	57	48.0	1
20	20.0	4	58	51.0	1
21	21.0	4	59	53.0	1
22	25.0	4	60	56.0	1
23	30.0	4	61	57.0	1
24	62.0	4	62	59.0	1
25	71.0	4	63	60.0	1
26	14.0	3	64	64.0	1
27	16.0	3	65	66.0	1
28	29.0	3	66	67.0	1
29	33.0	3	67	68.0	1
30	34.0	3	68	69.0	1
31	38.0	3	69	70.0	1
32	39.0	3	70	74.0	1
33	43.0	3	71	78.0	1
34	72.0	3	72	81.0	1
35	19.0	2	73	85.0	1
36	24.0	2	74	91.0	1
37	27.0	2	75	42.0 – 90.0	0
38	28.0	2			

Tabel 3. 10

Persentase detak jantung janin*k. Mean_value_of_long_variability*

Atribut ini menjelaskan Rata-rata nilai variabilitas jangka panjang dalam detak jantung janin.

No	Nilai Nominal	Jumlah	No	Nilai Nominal	Jumlah
1	0.0	37	11	1.5 – 9.9	96
2	5.2	13	12	0.1 – 9.0	81
3	6.7	10	13	0.2 – 8.9	78
4	7.4	10	14	0.4 - 9.7	54
5	9.5	10	15		

6	6.5 – 6.8	18			
7	4.8 – 8.7	64			
8	10.0 – 8.0	42			

Tabel Lanjutan

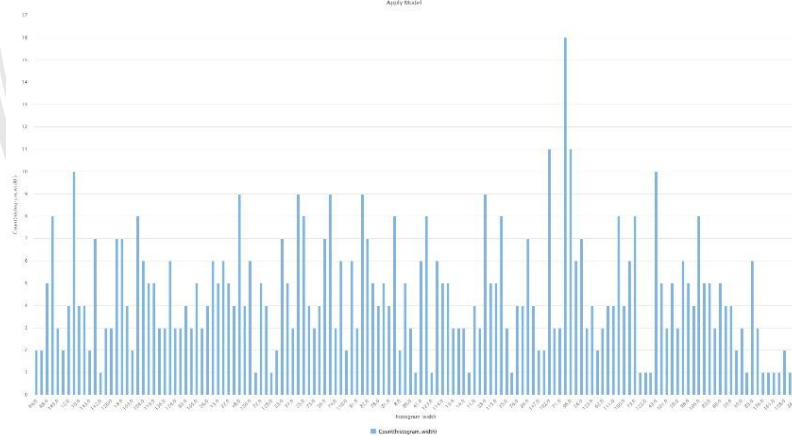
9	11.14 – 9.8	60			
10	10.2 – 8.4	65			

Tabel 3. 11

Atribut mean_value_of_long_variability

l. Histogram_Width

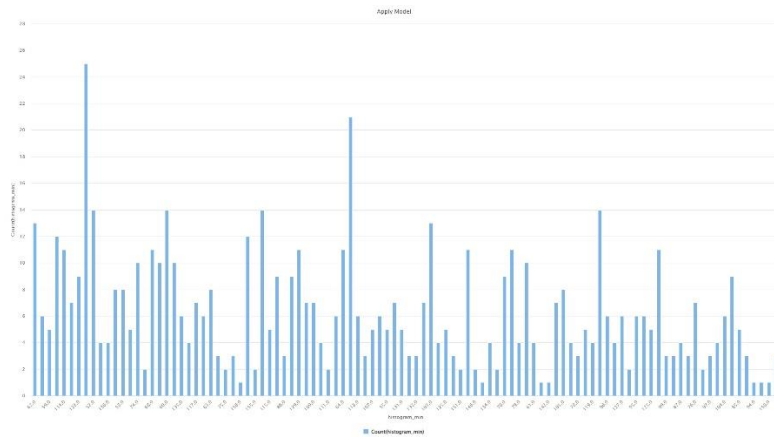
Atribut ini menjelaskan tentang lebar histogram yang menggambarkan distribusi detak jantung janin. Lebar yang besar bisa menunjukkan kesehatan yang baik. Hasil histogram bisa dilihat pada gambar 3.2



Gambar 3. 2
Histogram width

m. Histogram_min

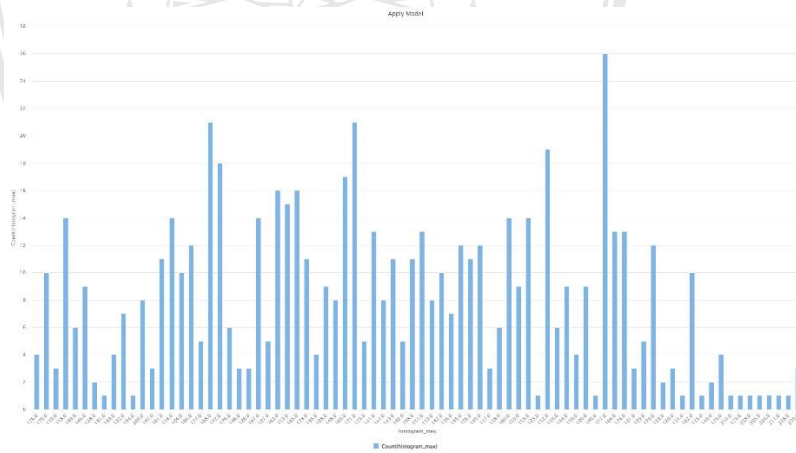
Atribut ini menjelaskan tentang nilai minimum dari histogram detak jantung janin. Hasil histogram bisa dilihat pada gambar 3.3



Gambar 3.3
Histogram min

n. *Histogram_max*

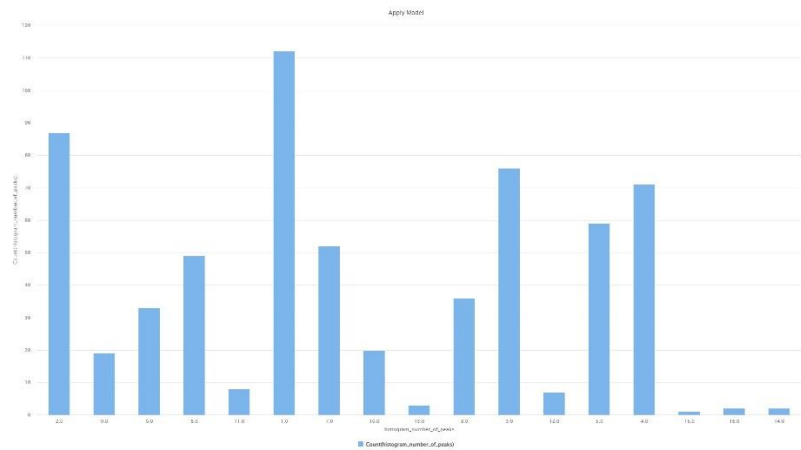
Atribut ini menjelaskan tentang nilai maksimum dari histogram detak jantung janin. Hasil histogram bisa dilihat pada gambar 3.4



Gambar 3.4
Histogram max

o. *Histogram number of peaks*

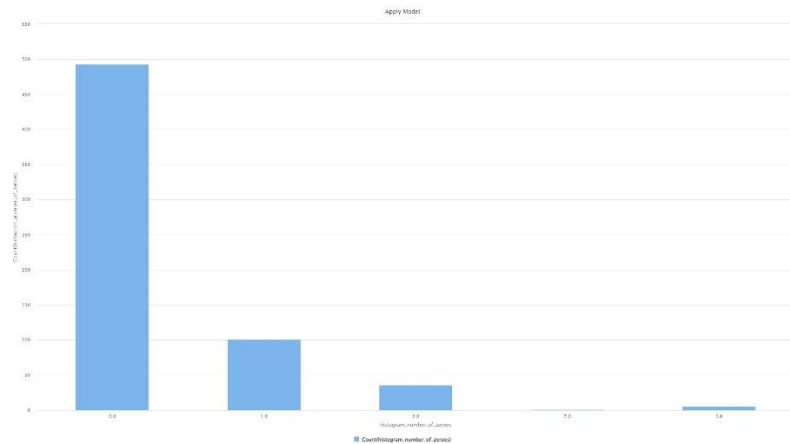
Atribut ini menjelaskan tentang jumlah puncak dalam histogram detak jantung janin. Hasil histogram bisa dilihat pada gambar 3.5



Gambar 3.5
Histogram number of peaks

p. Histogram number of zeroes

Atribut ini menjelaskan tentang jumlah nilai nol dalam histogram detak jantung janin. Hasil histogram bisa dilihat pada gambar 3.6



Gambar 3. 6
Histogram number of zeroes

q. Histogram mode

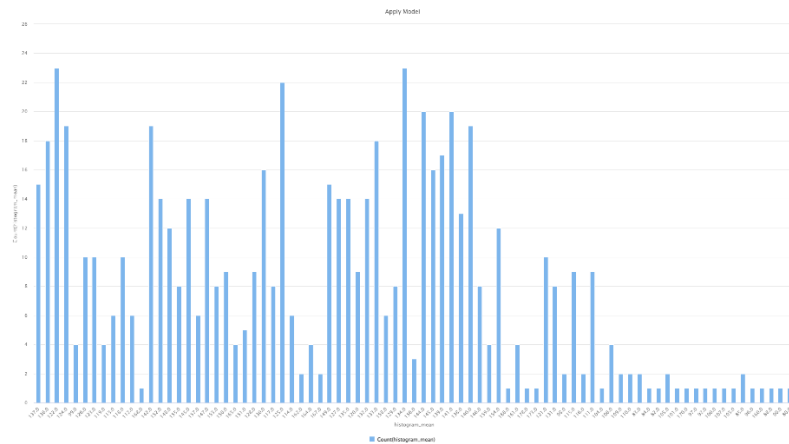
Atribut ini menjelaskan tentang modus dari histogram detak jantung janin. Hasil histogram bisa dilihat pada gambar 3.7



Gambar 3. 7
Histogram mode

r. Histogram mean

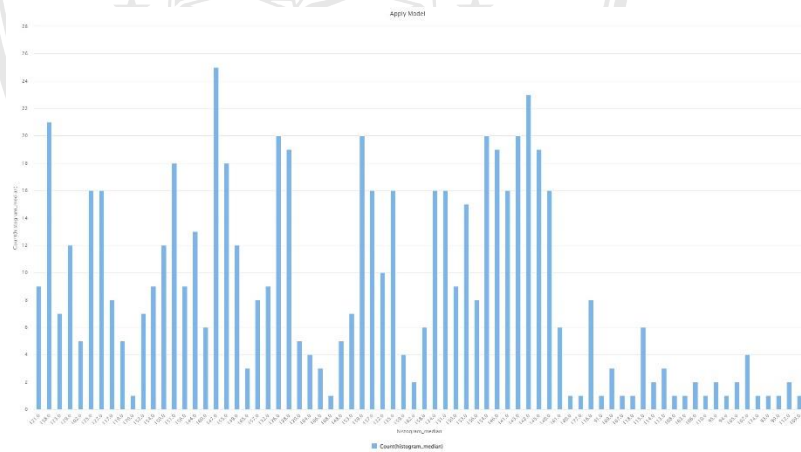
Atribut ini menjelaskan tentang nilai rata-rata dari histogram detak jantung janin. Hasil histogram bisa dilihat pada gambar 3.8



Gambar 3. 8
Histogram mean

s. *Histogram median*

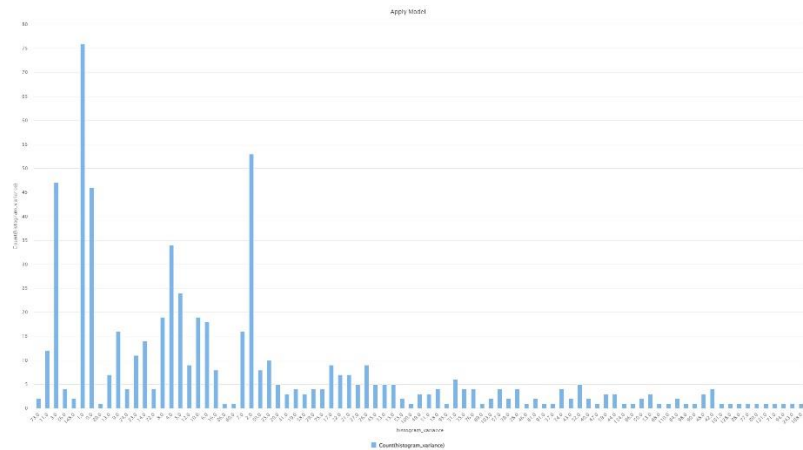
Atribut ini menjelaskan tentang nilai median dari histogram detak jantung janin. Hasil histogram bisa dilihat pada gambar 3.9



Gambar 3. 9
Histogram median

t. *Histogram variance*

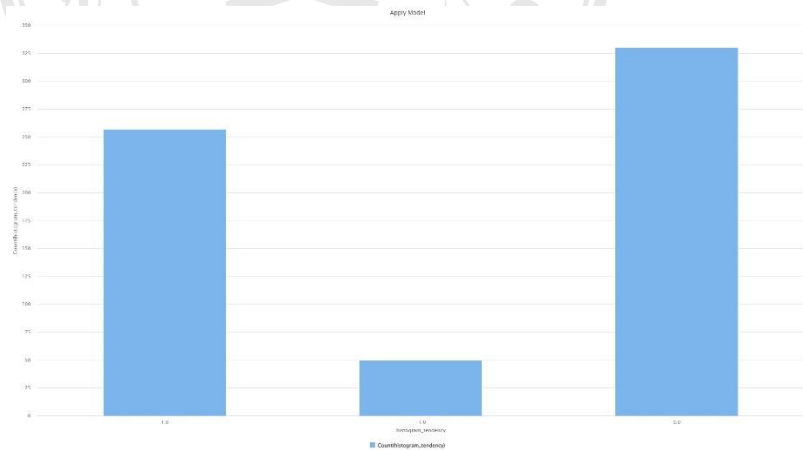
Atribut ini menjelaskan tentang variansi dari histogram detak jantung janin. Hasil histogram bisa dilihat pada gambar 3.10



Gambar 3. 10
Histogram variance

u. Histogram tendency

Atribut ini menjelaskan tentang kecenderungan dari histogram detak jantung janin (positif atau negatif). Hasil histogram bisa dilihat pada gambar 3.11



Gambar 3. 11
Histogram tendency

v. Class (fetal_health)

Dalam dataset yang digunakan di dalam penelitian ini terdapat 2.126 record yang terbagi menjadi tiga bagian. Data tersebut mempunyai 2.126 data record yang dikategorikan normal, suspect (kemungkinan), dan pathologic (kanker). Sebagaimana tabel 3.10

No	Nilai Nominal	Jumlah
1	1.0 (<i>Normal</i>)	1324
2	2.0 (<i>Suspect</i>)	236
3	3.0 (<i>Phatologic</i>)	141

Tabel 3. 12
Atribut fetal_health

3.2.3 Pemisahan Data

Pemisahan data menjadi training set dan testing set adalah langkah penting dalam proses pengembangan model machine learning. Ini dilakukan untuk memastikan bahwa model yang dibangun dapat menggeneralisasi dengan baik pada data yang belum pernah dilihat sebelumnya.

Adapun kegunaan dari data training adalah Subset data ini digunakan untuk melatih model. Model belajar pola, hubungan, dan karakteristik dari data ini. Selain data training juga ada kegunaan dari data testing yaitu Subset data ini digunakan untuk menguji model setelah pelatihan selesai. Data ini tidak pernah dilihat oleh model selama pelatihan dan digunakan untuk memberikan penilaian yang obyektif terhadap performa model[14].

3.3 Perancangan Sistem

3.3.1 Metode algoritma K-NN

Algoritma K-Nearest Neighbor (K-NN) adalah algoritma machine learning yang bersifat non-parametric dan lazy learning. Metode yang bersifat non-parametric memiliki makna bahwa metode tersebut tidak membuat asumsi apa pun tentang distribusi data yang mendasarinya. Dengan kata lain, tidak ada jumlah parameter atau estimasi parameter yang tetap dalam model, terlepas data tersebut berukuran kecil ataupun besar[15].

Berikut adalah alur perhitungan manual dengan algoritma K-NN.



Gambar 3. 12
Alur penghitungan algoritma K-NN

Apabila dataset sudah di input maka langkah pertama adalah Lakukan normalisasi data agar semua fitur berada dalam skala yang sama. Misalnya, gunakan normalisasi min-max. kemudian pilih nilai K misalnya K=3. Langkah ketiga adalah menghitung jarak Euclidean dengan rumus

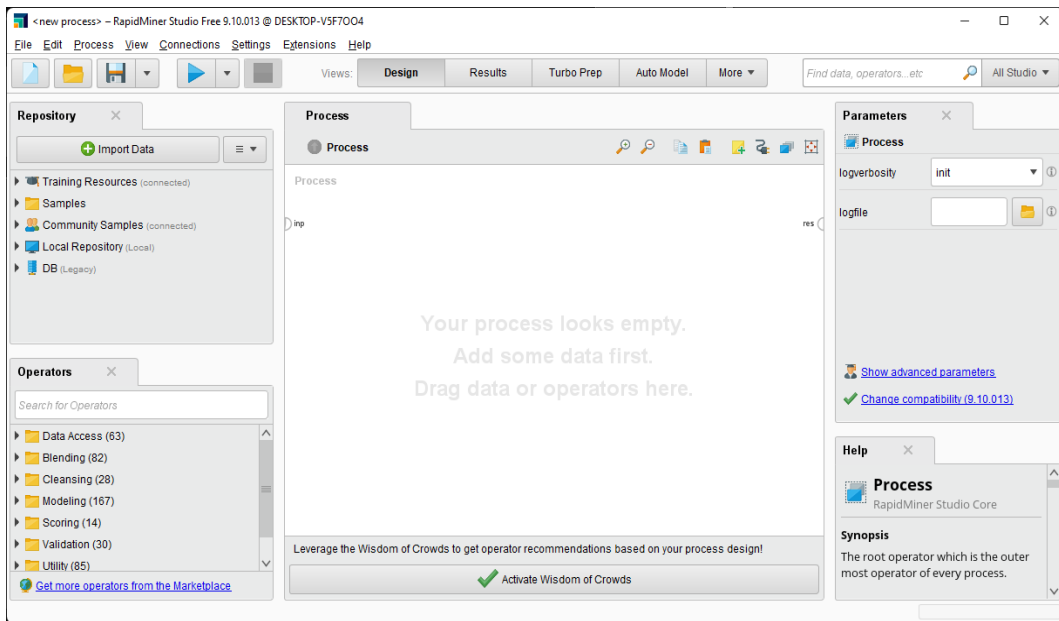
$$(A, B) = \sqrt{\sum_{i=1}^n (A_i - B_i)^2} .$$

Setelah menghitung jarak Euclidean langkah selanjutnya adalah menentukan K- Neighborsnya dengan cara mengurutkan dari nilai yang kecil sampai yang terbesar. Dari K neighbors terdekat tentukan label berdasarkan mayoritas dari K tetangga terdekat untuk mengevaluasi model dari algoritma K-NN

tersebut, setelah mengevaluasi mode langkah selanjutnya adalah menentukan kelas dari dataset, Dari seluruh perhitungan tersebut terbagi kepada perhitungan dengan *class* label *Normal*, *Suspect* dan *Phatologic* kemudian hasil yang terbesar dari perbandingan ketiga label tersebut merupakan hasil dari prediksi algoritma *K-Nearest Neighbors* tersebut.

3.4 Implementasi dan pengujian Metode

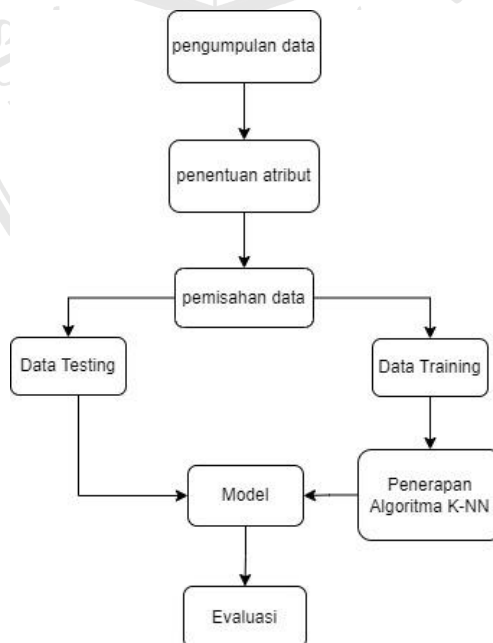
Penggunaan RapidMiner direkomendasikan untuk implementasi dan tinjauan aktuaria proyek ini. Perangkat lunak ini berguna untuk ekstraksi data menggunakan teknik data mining. Selain itu, RapidMiner dapat mengidentifikasi pola dari kumpulan data yang sangat besar dengan memanfaatkan teknik statistik, kecerdasan buatan, dan basis data. Model-model yang ditawarkan oleh RapidMiner adalah metode K – Nearest Neighbors, Naïve Bayes, Decision Tree, Neural Network, dan lainnya, yang berfokus pada proses Knowledge of Data Discovery. Berikut ini adalah *User Interface* dari aplikasi RapidMiner yang ditunjukkan di bawah ini.



Gambar 3. 13
User interface RapidMiner

Tahapan – tahapan yang akan dilakukan dalam implementasi *RapidMiner*

9.10 ini adalah sebagaimana berikut:



Gambar 3. 14
Kerangka pemikiran pada RapidMiner

Tahap pertama adalah import data dalam format csv yang telah disiapkan sebelumnya. Hal ini dilakukan dengan fitur *Read csv* dalam *RapidMiner 9.10*. Kemudian pelabelan dan penentuan tipe data dilakukan sebelum data siap untuk dihitung dengan algoritma *K-Nearest Neighbors* pada aplikasi *RapidMiner*. Kemudian pemilihan data *training* dan data *testing* dilakukan dengan menggunakan *tool split data* yang ada pada *RapidMiner*. Biasanya proporsi 70% untuk training dan 30% untuk testing. Kemudian Tambahkan operator cross-validation untuk memastikan model tidak overfitting dan memiliki kinerja yang baik pada data yang tidak terlihat. Kemudian tambahkan operator K-NN didalam operator cross-validation untuk melatih model KNN. Konfigurasi operator ini untuk memilih nilai K yang sesuai. pada penelitian ini penulis memakai K=5 karena ditemukan bahwa angka 5 memberikan performa terbaik berdasarkan akurasi validasi. Kemudian untuk mengevaluasi model perlu ditambahkan operator performance. Tujuan dari konfigurasi operator ini untuk menampilkan metric evaluasi seperti akurasi, precision, recall.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Penelitian

Hasil penelitian yang telah dilakukan dalam memprediksi kesehatan janin pada dataset *Fetal-Health* menggunakan metode *K-Nearest Neighbors*. Dalam penelitian ini diperoleh sebanyak 22 atribut dengan 2.126 record dari repositori online Kaggle.

4.2 Perhitungan Manual

4.2.1 Menentukan atribut dan label

Dataset yang telah disiapkan sebagaimana yang telah tertera pada gambar 4.1, data kesehatan janin yang disebutkan di bawah berisi 2.126 catatan *record*. Dan terdapat 22 atribut didalam dataset ini ,atribut '*fetal_health*'(fh) adalah label dari seluruh atribut yang ada

by	acc	fm	uc	ld	sd	pd	astv	mstv	Ptw	mvl	hw	hm	hax	hmp	hnz	hmd	han	hml	hvr	htd	fh
120	0	0	0	0	0	0	73	0,5	43	2,4	64	62	126	2	0	120	137	121	73	1	2
132	0,006	0	0,06	0,03	0	0	17	2,1	0	10,4	130	68	198	6	1	141	136	140	12	0	1
133	0,003	0	0,08	0,03	0	0	16	2,1	0	13,4	130	68	198	5	1	141	135	138	13	0	1
134	0,003	0	0,08	0,03	0	0	16	2,4	0	23	117	53	170	11	0	137	134	137	13	1	1
132	0,007	0	0,08	0	0	0	16	2,4	0	19,9	117	53	170	9	0	137	136	138	11	1	1
134	0,001	0	0,01	0,09	0	0,2	26	5,9	0	0	150	50	200	5	3	76	107	107	170	0	3
134	0,001	0	0,13	0,08	0	0,3	29	6,3	0	0	150	50	200	6	3	71	107	106	215	0	3

Gambar 4. 1

Atribut Dataset

fh
2
1
1
1
1
3
3

Gambar 4. 2

Label Dataset

bv	acc	fm	uc	ld	sd	pd	astv	mstv	ptw	mvl	hw	hm	hax	hnp	hnz	hmd	han	hmd	hvr	htd	fh
128	0,002	0	0,02	0,01	0	0	70	0,9	12	1,6	120	2	0	116	130	114	67	115	57	0	?

Gambar 4. 3

Data baru yang akan diprediksi

4.2.3 Menentukan nilai K terbaik

Menurut , pastikan algoritma *machine learning* yang ingin digunakan mendukung metric yang dipilih. Jika algoritma menggunakan metric berbeda. Hal tersebut dapat digunakan untuk memperoleh target untuk hasil *cross validation*. Untuk itu sebelum melakukan proses klasifikasi menggunakan algoritma KNN peneliti menggunakan parameter K=5, karena ditemukan bahwa angka 5 memberikan performa terbaik berdasarkan akurasi validasi.

4.2.4 Menghitung jarak antara data training dan data testing

$$d(X_1, X_2) = \sqrt{\sum_{i=1}^n (x_{1i} - x_{2i})^2}$$

Dalam menghitung jarak terdapat berbagai macam perhitungan salah satunya dengan Euclidian distance, setelah mengetahui $K=5$ maka dilakukanlah hitung jarak antara data baru dengan dataset.

a. Jarak ke data 1

$$\begin{aligned}
 d &= \sqrt{(128 - 120)^2 + (0,002 - 0)^2 + (0 - 0)^2 + (0,02 - 0)^2 + (0,01 - 0)^2 + (0 - 0)^2 + (0 - 0)^2 + (70 - 73)^2 + (0,9 - 0,5)^2 + (12 - 43)^2 + (1,6 - 2,4)^2 + (120 - 64)^2 + (2 - 62)^2 + (116 - 126)^2 + (0 - 2)^2 + (0 - 0)^2 + (114 - 120)^2 + (67 - 137)^2 + (115 - 121)^2 + (57 - 73)^2 + (0 - 1)^2} \\
 &= \sqrt{(8)^2 + (0,002)^2 + (0)^2 + (0,02)^2 + (0,01)^2 + (0)^2 + (0)^2 + (-3)^2 + (0,4)^2 + (-31)^2 + (-0,8)^2 + (56)^2 + (-60)^2 + (-10)^2 + (-2)^2 + (0)^2 + (-6)^2 + (-70)^2 + (-6)^2 + (-16)^2 + (1)^2} \\
 &= \sqrt{13.103,180104} = 114.469
 \end{aligned}$$

b. Jarak ke data 2

$$\begin{aligned}
 d &= \sqrt{(128 - 132)^2 + (0,002 - 0,006)^2 + (0 - 0)^2 + (0,02 - 0,06 + (0,01 - 0,03)^2 + (0 - 0)^2 + (0 - 0)^2 + (70 - 17)^2 + (0,9 - 2,1)^2 + (12 - 0)^2 + (1,6 - 10,4)^2 + (120 - 130)^2 + (2 - 68)^2 + (116 - 198)^2 + (0 - 6)^2 + (0 - 1)^2 + (114 - 141)^2 + (67 - 136)^2 + (115 - 140)^2 + (57 - 12)^2 + (0 - 0)^2} \\
 &= \sqrt{(-4)^2 + (-0,004)^2 + (0)^2 + (-0,04)^2 + (-0,02)^2 + (0)^2 + (0)^2 + (53)^2 + (-1,2)^2 + (12)^2 + (-8,8)^2 + (-10)^2 + (-66)^2 + (-82)^2 + (6)^2 + (1)^2 + (-27)^2 + (-69)^2 + (-35)^2 + (45)^2 + (0)^2} \\
 &= \sqrt{13.102,180} = 114,464
 \end{aligned}$$

c. Jarak ke data 3

$$d = \sqrt{\begin{aligned} &(128 - 133)^2 + (0,002 - 0,006)^2 + (0 - 0)^2 + (0,02 - 0,06 + \\ &(0,01 - 0,03)^2 + (0 - 0)^2 + (0 - 0)^2 + (70 - 17)^2 + (0,9 - 2,1)^2 + \\ &(12 - 0)^2 + (1,6 - 10,4)^2 + (120 - 130)^2 + (2 - 68)^2 + (116 - 198)^2 \\ &+(0 - 6)^2 + (0 - 1)^2 + (114 - 141)^2 + (67 - 136)^2 + (115 - 140)^2 \\ &+(57 - 12)^2 + (0 - 0)^2 \end{aligned}}$$

$$= \sqrt{\begin{aligned} &(-3)^2 + (-0,004)^2 + (0)^2 + (-0,04)^2 + (-0,02)^2 + (0)^2 + (0)^2 + \\ &(53)^2 + \\ &(-1,2)^2 + (12)^2 + (-8,8)^2 + (-10)^2 + (-66)^2 + (-82)^2 + (6)^2 + \\ &(1)^2 + (-27)^2 + (-69)^2 + (-35)^2 + (45)^2 + (0)^2 \end{aligned}}$$

$$= \sqrt{13.099,18} = 114,451$$

d. Jarak ke data 4

$$d = \sqrt{\begin{aligned} &(128 - 134)^2 + (0,002 - 0,006)^2 + (0 - 0)^2 + (0,02 - 0,06 + \\ &(0,01 - 0,03)^2 + (0 - 0)^2 + (0 - 0)^2 + (70 - 17)^2 + (0,9 - 2,1)^2 + \\ &(12 - 0)^2 + (1,6 - 10,4)^2 + (120 - 130)^2 + (2 - 68)^2 + (116 - 198)^2 \\ &+(0 - 6)^2 + (0 - 1)^2 + (114 - 141)^2 + (67 - 136)^2 + (115 - 140)^2 \\ &+(57 - 12)^2 + (0 - 0)^2 \end{aligned}}$$

$$= \sqrt{\begin{aligned} &(-6)^2 + (-0,004)^2 + (0)^2 + (-0,04)^2 + (-0,02)^2 + (0)^2 + (0)^2 + \\ &(53)^2 + \\ &(-1,2)^2 + (12)^2 + (-8,8)^2 + (-10)^2 + (-66)^2 + (-82)^2 + (6)^2 + \\ &(1)^2 + (-27)^2 + (-69)^2 + (-35)^2 + (45)^2 + (0)^2 \end{aligned}}$$

$$= \sqrt{13.093,18} = 114,425$$

e. Jarak ke data 5

$$\begin{aligned}
 d &= \sqrt{(128 - 132)^2 + (0,002 - 0,006)^2 + (0 - 0)^2 + (0,02 - 0,06 + \\
 &\quad (0,01 - 0,03)^2 + (0 - 0)^2 + (0 - 0)^2 + (70 - 17)^2 + (0,9 - 2,1)^2 + \\
 &\quad (12 - 0)^2 + (1,6 - 10,4)^2 + (120 - 130)^2 + (2 - 68)^2 + (116 - 198)^2 \\
 &\quad + (0 - 6)^2 + (0 - 1)^2 + (114 - 141)^2 + (67 - 136)^2 + (115 - 140)^2 \\
 &\quad + (57 - 12)^2 + (0 - 0)^2} \\
 &= \sqrt{(-4)^2 + (-0,004)^2 + (0)^2 + (-0,04)^2 + (-0,02)^2 + (0)^2 + (0)^2 + \\
 &\quad (53)^2 + \\
 &\quad (-1,2)^2 + (12)^2 + (-8,8)^2 + (-10)^2 + (-66)^2 + (-82)^2 + (6)^2 + \\
 &\quad (1)^2 + (-27)^2 + (-69)^2 + (-35)^2 + (45)^2 + (0)^2} \\
 &= \sqrt{13,094,18} = 114,321
 \end{aligned}$$

f. Jarak ke data 6

$$\begin{aligned}
 d &= \sqrt{(128 - 134)^2 + (0,002 - 0,006)^2 + (0 - 0)^2 + (0,02 - 0,06 + \\
 &\quad (0,01 - 0,03)^2 + (0 - 0)^2 + (0 - 0)^2 + (70 - 17)^2 + (0,9 - 2,1)^2 + \\
 &\quad (12 - 0)^2 + (1,6 - 10,4)^2 + (120 - 130)^2 + (2 - 68)^2 + (116 - 198)^2 \\
 &\quad + (0 - 6)^2 + (0 - 1)^2 + (114 - 141)^2 + (67 - 136)^2 + (115 - 140)^2 \\
 &\quad + (57 - 12)^2 + (0 - 0)^2} \\
 &= \sqrt{(-4)^2 + (-0,004)^2 + (0)^2 + (-0,04)^2 + (-0,02)^2 + (0)^2 + (0)^2 + \\
 &\quad (53)^2 + \\
 &\quad (-1,2)^2 + (12)^2 + (-8,8)^2 + (-10)^2 + (-66)^2 + (-82)^2 + (6)^2 + \\
 &\quad (1)^2 + (-27)^2 + (-69)^2 + (-35)^2 + (45)^2 + (0)^2} \\
 &= \sqrt{13,063,18} = 114,09
 \end{aligned}$$

g. Jarak ke data 7

$$\begin{aligned}
 d &= \sqrt{(128 - 134)^2 + (0,002 - 0,006)^2 + (0 - 0)^2 + (0,02 - 0,06 + \\
 &\quad (0,01 - 0,03)^2 + (0 - 0)^2 + (0 - 0)^2 + (70 - 17)^2 + (0,9 - 2,1)^2 + \\
 &\quad (12 - 0)^2 + (1,6 - 10,4)^2 + (120 - 130)^2 + (2 - 68)^2 + (116 - 198)^2 \\
 &\quad + (0 - 6)^2 + (0 - 1)^2 + (114 - 141)^2 + (67 - 136)^2 + (115 - 140)^2 \\
 &\quad + (57 - 12)^2 + (0 - 0)^2} \\
 &= \sqrt{(-4)^2 + (-0,004)^2 + (0)^2 + (-0,04)^2 + (-0,02)^2 + (0)^2 + (0)^2 + \\
 &\quad (53)^2 + \\
 &\quad (-1,2)^2 + (12)^2 + (-8,8)^2 + (-10)^2 + (-66)^2 + (-82)^2 + (6)^2 + \\
 &\quad (1)^2 + (-27)^2 + (-69)^2 + (-35)^2 + (45)^2 + (0)^2} \\
 &= \sqrt{13,063,18} = 114,09
 \end{aligned}$$

4.2.5 Identifikasi tetangga terdekat

Pada identifikasi kali ini adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut.

Data pembelajaran digambarkan ke ruang berdimensi banyak dengan tiap-tiap dimensi mewakili tiap ciri/fitur dari data. Klasifikasi data baru dilakukan dengan mencari label k tetangga terdekat. Label terbanyak yang muncul menjadi label data baru. Bila $k = 5$, data baru dilabeli dengan label tetangga terdekat. Jarak yang biasa dipakai adalah Jarak *Euclidean*.

Data Fetal health	Jarak	Label
Data 1	114,469	3
Data 2	114,465	3
Data 3	114,463	3

Tabel Lanjutan

Data 4	114,451	3
Data 5	114,425	3
Data 6	114,321	3
Data 7	114,09	3

Tabel 4. 1

Jarak Euclidean

4.2.6 Menghitung nilai Accuracy

Dari hasil yang dilakukan pada tahapan modeling akan diperoleh model yang selanjutnya model tersebut akan dievaluasi untuk mengukur kinerja dari algoritma KNN. Secara umum untuk menggambarkan kinerja dari klasifikasi adalah confusion matrix, matriks ini dibuat merupakan tabulasi silang sederhana dari kategori kelas yang diuji dan diprediksi dimana sel diagonal mewakili kelas menunjukkan kelas yang diprediksi benar.

a. Accuracy

Akurasi adalah skor dari matriks konfusi yang berisi jumlah suatu prediksi yang benar dari seluruh data. Hal ini karena tingkat nilai akurasi dapat ditentukan dari suatu hasil prediksi yang mendekati nilai yang terukur yang sebenarnya.

$$\text{Formulasi : } \frac{TP+TN}{TP+TN+FP+FN}$$

b. *Sensitivity*

Sentivitas adalah skor dari matriks konfusi yang mengukur dengan status orang sakit. Sensivitas adalah naama lain untuk mengikat menghitung kelas positif

$$\text{Formula : } \frac{TP}{TP+FN}$$

c. *Specificity*

$$\text{Formula : } \frac{TN}{TN+FP}$$

d. Presisi

$$\text{Formula : } \frac{TP}{TP+FP}$$

e. F1- score

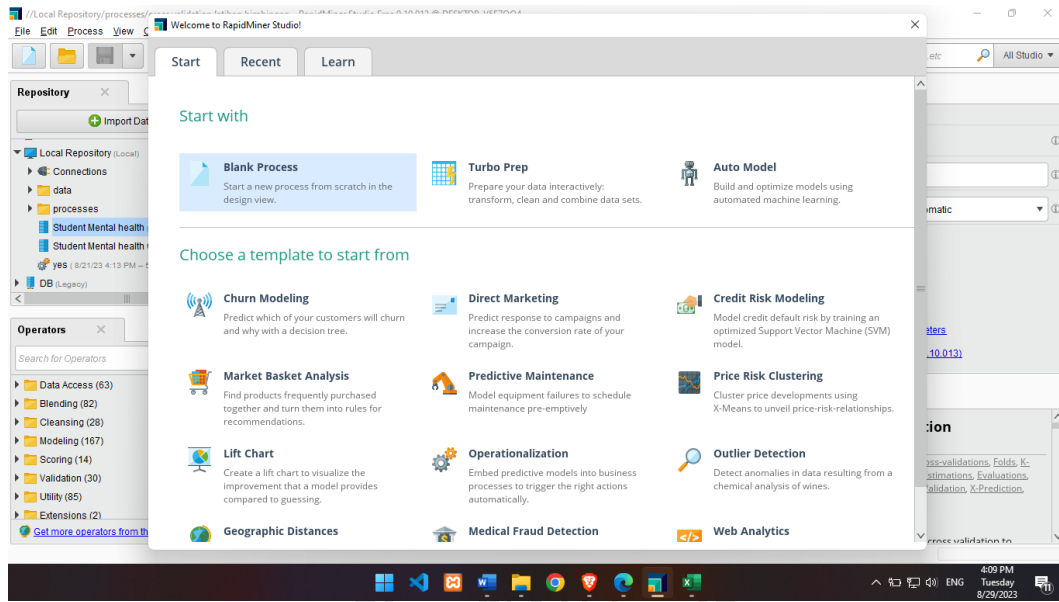
$$\text{Formula : } \frac{2TP}{2TP+FP+FN}$$

4.3 Perhitungan RapidMiner

Berikut ini adalah implementasi perhitungan prediksi Kesehatan janin menggunakan algoritma *K-Nearest Neighbors*. Berikut adalah Langkah – Langkah yang akan dilakukan dalam pengolahan data sampai hasil yang akan diperoleh dari prediksi tersebut.

4.3.1 Pemrosesan Data

Langkah awal adalah dengan membuka tampilan awal aplikasi *RapidMiner* dan memilih klik *File*, kemudian *new process*, dan klik *blank process*



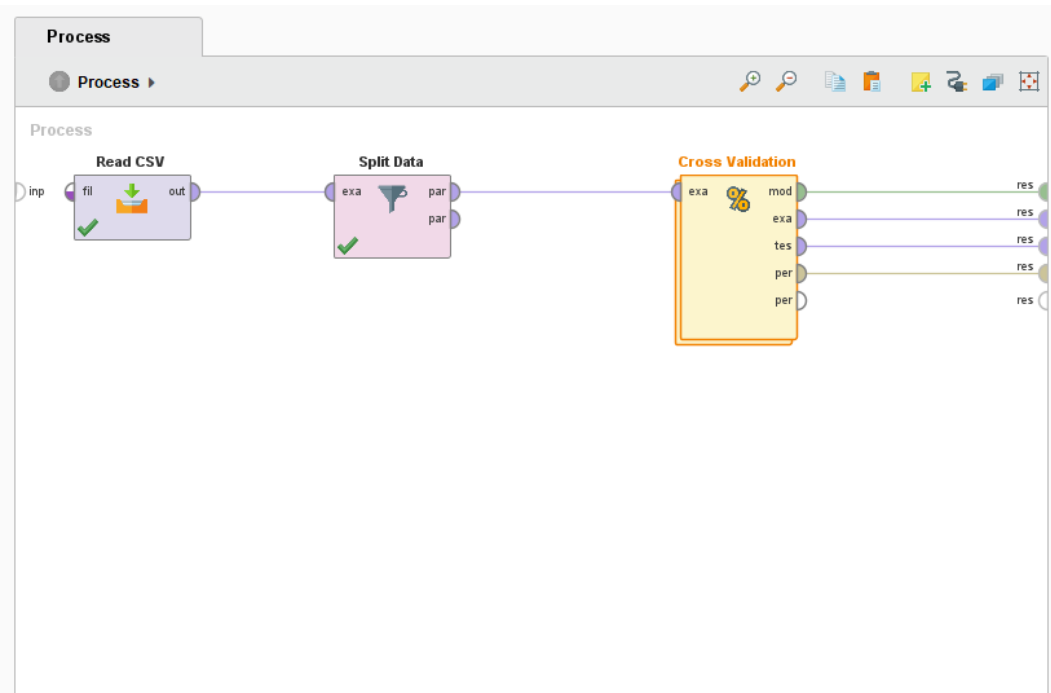
Gambar 4. 4

Penentuan jenis data

Kemudian cari pada *tool operator* dan *drag tool Read csv* untuk memasukkan data yang akan dihitung. Kemudian klik *Import Configuration Wizard* pada jendela kiri dan pilihlah data dalam bentuk csv yang akan dihitung. Kemudian dipilihlah kolom csv dari atribut yang akan digunakan. Selanjutnya adalah memilih tipe data dari data yang dimasukkan, apakah itu integer, polynominal atau binominal. Dan juga jangan lupa tentukan manakah atribut label dari data tersebut.

4.3.2 Perhitungan dan validasi

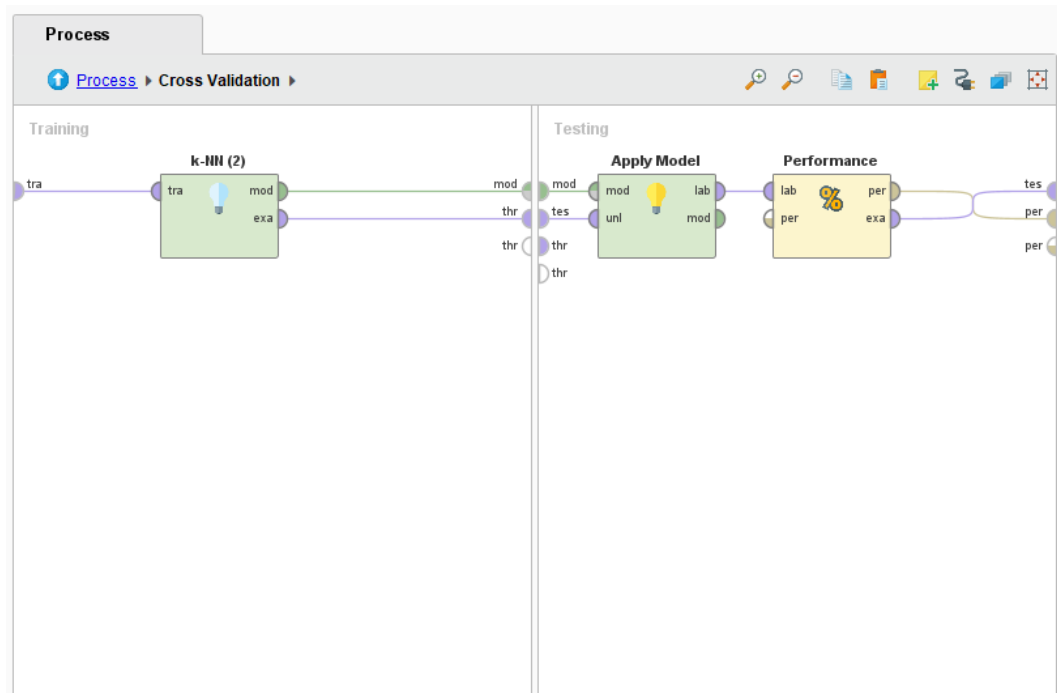
Dengan diinputkannya *operator split data* tersebut mengakhiri preprocessing data melalui RapidMiner tersebut, dan masuklah kita pada proses perhitungan menggunakan algoritma *K-Nearest Neighbors (KNN)*. Kemudian menghitung tingkat akurasi dengan *Confussion Matrix* yang terdapat pada operator *Cross Validation* sebagaimana gambar dibawah ini.



Gambar 4.5

Input Cross Validation pada RapidMiner

Kemudian klik 2x dan inputkan operator algoritma *K-Nearest Neighbors*, operator *Apply Model*, dan juga operator *performance* untuk menjalankan fungsi dari perhitungan algoritma tersebut. sebagaimana gambar dibawah ini.



Gambar 4. 6

Input operator KNN, Apply Model, dan Performance

4.3.3 Hasil Performance Vector

Hasil dari proses akan menghasilkan performance vector yang memperoleh nilai *accuracy*, *recall*, *precision*, *simple distribution*, dan AUC (*Area Under Curve*) sebagai berikut :

a. Accuracy

Dari pelatihan dan pengujian model KNN ini menghasilkan evaluasi akurasi yang tinggi, kita dapat melihat nilai akurasi sebagaimana tertera pada gambar 4.4 Yang menunjukkan bahwa pada penelitian ini memiliki akurasi yaitu sebesar 90,53% ini berarti menandakan bahwa dalam akurasi yang tinggi. Dan dalam penelitian ini memilih kesehatan janin

menggunakan metode K-NN berhasil memprediksi hasil dengan benar.

Nilai akurasi bisa dilihat pada gambar dibawah :

accuracy: 90.54% +/- 2.27% (micro average: 90.53%)

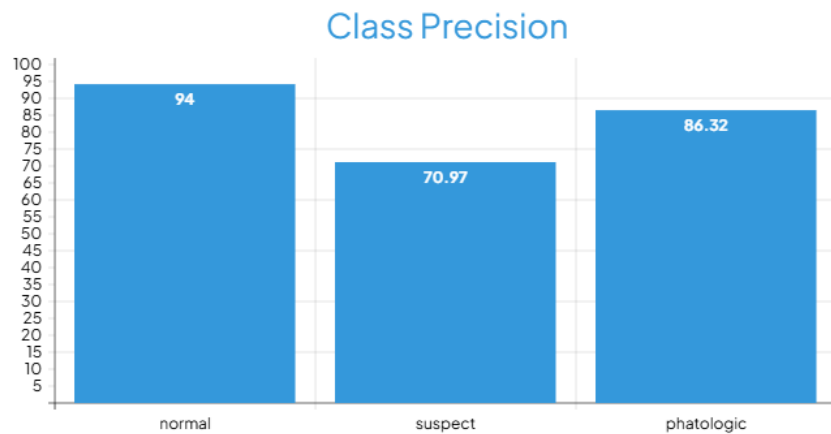
	true 2.0	true 1.0	true 3.0	class precision
pred. 2.0	154	37	26	70.97%
pred. 1.0	68	1285	14	94.00%
pred. 3.0	14	2	101	86.32%
class recall	65.25%	97.05%	71.63%	

Gambar 4. 7

Hasil akurasi

b. Precision

Pada gambar dibawah ini terlihat nilai akurasi klasifikasi berdasarkan menggunakan metode KNN mempunyai nilai masing – masing antara kelas 1,2, dan 3, sedangkan kelas 1 memperoleh nilai sebesar 94%, kelas 2 sebesar 70,97% dan kelas 3 sebesar 86,32%.



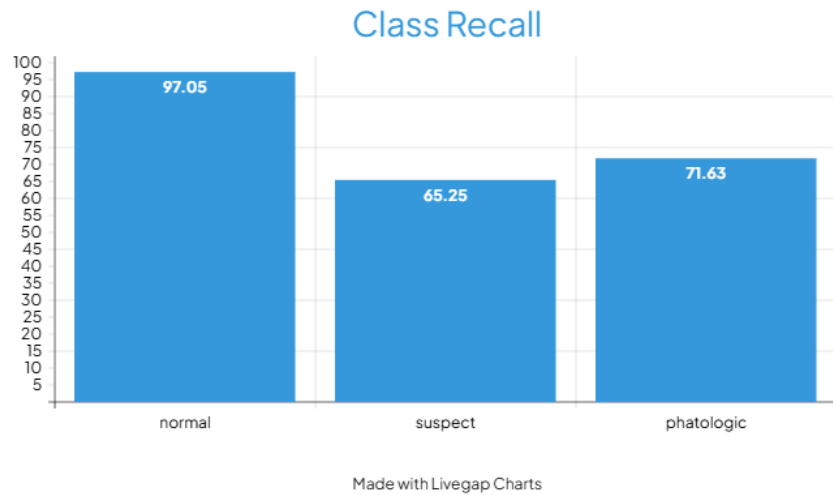
Gambar 4. 8

Hasil presisi

c. Recall

Pada gambar dibawah ini, terlihat nilai class recall dalam klasifikasi kesehatan janin menggunakan metode KNN. Terlihat niali kesesuaian

antara kelas normal dengan nilai 97,05%, kelas suspect dengan nilai 65,25%, dan kelas pathologic dengan nilai 71,63%



Gambar 4.9

Hasil recall



BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil pembahasan yang telah diuraikan pada bab sebelumnya, maka diperoleh kesimpulan sebagai berikut:

- a. implementasikan algoritma K-NN secara mandiri dan mendapatkan hasil akurasi yang cukup tinggi. Metode K-NN terbukti efektif dalam menentukan status kesehatan janin berdasarkan data medis yang relevan. K-NN dapat mengklasifikasikan status kesehatan janin dengan akurasi yang cukup tinggi.
- b. Pemilihan parameter dan fitur yang tepat sangat penting dalam implementasi K-NN. Fitur seperti detak jantung janin, gerakan janin, dan hasil pemeriksaan USG memberikan kontribusi signifikan terhadap akurasi klasifikasi. Data yang digunakan untuk melatih dan menguji model K-NN harus representative dan berkualitas cukup tinggi. Penggunaan data yang cukup dan beragam dapat meningkatkan performa model.
- c. Implementasi metode K-NN untuk penentuan status kesehatan janin dapat membantu tenaga medis dalam melakukan diagnosis awal yang cepat dan akurat, namun tetap diperlukan konfirmasi dari pemeriksaan medis lebih lanjut. Pengembangan lebih lanjut dapat mencakup optimasi parameter K-NN, penggunaan metode hybrid dengan algoritma lain, serta peningkatan kualitas dan kuantitas data untuk melatih model.

5.2 Saran

Berdasarkan hasil penelitian dan pembahasan serta kesimpulan yang telah dijelaskan, maka berikut ini merupakan sebuah saran yang diharapkan menjadi masukan dan pertimbangan untuk penelitian selanjutnya. Adapun beberapa saran sebagai berikut :

- a. Algoritma *K-nearest neighbors* ini hendaknya dapat diimplementasikan ke dalam sebuah aplikasi program yang dapat bermanfaat untuk membantu pemangku kesehatan menganalisa kesehatan janin, agar tujuan kesehatan janin dapat tercapai dan sesuai dengan sasaran.
- b. Untuk meningkatkan hasil kualitas penelitian ini agar lebih baik lagi, diharapkan untuk penelitian selanjutnya dapat mengembangkan dan menambah atau merubah atribut sehingga mendapatkan hasil klasifikasi yang lebih baik lagi.
- c. Diharapkan pula untuk penelitian selanjutnya dapat mencoba metode lain seperti metode C.45, dan lain – lain dalam pengklasifikasian sehingga kita dapat membandingkan metode mana yang memiliki tingkat akurasi yang lebih tinggi antara metode Algoritma *K-nearest neighbors* dan C.45 dalam pengklasifikasian. Dengan demikian, metode *K-NN* dapat menjadi alat bantu yang berguna dalam penentuan status kesehatan janin, mendukung tenaga medis dalam pengambilan keputusan yang lebih informative dan berbasis data

DAFTAR PUSTAKA

- [1] R. R. Dixit, "Predicting Fetal Health using Cardiotocograms: A Machine Learning Approach," Jan. 2022.
- [2] C. Sundar, M. Chitradevi, and G. Geetharamani, "Incapable of identifying suspicious records in CTG data using ANN based machine learning techniques," *J. Sci. Ind. Res. (India)*, vol. 73, no. 8, pp. 510–516, Aug. 2014.
- [3] T. Susanti, E. K. Suryaningsih, A. Kebidanan, and W. Buana, "PENGARUH PRENATAL YOGA TERHADAP KESEHATAN IBU DAN JANIN : SCOPING REVIEW The Effect Of Prenatal Yoga In Maternal And Fetal Wellness : Scoping Review," *Avicenna J. Heal. Res.*, vol. 6, no. 2, p. 112, Oct. 2023.
- [4] I. Sulihati, A. Syukur, and A. Marjuni, "Deteksi Kesehatan Janin Menggunakan Decision Tree dan Feature Forward Selection," *Build. Informatics, Technol. Sci.*, vol. 4, no. 3, pp. 1658–1664, Dec. 2022, doi: 10.47065/bits.v4i3.2672.
- [5] Y. Pratama, A. Prayitno, D. Azrian, N. Aini, Y. Rizki, and E. Rasywir, "Klasifikasi Penyakit Gagal Jantung Menggunakan Algoritma K-Nearest Neighbor," *Bull. Comput. Sci. Res.*, vol. 3, no. 1, pp. 52–56, Dec. 2022, doi: 10.47065/bulletincsr.v3i1.203.
- [6] J. Homepage, S. R. Cholil, T. Handayani, R. Prathivi, and T. Ardianita,

- “IJCIT (Indonesian Journal on Computer and Information Technology) Implementasi Algoritma Klasifikasi K-Nearest Neighbor (KNN) Untuk Klasifikasi Seleksi Penerima Beasiswa,” *IJCIT (Indonesian J. Comput. Inf. Technol.*, vol. 6, no. 2, pp. 118–127, Jul. 2021.
- [7] Y. Salini, S. N. Mohanty, J. V. N. Ramesh, M. Yang, and M. M. V. Chalapathi, “Cardiotocography Data Analysis for Fetal Health Classification Using Machine Learning Models,” *IEEE Access*, vol. 12, no. February, pp. 26005–26022, Feb. 2024, doi: 10.1109/ACCESS.2024.3364755.
- [8] R. Rahmadini, Enjel Erika LorencisLubis, Aji Priansyah, Yolanda R.W.N, and Tuti Meutia, “Penerapan Data Mining Untuk Memprediksi Harga Bahan Pangan Di Indonesia Menggunakan Algoritma K-Nearest Neighbor,” *J. Mhs. Akunt. Samudra*, vol. 4, no. 4, pp. 223–235, Aug. 2023, doi: 10.33059/jmas.v4i4.7074.
- [9] S. Aldana and J. Sasongko Wibowo, “Penerapan Data Mining Terhadap Klasifikasi Pasien Penderita Penyakit Liver Menggunakan Metode K-Nearest Neighbor,” *J. Ilm. Komput.*, vol. 20, no. 1, pp. 124–132, Feb. 2024.
- [10] R. satria Wahono, “Data Mining,” *Data Mining*, May 2020. <https://romisatriawahono.net/dm/>.
- [11] M. S. Haris *et al.*, *PENGANTAR DATA SCIENCE*, Pertama. makassar: Penerbit Widina, 2023.
- [12] Aprilla Dennis, *Belajar Data Mining dengan RapidMiner*, vol. 5, no. 4.

2013.



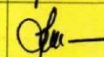
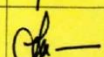
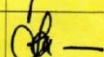
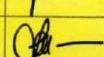
- [13] S. R. Annisa, “Modul Pengantar Aplikasi Komputer (PAK 240),” medan, Sep. 2019.
- [14] G. J. M. Rosa, “The Elements of Statistical Learning: Data Mining, Inference, and Prediction by HASTIE, T., TIBSHIRANI, R., and FRIEDMAN, J.” Oxford University Press, Aug. 2010.
- [15] W. Ode Nurhayah Kadir, B. Pramono, and Stiswaty, “Penerapan Data Mining Dengan Metode K-Nearest Neighbor(KNN) Untuk Mengelompokan Minat Konsumen Asuransi (Pt. Jasaraharja Putera),” *semanTIK*, vol. 5, no. 1, pp. 97–104, Jun. 2019, [Online]. Available: <http://ojs.uho.ac.id/index.php/semantik/article/view/6141>.

LAMPIRAN

Pembimbing I : Ahmad Homaidi, M.Kom

NO	TANGGAL	CATATAN	PARAF
1.	20 Feb 2024	Pengajuan Judul	
2.	23 Feb 2024	Bab I & II	
3.	27 Mei 2024	Konsultasi Pata	
4.	04 Juli 2024	Bab III Revisi	
5.	20 Juli 2024	Acc Bab III	
6.	04 Agustus 2024	Bab IV & Bab V	
7.	27 Agustus 2024	Acc	

Pembimbing II : Luleman Fakhri Lidimillah, M.Kom

NO	TANGGAL	CATATAN	PARAF
1	23 Feb 2024	Bab I dan II	
2	30 Mei 2024	Bab II	
3	15 Juli 2024	Bab III	
4	10 Agustus 2024	Perbaiki data	
5	19 Agustus 2024	Bab IV & Bab V	
6	27 Agustus 2024	Acc	



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET, DAN TEKNOLOGI
UNIVERSITAS JENDERAL SOEDIRMAN
FAKULTAS TEKNIK
JURUSAN INFORMATIKA
PANITIA SENIKO (SEMINAR NASIONAL INFORMATIKA DAN KOMPUTER)
 Jl. Mayor Jend. Sungkono KM 5 Blater Purbalingga 53371, Telp./Fax.: (0281) 6596700
 Laman: <https://seniko.ft.unsoed.ac.id>

Nomor : 042/SENIKO/VI/2024 15 Juni 2024
 Lampiran : 3 lembar
 Hal : Undangan Peserta SENIKO

Yth. Maulidatul Mawaddah
 Di Universitas Ibrahimiy
 Dengan hormat,

Melalui surat ini kami menginformasikan kembali bahwa SENIKO (Seminar Nasional Informatika dan Komputer) akan diselenggarakan pada:


Hari/Tanggal : Senin, 24 Juni 2024
 Waktu : Pukul 08.00 WIB - selesai
 Tempat : Java Heritage Hotel Purwokerto

Sehubungan dengan itu, kami mengundang Anda untuk menghadiri acara tersebut sesuai dengan bentuk kehadiran (daring/luring) yang telah Anda pilih pada saat pendaftaran acara. Sementara itu, untuk konfirmasi kehadirannya Anda dipersilakan untuk mengisi data RSVP pada tautan berikut ini: <https://bit.ly/rsvp-peserta-seniko> paling lambat tanggal 22 Juni 2024. Pengisian RSVP tersebut hanya diperlukan jika Anda belum pernah mengisinya sebelumnya. Demikian surat ini kami sampaikan. Atas perhatian dan kerja samanya kami ucapkan terima kasih.

Hormat kami,



Ketua Panitia,
Mochammad Agri Triansyah, S.Kom., M.Kom.
 NIP. 199408122023211023

Sekretaris,

Nur Alfi Ekowati, S.Kom., M.Sc.
 NIP. 199001302023212032

Mengetahui,

Ketua Jurusan Informatika
 Universitas Jenderal Soedirman,

Dr. Ir. Lasmedi Afuan, S.T., M.Cs., IPM.
 NIP. 198505102008121002



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN,
RISET DAN TEKNOLOGI

UNIVERSITAS JENDERAL SOEDIRMAN
JURNAL TEKNIK INFORMATIKA (JUTIF)

Jalan Mayor Jenderal Sungkono Km. 5, Blater Purbalingga 53371
Laman : <http://jutif.if.unsoed.ac.id>, email : jutif.ft@unsoed.ac.id

21 Juni 2024

No : 2173/LoA/JUTIF/IV/2024
Lampiran : -
Hal : Surat Pemberitahuan Bukti Penerimaan Paper

Kepada
Yth

1. Maulidatul Mawaddah, Information Technology, Faculty of Science and Technology, Universitas Ibrahimi, Indonesia
2. Ahmad Homaidi, Information Technology, Faculty of Science and Technology, Universitas Ibrahimi, Indonesia
3. Lukman Fakhid Lidimillah, Information Technology, Faculty of Science and Technology, Universitas Ibrahimi, Indonesia

Jurnal Teknik Informatika (JUTIF) adalah jurnal nasional terdaftar di LIPI dengan P-ISSN : 2723-3863 dan e-ISSN : 2723-3871 serta terakreditasi SINTA 3 berdasarkan Surat Keputusan Direktur Jenderal Pendidikan Tinggi, Riset, Dan Teknologi Kementerian Pendidikan, Kebudayaan, Riset, Dan Teknologi Republik Indonesia Nomor 164/E/KPT/2021.

Kami mengucapkan terima kasih atas pengiriman artikel ilmiah ke **Jurnal Teknik Informatika (JUTIF)** dengan data artikel sebagai berikut:

Judul *IMPLEMENTATION OF THE K-NEAREST NEIGHBORS METHOD FOR DETERMINING FETAL HEALTH STATUS*

Penulis
1. Maulidatul Mawaddah
2. Ahmad Homaidi
3. Lukman Fakhid Lidimillah

Berdasarkan hasil review, artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan pada **Jurnal Teknik Informatika (JUTIF) UNIVERSITAS JENDERAL SOEDIRMAN (UNSOED)** Volume 5 Nomor 4 Agustus 2024.

Demikian surat pemberitahuan ini, kami mengucapkan terima kasih atas partisipasi dan kerja sama yang baik.

Chief Editor,



Dr. Lasmedi Afuan, S.T.,M.Cs
NIP. 19850510 200812 1 002



SERTIFIKAT Penghargaan

SERTIFIKAT INI DIBERIKAN KEPADA

Maulidatul Mawaddah

ATAS PARTISIPASINYA SEBAGAI PRESENTER DALAM ARTIKEL IMPLEMENTATION OF THE K-NEAREST NEIGHBORS METHOD FOR DETERMINING FETAL HEALTH STATUS PADA ACARA SEMINAR NASIONAL INFORMATIKA DAN KOMPUTER (SENIKO) YANG DISELENGGARAKAN PADA TANGGAL 24 JUNI 2024.

Dr. Ir. Lasmedi Afuan, S.T., M.Cs., IPM.
Ketua Jurusan

Mochammad Agri Triansyah, S.Kom., M.Kom.
Ketua Panitia

Co-Hosted by:

Supported by:

Media Partner:

CURICULUM VITAE

Nama Lengkap : Maulidatul Mawaddah
NPM : 2020503050
Tempat, Tanggal Lahir : Situbondo, 26 Mei 2002
Program Studi : Teknologi Informasi
Nama Orang Tua
Ayah : Ahmad Subairi
Ibu : Eti Fidiyani
Latar Belakang Pendidikan
SD/MI : SDN Rambipuji 02
SMP/MTs : SMP Ibrahimy 03
SMA/SMK/MA : SMK Ibrahimy 1
Latar Organisasi : - BEM Universitas Ibrahimy
- IKSASS Rayon Jember
Alamat Rumah : Rambipuji - Jember
No. Telepon : 081359638657
E-mail : mauidatulm271@gmail.com