

**PENERAPAN ALGORITMA NAÏVE BAYES DAN GREEDY  
FORWARD SELECTION UNTUK PREDIKSI GAGAL  
JANTUNG**

**PROPOSAL TUGAS AKHIR**



Diajukan oleh :

Dwiki Suastu Aditia

8020190297

Untuk Persyaratan Penelitian Dan Penulisan Tugas Akhir

Sebagai Akhir Proses Studi Strata 1

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS ILMU KOMPUTER**

**UNIVERSITAS DINAMIKA BANGSA**

**JAMBI**

**2022**

## **IDENTITAS PROPOSAL PENELITIAN**

Judul Proposal : **PENERAPAN ALGORITMA NAÏVE BAYES DAN  
GREEDY FORWARD SELECTION UNTUK  
PREDIKSI GAGAL JANTUNG**

Program Studi : TI

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

a. Nama : Dwiki Suastu Aditia

b. NIM : 8020190297

c. Jenis Kelamin : Laki - Laki

d. Tempat/Tgl. Lahir : Jambi, 19 Januari 2001

e. Alamat : Talang Bakung, Jambi

f. No. Telepon : 08877876489

g. Email : dwikisuastuaditia@gmail.com

## PERNYATAAN HASIL EVALUASI

NAMA/NIM : Dwiki Suastu Aditia / 8020190297

PRODI : TI / ~~SI~~ / ~~SK~~ \*)

JUDUL : **PENERAPAN ALGORITMA NAÏVE BAYES DAN  
GREEDY FORWARD SELECTION UNTUK  
PREDIKSI GAGAL JANTUNG**

Hasil Evaluasi : Disetujui / Disetujui dengan perbaikan / Ditolak \*)

### 1. Catatan

Alasan Penolakan Proposal Metode Penelitian :

- ☐ Proyek Metode Penelitian tidak relevan dengan Program Studi
- ☐ Pernah ada topik sejenis
- ☐ Metode utama telah banyak dipakai
- ☐ Metode yang dipakai tidak jelas
- ☐ Masalah terlalu sempit
- ☐ -----

\*) Coret yang tidak perlu

# PENDAHULUAN

## 1.1 LATAR BELAKANG MASALAH

Pada saat ini, kita hidup di zaman modernisasi. Modernisasi sendiri merujuk pada bentuk transformasi dari keadaan yang kurang maju atau kurang berkembang ke arah yang lebih baik dengan harapan membuat kehidupan menjadi lebih baik. Di zaman modernisasi seperti sekarang, manusia sangat bergantung pada teknologi. Perkembangan teknologi yang begitu pesat merambah ke berbagai sektor termasuk kesehatan. Dengan bantuan teknologi, dapat membantu dalam banyak pekerjaan di bidang medis. Selain memiliki potensi dalam memfilter data dan mengolah menjadi informasi, juga mampu menyimpan dalam jumlah kapasitas jauh lebih besar. Konvergensi dengan teknologi informasi juga memungkinkan data kesehatan di-share dengan mudah dan cepat.

Gagal jantung susah dikenali secara klinis, karena beragamnya keadaan klinis serta tidak spesifik serta hanya sedikit tanda – tanda klinis pada tahap awal penyakit. Perkembangan terkini memungkinkan untuk mengenali gagal jantung secara dini serta perkembangan pengobatan yang memperbaiki gejala klinis, kualitas hidup, penurunan angka perawatan, memperlambat progresifitas penyakit dan meningkatkan kelangsungan hidup.

Algoritma Naïve Bayes adalah pengklasifikasi probabilitas sederhana yang menghitung sekumpulan probabilitas dengan menghitung frekuensi dan kombinasi nilai dalam kumpulan data yang diberikan. Klasifikasi menggunakan Naïve Bayes dapat menghasilkan akurasi yang tinggi dan cepat ketika diaplikasikan pada data yang besar serta mudah dalam penerapannya. Seleksi fitur menggunakan Greedy Forward Selection sangat efisien, sederhana dan tidak membutuhkan waktu lama dalam prosesnya.

Menurut World Health Organization (WHO), penyakit kardiovaskular akan menjadi penyebab terbanyak kasus kematian di seluruh dunia. Di Indonesia, penyakit gagal jantung kongestif telah menjadi pembunuh nomor satu. Prevalensi penyakit jantung di Indonesia dari tahun ke tahun semakin meningkat. Meski menjadi pembunuh utama, tetapi masih sedikit sekali orang yang tahu tentang

penyakit gagal jantung dan faktor risikonya. Faktor risiko terjadinya gagal jantung yang paling sering adalah usia lanjut. Memiliki faktor risiko lebih dari satu seperti hipertensi, diabetes mellitus, dan merokok maka akan mempunyai peluang yang lebih besar terkena penyakit gagal jantung dibanding yang tidak. Faktor risiko tersebut berperan penting untuk terjadinya penyakit gagal jantung. Prognosis dari penyakit gagal jantung akan buruk bila dasar atau penyebabnya tidak dapat diperbaiki. Apabila faktor risiko dapat diketahui maka akan lebih mudah untuk dilakukannya tindakan pencegahan. Dengan latar belakang yang penulis kemukakan maka penulis ingin merancang “Penerapan Algoritma Naïve Bayes Dan Greedy Forward Selection Untuk Prediksi Gagal Jantung”

## **1.2 RUMUS MASALAH**

Berdasarkan latar belakang yang sudah dijelaskan, maka dapat diambil suatu rumusan masalah yaitu sebagai berikut :

1. Bagaimana menerapkan algoritma naïve bayes dan greedy forward selection untuk prediksi gagal jantung.
2. Berapa tingkat akurasi dari penerapan algoritma naïve bayes dan greedy forward selection untuk prediksi gagal jantung

## **1.3 BATASAN MASALAH**

Agar dalam penelitian ini dapat berjalan dengan baik dan terarah penulis menetapkan ruang lingkup penelitian meliputi :

1. Pada penelitian ini peneliti menggunakan metode klasifikasi dengan metode clustering dengan algoritma k-means.
2. Dataset yang digunakan berasal dari UCI Machine Learning dari penelitian Davide Chicco (Krembil Research Institute, Toronto, Canada) pada tahun 2020.
3. Atribut yang di gunakan yaitu : age, anaemia, creatinine\_phosphokinase, diabetes, ejection\_fraction, high\_blood\_pressure, platelets, serum\_creatinine, serum\_sodium, sex, smoking, time, dan death\_event.
4. Alat bantu (tools) yang digunakan pada penelitian ini adalah WEKA dan Rapidminer.

## **1.4 TUJUAN DAN MANFAAT PENELITIAN**

### **1.4.1 Tujuan Penelitian**

Berdasarkan permasalahan yang tertera diatas, maka dapat disimpulkan bahwa tujuan dari penelitian ini adalah :

1. Menerapkan algoritma naïve bayes dan greedy forward selection untuk prediksi gagal jantung.
2. Mengetahui tingkat akurasi dari penerapan algoritma naïve bayes dan greedy forward selection untuk prediksi gagal jantung.

### **1.4.2 Manfaat Penelitian**

Serta manfaat yang akan didapat dari penelitian, yaitu:

1. Dapat melihat bagaimana pengaruh dari atribut yang memiliki banyak data dalam memprediksi gagal jantung.
2. Penulis dapat menambah ilmu dan wawasan baru mengenai menerapkan algoritma naïve bayes dan greedy forward selection untuk prediksi gagal jantung.
3. Dapat digunakan sebagai acuan untuk melakukan penelitian selanjutnya.

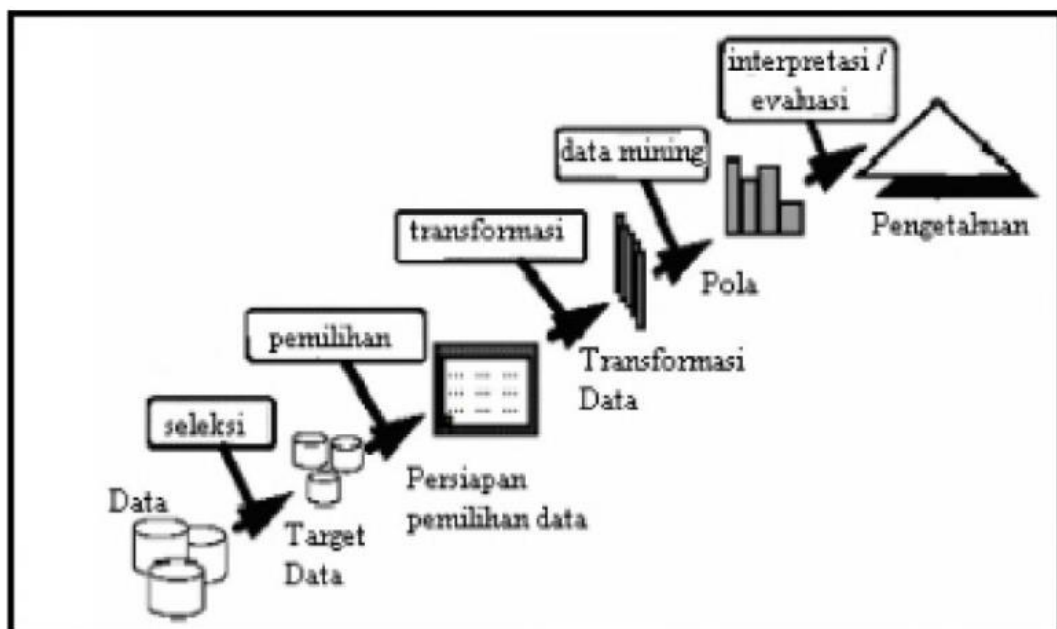
## LANDASAN TEORI

### 2.1 DATA MINING

Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar.

Menurut A. Muzakir and R. A. Wulandari “Data mining adalah proses mencari pola atau informasi menarik dalam data terpilih dengan menggunakan teknik atau metode tertentu. Teknik, metode atau algoritma data mining sangat bervariasi. Pemilihan metode atau algoritma yang tepat sangat bergantung pada tujuan dan proses Knowledge Discovery in Database (KDD).” [2]

Menurut Dicky Nofriansyah dan Gunadi Widi Nurcahyo “proses Data Mining yang biasa disebut Knowledge Discovery Database(KDD) terdapat beberapa proses” [3] seperti terlihat pada gambar 2.1



Gambar 2. 1 Langkah-langkah Data Mining

1. Seleksi Data (Selection)

Selection (seleksi/pemilihan) data dari merupakan sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam Knowledge Discovery Database (KDD) dimulai.

2. Pemilihan Data (Preprocessing/Cleaning)

Proses Preprocessing mencakup antara lain membuang duplikasi data, memeriksa data yang inkonsisten, dan (tipografi).

3. Transformasi (Transformation)

Pada fase ini yang dilakukan adalah mentransformasi bentuk data yang belum memiliki entitas yang jelas kedalam bentuk data yang valid atau siap untuk dilakukan proses Data Mining.

4. Data Mining

Pada fase ini yang dilakukan adalah menerapkan algoritma atau metode pencarian pengetahuan.

5. Interpretasi/Evaluasi (Interpretation/Evaluation)

Pada fase terakhir ini yang dilakukan adalah proses pembentukan keluaran yang mudah dimengerti yang bersumber pada proses Data Mining Pola informasi

Menurut Budi Santosa & Ardian Umam “Tugas-tugas yang biasa dilakukan *data mining*” [4]:

- a. Klustering

Mengelompokkan obyek ke dalam beberapa kelompok berdasarkan kemiripan antar obyek, dimana dalam satu kluster harus berisi obyek yang saling mirip dan antar kluster obyek saling tidak mirip. Klustering ini tidak memerlukan data pelatihan yang sudah diberi label.



b. Klasifikasi

Melakukan pengelompokan obyek berdasarkan kelompok yang sudah ada. Berbeda dengan klastering, klasifikasi ini memerlukan data pelatihan yang sudah diberi label kelompok/kelas.

c. Regresi/Estimasi

Regresi pada dasarnya mirip dengan klasifikasi, yakni memerlukan data pelatihan yang sudah diberi label. Bedanya, output klasifikasi adalah nilai diskrit, sedangkan output dari regresi adalah nilai kontinyu. Regresi ini mencari model hubungan antar atribut predictor dan atribut dependen, dimana atribut dependennya juga berupa nilai kontinyu.

d. Asosiasi

Melakukan asosiasi antar obyek dalam suatu set data, biasanya data transaksional. Asosiasi dilakukan dengan menghitung berapa kali dalam suatu set data suatu transaksi yang mengandung dua item atau lebih yang berhubungan. Sering ada yang menyebutnya Market Basket

e. Analytics.

Berdasarkan pengertian diatas dapat disimpulkan bahwa data mining merupakan proses mengekstrak dan mengidentifikasi data yang kita peroleh menjadi informasi yang dapat membantu untuk proses Klastering, Klasifikasi, Regresi/Estimasi dan Asosiasi.

Berdasarkan pengertian diatas dapat disimpulkan bahwa *data mining* merupakan proses mengekstrak dan mengidentifikasi data yang kita peroleh menjadi informasi yang dapat membantu untuk proses Klastering, Klasifikasi, Regresi/Estimasi dan Asosiasi.

## 2.2 KLASIFIKASI

Klasifikasi adalah tipe analisis data yang dapat membantu orang menentukan kelas label dari sampel yang ingin di klasifikasi.

Menurut Senna Hendrian “Klasifikasi merupakan Metode supervised learning, metode yang mencoba menemukan hubungan antara atribut masukan dan atribut target.” [5]

Beberapa algoritma yang dapat digunakan dalam klasifikasi *data mining* adalah sebagai berikut [6] :

A. Neural Network.

Neural Network (Jaringan Saraf Tiruan) adalah prosesor tersebar paralel yang sangat besar dan memiliki kecenderungan untuk menyimpan pengetahuan yang bersifat pengalaman dan membuatnya siap untuk digunakan.

B. Decision Tree.

Decision tree sendiri merupakan metode klasifikasi dan prediksi yang sangat kuat dan banyak diminati. Dalam decision tree ini data yang berupa fakta dirubah menjadi sebuah pohon keputusan yang berisi aturan dan tentunya dapat lebih mudah dipahami dengan bahasa alami. Model pohon keputusan banyak digunakan pada kasus data dengan output yang bernilai diskrit. Walaupun tidak menutup kemungkinan dapat juga digunakan untuk kasus data dengan atribut numeric.

C. Naive Bayes.

Naïve Bayes merupakan sebuah model klasifikasi statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Naïve Bayes didasarkan pada teorema bayes yang memiliki kemampuan klasifikasi serupa dengan decision tree dan neural network.

D. K-Nearest Neighbor.

Algoritma k-nearest neighbor (k-NN atau KNN) adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut, Ketepatan algoritma k-NN ini sangat dipengaruhi oleh ada atau tidaknya fitur-fitur yang tidak relevan, atau jika bobot fitur tersebut tidak setara dengan relevansinya terhadap klasifikasi.

E. Logistic Regression.

Regresi logistik (Logistic regression) adalah bagian dari analisis regresi yang digunakan ketika variabel dependen (respon) merupakan variabel dikotomi.

Jadi dapat disimpulkan bahwa klasifikasi merupakan kegiatan yang digunakan untuk mengelompokkan data kedalam kelas-kelas tertentu dalam pengerjaan klasifikasi dapat menggunakan beberapa metode algoritma seperti *Neural Network, Decision Tree, Naïve Bayes, K-Nearest Neighbor, Logistic Regression*.

### 2.3 ALGORITMA NAÏVE BAYES

Naïve Bayes classifier (NBC) merupakan salah satu metode pembelajaran mesin yang memanfaatkan perhitungan probabilitas dan statistic yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya.

Menurut AR Walad Mahfuzhi “Bayes merupakan teknik prediksi berbasis probabilistic sederhana yang berdasar pada penerapan teorema bayes (atau aturan bayes) dengan asumsi idependensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam Naïve Bayes, model yang digunakan adalah *model fitur independent*.”[7]

Menurut Sartika & Sansuse “Naïve Bayes merupakan perhitungan teorema bayes yang paling sederhana, karena mampu mengurangi kompleksitas komputasi menjadi multiplikasi sederhana dari probabilitas. Selain itu, algoritma Naïve Bayes juga mampu menangani set data yang memiliki banyak atribut. Persamaan dari Naïve Bayes sebagai berikut” [8] :

$$P ( C_i | X = \frac{P(X | C_i)}{P(X)} \quad (2.1)$$

Penjelasan dari formula tersebut adalh sebagai berikut :

$X$  = Kriteria Suatu kasus bedasarkan masukan.

$C_i$  = Kelas solusi pola ke- $i$ , dimana  $i$  adalah jumlah label kelas.

$P(C_i | X)$  = Probabilitas kemunculan label kelas  $C_i$  dengan kriteria masukan  $X$ .

$P(X | C_i)$  = Probabilitas kriteria masukan  $X$  dengan label kelas  $C_i$ .

$P(X)$  = Probabilitas label kelas  $C_i$ .

Berdasarkan pengertian diatas dapat disimpulkan bahwa algoritma *naïve bayes* merupakan salah satu metode algoritma yang memiliki independensi yang kuat (naif) yang digunakan untuk memprediksi kemungkinan dimasa yang akan datang melalui pengalaman dari masa lampau.

Pada pengukuran kinerja menggunakan confusion matrix, terdapat 4 (empat) istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut 15 adalah True Positive (TP), True Negative (TN), False Positive (FP) dan False Negative (FN). Nilai True Negative (TN) merupakan jumlah data negatif yang terdeteksi dengan benar, sedangkan False Positive (FP) merupakan data negatif namun terdeteksi sebagai data positif. Sementara itu, True Positive (TP) merupakan data positif yang terdeteksi benar. False Negative (FN) merupakan kebalikan dari True Positive, sehingga data positif, namun terdeteksi sebagai data negative.

Pada jenis klasifikasi *binary* yang hanya memiliki 2 keluaran kelas, *confusion matrix* dapat disajikan seperti pada table 2.1 :

**Table 2. 1 Tabel *Confusion Matrix* [9]**

| Correct Classification | Classified as   |                 |
|------------------------|-----------------|-----------------|
|                        | Predicted “ + “ | Predicted “ - “ |
| Actual “ + ”           | True Positives  | False Negatives |
| Actual “ – “           | False Positives | True Negatives  |

Berdasarkan nilai True Negative (TN), False Positive (FP), False Negative (FN), dan True Positive (TP) dapat diperoleh nilai akurasi, presisi dan recall. Nilai akurasi menggambarkan seberapa akurat sistem dapat mengklasifikasikan data secara benar. Dengan kata lain, nilai akurasi merupakan perbandingan antara data yang terklasifikasi benar dengan keseluruhan data. Nilai akurasi dapat diperoleh

dengan Persamaan 1. Nilai presisi menggambarkan jumlah data kategori positif yang diklasifikasikan secara benar dibagi dengan total data yang diklasifikasi positif. Presisi dapat diperoleh dengan Persamaan 2. Sementara itu, recall menunjukkan berapa persen data kategori positif yang terklasifikasikan dengan benar oleh sistem. Nilai recall diperoleh dengan Persamaan 3.

$$Akurasi = \frac{TP+TN}{TP+TN+FP+FN} * 100\% \quad (2.2)$$

$$Presisi = \frac{TP}{FP+TP} * 100\% \quad (2.3)$$

$$Recall = \frac{TP}{FN+TP} * 100\% \quad (2.4)$$

dimana :

- TP adalah *True Positive*, yaitu jumlah data positif yang terklasifikasi dengan benar oleh system.
- TN adalah *True Negative*, yaitu jumlah data negatif yang terklasifikasi dengan benar oleh sistem.
- FN adalah *False Negative*, yaitu jumlah data negatif namun terklasifikasi salah oleh sistem.
- FP adalah *False Positive*, yaitu jumlah data positif namun terklasifikasi salah oleh system

## 2.4 WEKA

*WEKA* merupakan sebuah perangkat lunak yang menerapkan berbagai algoritma *machine learning* untuk melakukan beberapa proses yang berkaitan dengan system temu Kembali informasi atau *data mining*.

Menurut Witten, Frank & Hall "The Waikato Environment for Knowledge Analysis (Weka) adalah rangkaian lengkap perpustakaan kelas Java yang mengimplementasikan banyak state-of-the-art pembelajaran mesin dan algoritma data mining. Weka tersedia secara bebas di World Wide Web dan menyertai teks baru pada dokumen data mining dan sepenuhnya menjelaskan semua algoritma yang dikandungnya. Aplikasi yang ditulis menggunakan library class pada Weka yang dapat dijalankan pada komputer manapun dengan kemampuan browsing

Web, ini memungkinkan pengguna untuk menerapkan teknik pembelajaran mesin untuk data mereka sendiri terlepas dari platform komputer.”[10]



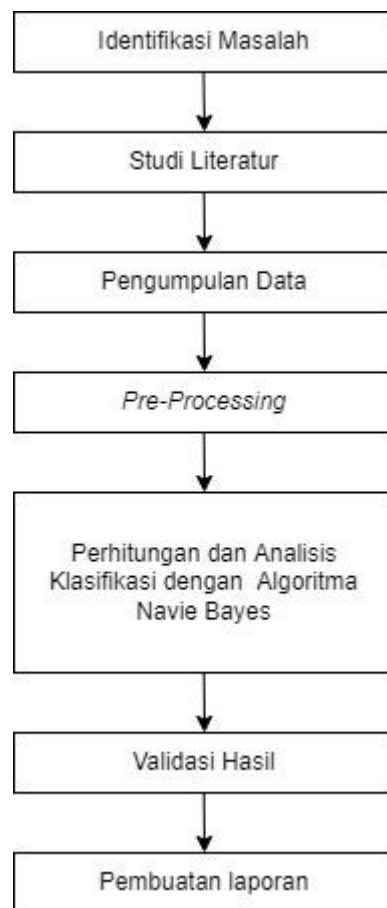
**Gambar 2. 2 Tampilan Halaman Utama Weka**

Berdasarkan pengertian diatas dapat disimpulkan bahwa weka merupakan rangkaian kelas java yang digunakan untuk melakukan beberapa proses *machine learning* dan algoritma *data mining*.

## METODOLOGI PENEITIAN

### 3.1 KERANGKA KERJA PENELITIAN

Kerangka kerja penelitian merupakan tahapan-tahapan yang dilakukan selama mengerjakan penelitian. Kerangka kerja penelitian tepat mempermudah pencapaian hasil penelitian, dapat menyelesaikan penelitian tepat waktu dan penelitian dapat berjalan sesuai dengan yang diharapkan. Adapun kerangka kerja penelitian yang digunakan dapat dilihat pada gambar 3.1



**Gambar 3. 1 Kerangka Kerja Penelitian**

Berdasarkan kerangka kerja penelitian yang telah digambarkan 3.1, maka dapat diuraikan pembahasan masing-masing tahap dalam penelitian adalah sebagai berikut :

1. Identifikasi Masalah

Pada tahapan ini penulis melakukan identifikasi masalah terhadap klasifikasi penentuan rating video game. Identifikasi ini bertujuan untuk mengkaji permasalahan yang ada pada pemain.

2. Studi Literatur

Pada tahapan ini penulis menambah wawasan guna mendapatkan sebuah topik yang layak diangkat sebagai sebuah penelitian dengan mempelajari dan memahami teori dan konsep dimana penulis banyak melakukan penelitian pada buku, jurnal, *paper*, dan berbagai sumber diantaranya Permustakaan UNAMA Jambi.

3. Pengumpulan Data

Pada pengumpulan data penulis melakukan pengamatan di web UCI Machine Learning Repository

4. *Pre-Processing*

Pada tahapan ini peneliti melakukan persiapan data semua game. Setelah itu peneliti melakukan seleksi atribut yang berpengaruh dalam *classification* untuk menentukan kategori pemain yang kemungkinan memilih game yang sudah ditentukan.

- a. *Data Selection*

Pemilihan (seleksi) data baru sekumpulan data operasional perlu dilakukan sebelum tahap penggalian informasi dalam *data mining* dimulai. Data hasil seleksi yang akan digunakan untuk proses *data mining*, disimpan dalam suatu berkas, terpisah dari basis data operasional.

- b. *Transformation*

Pada tahapan ini penulis melakukan transformasi data yaitu proses pengubahan data ke dalam format yang sesuai untuk diproses dalam *data mining*. Data di rubah dengan format *comma seprated files*



(CSV) atau Andrew's Ridiculos file format (ARFF) sebelum bisa diaplikasikan ke Weka 3.8.6. Data tersebut berupa data video game yang telah di *cleaning* pada tahap sebelumnya yang ditransformasikan yaitu mengubah format yang awalnya dari data excel (XLS) ke CSV atau arff.

5. Perhitungan dan Analisis dengan Algoritma *Naïve Bayes*

Pada tahap ini penulis melakukan perhitungan *Naïve Bayes* dengan menggunakan tools WEKA terhadap data yang telah didapatkan dari Puskesmas Talang Bakung.

6. Validasi Hasil

Validasi hasil menunjukan kedekatan hasil pengukuran dengan nilai *Accuracy*, *Prevision*, dan *Recall*.

7. Pembuatan Laporan

Setelah semua tahapan penelitian dilakukan, maka akan dibuat laporan sebagai dokumentasi penelitian agar dapat dimanfaatkan pada waktu yang akan datang. Baik oleh peneliti sendiri maupun peneliti lainnya.

### 3.2 ALAT BANTU PENELITIAN

Dalam melakukan penelitian, penulis juga memerlukan alat bantu, Alat bantu yang digunakan, yaitu :

1. Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan peneliti sebagai pengembangan adalah sebagai berikut :

- a. 1 unit Leptop Asus dengan Processor AMD Athlin Gold 3150U dengan kecepatan 2,4GHz.
- b. Memory RAM 8 GB.
- c. 512 GB SSD.
- d. 1 unit Flashdisk Sandisk 8GB.

2. Perangkat Lunak (*Software*)

Perangkat yang digunakan oleh pengembang dalam penelitian yaitu :

- a. Sistem operasi Windows 11 Home Single Language
- b. Microsoft Word dan Excel 2010.
- c. Alat bantu WEKA 3.8.6
- d. RapidMiner

## DAFTAR ISI

- [1] T. W. Candra Agustina, "A-Glucosidase Inhibitory Activity of Alginate Hydrolytes Produced By Acid," *J. Chinese Inst. Food Sci. Technol.*, vol. 15, no. 5, pp. 33–39, 2015, doi: 10.16429/j.1009-7848.2015.05.005.
- [2] A. Muzakir and R. A. Wulandari, "Model Data Mining sebagai Prediksi Penyakit Hipertensi Kehamilan dengan Teknik Decision Tree," *Sci. J. Informatics*, vol. 3, no. 1, pp. 19–26, 2016, doi: 10.15294/sji.v3i1.4610.
- [3] Dicky Nofriansyah dan Gunadi Widi Nurcahyo, *Algoritma Data Mining dan Pengujian*. yogyakarta: Deepublish, 2015. [Online]. Available: [https://www.google.co.id/books/edition/Algoritma\\_Data\\_Mining\\_Dan\\_Pengujian/Fn-QDwAAQBAJ?hl=id&gbpv=1&printsec=frontcover](https://www.google.co.id/books/edition/Algoritma_Data_Mining_Dan_Pengujian/Fn-QDwAAQBAJ?hl=id&gbpv=1&printsec=frontcover)
- [4] B. & A. U. Santosa, *Data Mining dan Big Data Analytic : Teori dan Implementasi Menggunakan Python & Apache Spark*. Yogyakarta, 2019. [Online]. Available: [https://books.google.co.id/books/about/Data\\_Mining\\_dan\\_Big\\_Data\\_Analytics\\_Teori.html?id=w1nUDwAAQBAJ&redir\\_esc=y](https://books.google.co.id/books/about/Data_Mining_dan_Big_Data_Analytics_Teori.html?id=w1nUDwAAQBAJ&redir_esc=y)
- [5] S. Hendrian, "Algoritma Klasifikasi Data Mining Untuk Memprediksi Siswa Dalam Memperoleh Bantuan Dana Pendidikan," *Fakt. Exacta*, vol. 11, no. 3, pp. 266–274, 2018, doi: 10.30998/faktorexacta.v11i3.2777.
- [6] S. Dewi, "Komparasi 5 Metode Algoritma Klasifikasi Data Mining Pada Prediksi Keberhasilan Pemasaran Produk Layanan Perbankan," *Techno Nusa Mandiri*, vol. 13, no. 1, pp. 60–66, 2016.
- [7] M. Halim, "Pembuatan Aplikasi Pendukung Keputusan Beasiswa Tidak Mampu Dengan Metode Naive Bayes," *JSai (Journal Sci. Appl. Informatics)*, vol. 1, no. 1, pp. 24–31, 2018, doi: 10.36085/jsai.v1i1.6.
- [8] D. Sartika and D. I. Sensuse, "Perbandingan Algoritma Klasifikasi Naive Bayes, Nearest Neighbour, dan Decision Tree pada Studi Kasus Pengambilan Keputusan Pemilihan Pola Pakaian," *JATISI (Jurnal Tek. Inform. dan Sist. Informasi)*, vol. 3, no. 2, pp. 151–161, 2017, [Online]. Available: <https://jurnal.mdp.ac.id/index.php/jatisi/article/view/78>
- [9] M. F. Rahman, D. Alamsah, M. I. Darmawidjadja, and I. Nurma, "Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN)," *J. Inform.*, vol. 11, no. 1, p. 36, 2017, doi: 10.26555/jifo.v11i1.a5452.
- [10] M. A. H. Ian H. Witten, Eibe Frank, *Data Mining: Practical Machine Learning Tools and Techniques, 3/E*. Burlington: Morgan Kaufmann, 2011. [Online]. Available: <https://openlibrary.telkomuniversity.ac.id/pustaka/29732/data-mining-practical-machine-learning-tools-and-techniques-3-e.html>

| NO  | anaemia | creatinine | diabetes | ejection_f | high_bloo | platelets | serum_cr | serum_so | sex | smoking | time | DEATH_EV |     |
|-----|---------|------------|----------|------------|-----------|-----------|----------|----------|-----|---------|------|----------|-----|
| 1   | 0       | 582        | 0        | 20         | 1         | 265000    | 1.9      |          | 130 | 1       | 0    | 4        | 1   |
| 2   | 0       | 7861       | 0        | 38         | 0         | 263358.03 | 1.1      |          | 136 | 1       | 0    | 6        | 1   |
| 3   | 0       | 146        | 0        | 20         | 0         | 162000    | 1.3      |          | 129 | 1       | 1    | 7        | 1   |
| 4   | 1       | 111        | 0        | 20         | 0         | 210000    | 1.9      |          | 137 | 1       | 0    | 7        | 1   |
| 5   | 1       | 160        | 1        | 20         | 0         | 327000    | 2.7      |          | 116 | 0       | 0    | 8        | 1   |
| 6   | 1       | 47         | 0        | 40         | 1         | 204000    | 2.1      |          | 132 | 1       | 1    | 8        | 1   |
| 7   | 1       | 246        | 0        | 15         | 0         | 127000    | 1.2      |          | 137 | 1       | 0    | 10       | 1   |
| 8   | 1       | 315        | 1        | 60         | 0         | 454000    | 1.1      |          | 131 | 1       | 1    | 10       | 1   |
| 9   | 0       | 157        | 0        | 65         | 0         | 263358.03 | 1.5      |          | 138 | 0       | 0    | 10       | 1   |
| 10  | 1       | 123        | 0        | 35         | 1         | 388000    | 9.4      |          | 133 | 1       | 1    | 10       | 1   |
| 11  | 1       | 81         | 0        | 38         | 1         | 368000    |          | 4        | 131 | 1       | 1    | 10       | 1   |
| 12  | 0       | 231        | 0        | 25         | 1         | 253000    | 0.9      |          | 140 | 1       | 1    | 10       | 1   |
| 13  | 1       | 981        | 0        | 30         | 0         | 136000    | 1.1      |          | 137 | 1       | 0    | 11       | 1   |
| 14  | 1       | 168        | 0        | 38         | 1         | 276000    | 1.1      |          | 137 | 1       | 0    | 11       | 1   |
| 15  | 1       | 80         | 0        | 30         | 1         | 427000    |          | 1        | 138 | 0       | 0    | 12       | 0   |
| 16  | 1       | 379        | 0        | 50         | 0         | 47000     | 1.3      |          | 136 | 1       | 0    | 13       | 1   |
| 17  | 1       | 149        | 0        | 38         | 0         | 262000    | 0.9      |          | 140 | 1       | 0    | 14       | 1   |
| 18  | 0       | 582        | 0        | 14         | 0         | 166000    | 0.8      |          | 127 | 1       | 0    | 14       | 1   |
| 19  | 1       | 125        | 0        | 25         | 1         | 237000    |          | 1        | 140 | 0       | 0    | 15       | 1   |
| ... | ...     | ...        | ...      | ...        | ...       | ...       | ...      | ...      | ... | ...     | ...  | ...      | ... |
| 280 | 1       | 1051       | 1        | 30         | 0         | 232000    | 0.7      |          | 136 | 0       | 0    | 246      | 0   |
| 281 | 0       | 84         | 1        | 38         | 0         | 451000    | 1.3      |          | 136 | 0       | 0    | 246      | 0   |
| 282 | 0       | 2695       | 1        | 40         | 0         | 241000    |          | 1        | 137 | 1       | 0    | 247      | 0   |
| 283 | 0       | 582        | 0        | 40         | 0         | 51000     | 2.7      |          | 136 | 1       | 1    | 250      | 0   |
| 284 | 0       | 64         | 0        | 30         | 0         | 215000    | 3.8      |          | 128 | 1       | 1    | 250      | 0   |
| 285 | 0       | 1688       | 0        | 38         | 0         | 263358.03 | 1.1      |          | 138 | 1       | 1    | 250      | 0   |
| 286 | 1       | 54         | 0        | 40         | 0         | 279000    | 0.8      |          | 141 | 1       | 0    | 250      | 0   |
| 287 | 1       | 170        | 1        | 40         | 0         | 336000    | 1.2      |          | 135 | 1       | 0    | 250      | 0   |
| 288 | 0       | 253        | 0        | 35         | 0         | 279000    | 1.7      |          | 140 | 1       | 0    | 250      | 0   |
| 289 | 0       | 582        | 1        | 55         | 0         | 543000    |          | 1        | 132 | 0       | 0    | 250      | 0   |
| 290 | 0       | 892        | 1        | 35         | 0         | 263358.03 | 1.1      |          | 142 | 0       | 0    | 256      | 0   |
| 291 | 1       | 337        | 0        | 38         | 0         | 390000    | 0.9      |          | 144 | 0       | 0    | 256      | 0   |
| 292 | 0       | 615        | 1        | 55         | 0         | 222000    | 0.8      |          | 141 | 0       | 0    | 257      | 0   |
| 293 | 0       | 320        | 0        | 35         | 0         | 133000    | 1.4      |          | 139 | 1       | 0    | 258      | 0   |
| 294 | 0       | 190        | 1        | 38         | 0         | 382000    |          | 1        | 140 | 1       | 1    | 258      | 0   |
| 295 | 1       | 103        | 1        | 35         | 0         | 179000    | 0.9      |          | 136 | 1       | 1    | 270      | 0   |
| 296 | 0       | 61         | 1        | 38         | 1         | 155000    | 1.1      |          | 143 | 1       | 1    | 270      | 0   |
| 297 | 0       | 1820       | 0        | 38         | 0         | 270000    | 1.2      |          | 139 | 0       | 0    | 271      | 0   |
| 298 | 0       | 2060       | 1        | 60         | 0         | 742000    | 0.8      |          | 138 | 0       | 0    | 278      | 0   |
| 299 | 0       | 2413       | 0        | 38         | 0         | 140000    | 1.4      |          | 140 | 1       | 1    | 280      | 0   |
| 300 | 0       | 196        | 0        | 45         | 0         | 395000    | 1.6      |          | 136 | 1       | 1    | 285      | 0   |