**JUDUL**

**IMPLEMENTASI ALGORITMA KNN UNTUK KLASIFIKASI PENYAKIT HEPATITIS**

**PROPOSAL PROYEK AKHIR**

**Disusun Oleh:**

**Kelompok 5**

1. **FOUNTAIN LALAMENTIK (21013012)**
2. **JUAN KARUNDENG (21013025)\***
3. **JOHN JONG (21013040)**
4. **MARCHELINO RACO (21013047)**

A green and blue logo with a star and a book

Description automatically generated

**DATA MINING DAN WAREHOUSE**

**PROGRAM STUDI TEKNIK INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS KATOLIK DE LA SALLE**

**MANADO**

**2023**

# DAFTAR ISI

[DAFTAR ISI i](#_Toc148551060)

[BAB I 1](#_Toc148551061)

[PENDAHULUAN 1](#_Toc148551062)

[1.1 Latar Belakang 1](#_Toc148551063)

[1.2 Rumusan Masalah 2](#_Toc148551064)

[1.3 Tujuan 2](#_Toc148551065)

[1.4 Manfaat 2](#_Toc148551066)

[1.5 Batasan Masalah 2](#_Toc148551067)

[BAB II 4](#_Toc148551068)

[TINJAUAN PUSTAKA 4](#_Toc148551069)

[2.1 Hepatitis 4](#_Toc148551070)

[2.2 Data Mining 6](#_Toc148551071)

[2.3 Teknik Klasifikasi 6](#_Toc148551072)

[2.4 Algoritma *K-Nearest Neighbour* (KNN) 6](#_Toc148551073)

[2.5 *Tools* 7](#_Toc148551074)

[BAB III 8](#_Toc148551075)

[METODOLOGI PENELITIAN 8](#_Toc148551076)

[3.1 Data 8](#_Toc148551077)

[3.2 Metode Usulan 8](#_Toc148551078)

[3.3 Metode Pembanding 8](#_Toc148551079)

[3.4 Evaluasi Performa 8](#_Toc148551080)

[3.5 Experimental Setup 8](#_Toc148551081)

[DAFTAR PUSTAKA 9](#_Toc148551082)

# BAB I

# PENDAHULUAN

## 1.1 Latar Belakang

Topik yang akan diangkat pada penelitian kali ini adalah implementasi Algoritma *K*-*Nearest Neighbor* (KNN) untuk klasifikasi penyakit hepatitis. Hepatitis merupakan penyakit yang meradang pada hati manusia. Penyakit hepatitis ini memiliki beberapa jenis, yaitu hepatitis A, hepatitis B, hepatitis C, hepatitis D, dan hepatitis E [1]. Berdasarkan Riskesdas, pada tahun 2013 sekitar 18 juta orang di Indonesia menderita penyakit hepatitis B dan sekitar 2,5 juta orang di Indonesia menderita penyakit hepatitis C. Hepatitis B dan C harus menjadi perhatian bagi kita semua, karena hepatits B dan C dapat menyebabkan sironis hati, kanker hati dan bisa menyebabkan kematian [2]. Selain itu, untuk penyakit hepatits B, C, dan D yang bersifat kronis tidak bisa disembuhkan [3]. Oleh karena itu, kita tidak boleh meremehkan penyakit hepatitis yang bisa menyebabkan kematian.

Data mining adalah kegiatan mengumpulkan dan mengolah data, yang nantinya akan diekstrak informasi yang penting dari data tersebut [4]. Teknik klasifikasi pada data mining adalah proses untuk menemukan kesamaan karakteristik dalam suatu kelompok atau kelas [5]. Teknik klasifikasi dapat membantu dalam memprediksikan hasil dari suatu data yang sudah dikumpulkan, misalnya dengan data yang sudah dikumpul, maka kita bisa menggunakan teknik klasifikasi untuk melihat hasil dari data tersebut. Contohnya dengan adanya data-data yang sudah dikumpul, kita bisa lihat hasil akhir dari data tersebut, misalnya kita bisa periksa orang tersebut menderita penyakit tersebut atau tidak, kita bisa lihat orang yang memasukkan data akan hidup atau mati. Algoritma KNN adalah algoritma yang sederhana dan mudah untuk diterapkan dalam menyelesaikan masalah klasifikasi dan regresi, kali ini Algoritma KNN ini akan digunakan untuk menyelesaikan masalah klasifikasi [6]. Algoritma KNN sudah diimplementasikan untuk memprediksi penyakit diabetes [7]. Dan pada penelitian kali ini, kita akan mengimplementasi Algoritma KNN untuk klasifikasi penyakit hepatitis.

## 1.2 Rumusan Masalah

Bagaimana mengimplementasikan Algoritma KNN untuk klasifikasi penyakit hepatitis?

## 1.3 Tujuan

Tujuan penelitian ini adalah mengimplementasikan Algoritma KNN untuk klasifikasi penyakit hepatitis, sehingga hasil dari implementasi tersebut dapat membantu banyak orang dalam mengklasifikasikan penyakit hepatitis.

## 1.4 Manfaat

Manfaat adanya penelitian ini ditunjukan kepada tenaga medis atau masyarakat umum dan pengembang aplikasi.

1. Tenaga medis atau masyarakat umum, dengan adanya implementasi Algoritma KNN, hal ini mempermudah tenaga medis atau masyarakat umum dalam mengklasifikasikan penyakit hepatitis.
2. Pengembang aplikasi, dengan mengimplementasi Algoritma KNN, hal ini sangat berguna bagi pengembang aplikasi dalam memahami penggunaan Algoritma KNN dalam suatu aplikasi.

## 1.5 Batasan Masalah

Batasan masalah dalam penelitian ini, yaitu :

1. Sumber data yang diambil dari website UCI Machine Learning Repository.
2. Dataset yang diambil hanya dataset penyakit hepatitis.
3. Atribut dari dataset hanya 20 atribut.
4. Salah satu atribut, yaitu kelas/*class*. Kelas tersebut hanya memiliki 2 *value*, yaitu *live* (hidup) dan *die* (mati).
5. Penelitian ini hanya membahas tentang penyakit hepatitis, data mining, teknik klasifikasi, Algoritma KNN, dan Algoritma *Naïve-Bayes*.
6. Penelitian ini hanya membuat 1 aplikasi dan hanya menggunakan 1 algoritma, dalam hal ini Algoritma KNN.
7. Algoritma *Naïve-Bayes* hanya akan menjadi pembanding, yang nantinya hanya akan digunakan pada aplikasi Weka sebagai media untuk pembanding.
8. Untuk pembagian data *training* dan data tes, hanya menggunakan metode *split percentage*.
9. Nilai dari pembagian data *training* dan data tes dengan *split percentage*, hanya digunakan 60% dan 40%.
10. Data *training* yang dicoba hanya 60%, tidak ada data *training* untuk 40%.
11. Data tes hanya 40%, tidak ada data tes untuk 60%.

# BAB II

# TINJAUAN PUSTAKA

## 2.1 Hepatitis

Hepatitis adalah penyakit yang meradang organ tubuh hati pada manusia. Penyebab dari penyakit ini terdiri atas beberapa hal, yaitu kebiasaan mengonsumsi alkohol, infeksi virus, pemakaian obat-obat tertentu, penyakit autoimun, dan karena infeksi cacing hati. Jika terkena hepatitis karena disebabkan oleh infeksi virus, maka hepatitis tersebut bisa menular ke orang lain [1].

Gejala dari penyakit hepatitis, yaitu berupa demam, nyeri sendi, sakit perut, mudah Lelah, feses berwarna pucat, urin berwarna gelap, hilangnya nafsu makan, penurunan berat badan, dan penyakit kuning. Jika kondisi ini berlangsung selama 6 bulan, berarti hepatitis akut. Tetapi jika kondisi ini berlangsung lebih dari 6 bulan, berarti hepatitis kronis [1].

Faktor yang dapat meningkatkan risiko seseorang untuk terkena penyakit hepatitis, yaitu tidak mencuci tangan sesudah menggunakan toilet, tidak mencuci tangan sebelum mengolah sebuah makanan, tidak mencuci tangan sebelum makan, mengonsumsi makanan yang tidak matang atau makanan yang sudah terkontaminasi virus hepatitis, kebiasaan berbagi barang pribadi, berhubungan seksual dengan penderita hepatitis, mempunyai lebih dari satu pasangan seksual, melakukan hubungan seksual sesama jenis, menderita penyakit HIV, sering menerima transfusi darah (apalagi jika pendonor tidak menjalani pemeriksaan yang ketat atau alat yang digunakan tidak higienis), mengonsumsi beberapa jenis obat yang dimana obat tersebut mengandung paracetamol atau meminum obat herbal tertentu tanpa konsultasi dengan dokter, bekerja di pusat pengolahan air dan limbah, dan bekerja sebagai tenaga kesehatan [1].

Pada dataset hepatitis yang kami miliki, berikut beberapa variabel yang ada dalam dataset hepatitis tersebut :

1. *AGE*, atribut ini untuk memberi keterangan mengenai usia pasien.
2. *SEX*, atribut ini untuk memberi keterangan mengenai jenis kelamin pasien.
3. *STEROID*, atribut ini untuk memberi keterangan mengenai seorang pasien yang sudah menerima atau sedang menjalani terapi steroid.
4. *ANTIVIRALS*, atribut ini untuk memberi keterangan mengenai seorang pasien yang sudah menerima obat antivirals.
5. *FATIGUE*, atribut ini untuk memberi keterangan kalau seorang pasien mengalami gejala kelelahan.
6. *MALAISE*, atribut ini untuk memberi keterangan kalau seorang pasien mengalami gejala ketidaknyaman secara umum.
7. *ANOREXIA*, atribut ini untuk memberi keterangan kalau seorang pasien kehilangan nafsu makan.
8. *LIVER\_BIG*, atribut ini untuk memberi keterangan ukuran hati pasien (jika hati besar, maka “Yes”).
9. *LIVER\_FIRM*, atribut ini untuk memberi keterangan mengenai hati pasien keras ketika disentuh.
10. *SPLEEN\_PALPABLE*, atribut ini untuk memberi keterangan mengenai limpa pasien dapat diraba.
11. *SPIDERS*, atribut ini untuk memberi keterangan mengenai pembuluh darah pasien yang membesar di kulit seperti spider angiomas.
12. *ASCITES*, atribut ini untuk memberi keterangan mengenai pasien yang memiliki cairan yang terakumulasi dalam rongga perut.
13. *VARICES*, atribut ini untuk memberi keterangan mengenai pasien yang memiliki pembuluh varises.
14. *BILIRUBIN*, atribut ini untuk memberi keterangan mengenai tingkat bilirubin dalam darah, hal ini untuk menjadi indikator fungsi hati.
15. *ALK\_PHOSPHATE*, atribut ini untuk memberi keterangan mengenai tingkat fosfat alkali fostase dalam darah, dimana hal ini juga berhubungan dengan kesehatan hati.
16. *SGOT*, atribut ini untuk mengukur tingkat SGOT (Serum Glutamic Oxaloacetic Transaminase) dalam darah, hal ini menjadi indikator kerusakan hati.
17. *ALBUMIN*, atribut ini untuk memberi keterangan mengenai tingkat albumin darah, yang dimana ini adalah protein yang diproduksi oleh hati.
18. *PROTIME*, atribut ini untuk mengukur protrombin time, yang dimana hal ini di tes dengan seberapa cepat darah membeku.
19. *HISTOLOGY*, atribut ini untuk memberi keterangan mengenai seorang pasien sudah mengikuti pemeriksaan histologi hati.
20. *CLASS*, atribut ini untuk memberi keterangan mengenai seorang pasien hidup ada mati [8].

## 2.2 Data Mining

Data mining adalah kegiatan mengumpulkan dan mengolah data, yang nantinya akan diekstrak informasi yang penting dari data tersebut. Tujuan dari data mining, yaitu sebagai sarana menjelaskan, konfirmasi, dan eksplorasi. Metode data mining juga memiliki beberapa metode, yaitu asosiasi, regresi, klasifikasi, dan *clustering* [4]*.*

## 2.3 Teknik Klasifikasi

Teknik klasifikasi pada data mining adalah proses untuk menemukan kesamaan karakteristik dalam suatu kelompok atau kelas. Teknik klasifikasi bertujuan untuk memberi kelas dari objek tertentu, yang dimana labelnya belum diketahui. Ada beberapa metode yang digunakan untuk teknik klasifikasi, yaitu *Logistic Regression*, *Naïve-Bayes*, *Decision Tree*, *Random Forest*, *K-Nearest Neighbour*, dan *Artificial Neural Network* [5].

## 2.4 Algoritma *K-Nearest Neighbour* (KNN)

Klasifikasi adalah salah satu metode yang digunakan pada data mining. Klasifikasi merupakan masalah yang penting pada bidang data *science* dan *machine learning*. Algoritma KNN ini merupakan salah satu algoritma tertua, tetapi memiliki tingkat akurasi yang tinggi. Algoritma KNN merupakan algoritma yang sederhana dan mudah untuk diterapkan dalam menyelesaikan masalah klasifikasi. Tujuan utama dari Algoritma KNN, yaitu memberikan identifikasi pada tetangga terdekat, berdasarkan kueri awal yang diberikan. Algoritma KNN memiliki prinsip, dimana titik data yang berdekatan dengan data lainnya, mereka berada pada kelas yang sama [6].

Algoritma KNN memiliki kelebihan dan kekurangan. Kelebihannya mudah bagi kita untuk menerapkan Algoritma KNN, mudah untuk beradaptasi , dan hanya memiliki sedikit hyperparameter (nilai k dan nilai metrik jarak) [6].

Sedangkan kelemahan Algoritma KNN, yaitu algoritma ini tidak berfungsi jika digunakan pada dataset yang memiliki ukuran besar, kurang disarankan untuk digunakan pada dimensi tinggi, harus memberikan skala fitur, dan sangat sensitif pada nilai yang hilang dari data [6].

## 2.5 *Tools*

*Tools* yang digunakan yaitu bahasa pemrograman *Python*. Untuk media pembuatan *coding*, menggunakan aplikasi Visual Studio Code. Dan untuk *library* yang akan digunakan, yaitu *NumPy*, *Pandas*, dan *Scikit-Learn* (sklearn) [9].

# BAB III

# METODOLOGI PENELITIAN

## 3.1 Data

Data yang digunakan adalah dataset tentang hepatitis yang bersumber dari UCI Machine Learning Repository. Dataset tersebut memiliki 20 atribut dan 155 records [8].

## 3.2 Metode Usulan

Metode usulan yang digunakan, yaitu Algoritma KNN. Dengan parameter yang digunakan, yaitu K = 3. Alasan tidak memakai Algoritma *Naïve-Bayes* pada metode usulan, ini dikarenakan Algoritma *Naïve-Bayes* bersifat asumsi independensi, yang dimana hal ini tidak cocok untuk dataset yang bersifat medis.

## 3.3 Metode Pembanding

Metode pembanding yang digunakan, yaitu Algoritma *Naïve-Bayes*. Dengan parameter yang digunakan, yaitu parameter standar dari aplikasi Weka.

## 3.4 Evaluasi Performa

Indikator yang dilihat untuk dibandingkan dari kedua metode tersebut, yaitu akurasi dan waktu.

## 3.5 Experimental Setup

Untuk pembagian data *training* dan data test akan digunakan *split percentage*, dengan nilai 60% : 40%. Jadi untuk data *training* sebesar 60% dan data tes sebesar 40%.

# DAFTAR PUSTAKA

|  |  |
| --- | --- |
| [1] | dr.Pittara, "Hepatitis," 9 Mei 2022. [Online]. Available: https://www.alodokter.com/hepatitis. [Accessed 17 Oktober 2023]. |
| [2] | P. K. RI, "Hepatitis Can't Wait," 18 Agustus 2022. [Online]. Available: http://p2p.kemkes.go.id/2022/08/18/. [Accessed 18 10 2023]. |
| [3] | d. M. D. C. Pane, "Apakah Hepatitis Bisa Sembuh?," 6 Juli 2023. [Online]. Available: https://www.alodokter.com/apakah-hepatitis-bisa-sembuh#:~:text=Agar%20lekas%20sembuh%2C%20dokter%20biasanya,atau%20kurang%20dari%206%20bulan. [Accessed 18 Oktober 2023]. |
| [4] | R. Setiawan, "Apa itu Data Mining dan Bagaimana Metodenya?," 30 Oktober 2021. [Online]. Available: https://www.dicoding.com/blog/apa-itu-data-mining/. [Accessed 18 Oktober 2023]. |
| [5] | N. H. Amrullah, "Metode Klasifikasi Dalam Data Mining," 20 Juni 2022. [Online]. Available: https://www.kompasiana.com/najahamrullah1322/62b008fe38350056ba000382/metode-klasifikasi-dalam-data-mining. [Accessed 18 Oktober 2023]. |
| [6] | ADMINLP2M, "Algoritma K-Nearest Neighbors (KNN) – Pengertian dan Penerapan," 16 Februari 2023. [Online]. Available: https://lp2m.uma.ac.id/2023/02/16/algoritma-k-nearest-neighbors-knn-pengertian-dan-penerapan/. [Accessed 18 Oktober 2023]. |
| [7] | M. a. F. N.Azizah, "Penerapan Algoritma Klasifikasi K-Nearest Neighbor pada Penyakit Diabetes," 7 Agustus 2023. [Online]. Available: https://prosidingsnsa.statistics.unpad.ac.id/?journal=prosidingsnsa&page=article&op=view&path%5B%5D=344. [Accessed 18 Oktober 2023]. |
| [8] | R. M.Kelly, "Hepatitis," 31 Oktober 1998. [Online]. Available: https://archive.ics.uci.edu/dataset/46/hepatitis. [Accessed 18 Oktober 2023]. |
| [9] | R. Sudewo, "5 Rekomendasi Library Python yang Paling Populer Digunakan," 1 November 2022. [Online]. Available: https://danacita.co.id/blog/5-rekomendasi-library-python-yang-paling-populer/. [Accessed 18 Oktober 2023]. |