

**IMPLEMENTASI METODE NAÏVE BAYES UNTUK
MENENTUKAN PASIEN PENDERITA KANKER PAYUDARA**

PROPOSAL TUGAS AKHIR



Diajukan Oleh :

Raden Tio Putra Sudewo

8020190235

Untuk Persyaratan Penelitian Dan Penulisan Tugas Akhir

Sebagai Akhir Proses Studi Strata 1

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS DINAMIKA BANGSA
2022**

PERNYATAAN HASIL EVALUASI

NAMA DAN NIM : Raden Tio Putra Sudewo (8020190235)

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

JUDUL : Implementasi Metode Naïve Bayes Untuk Menentukan
Pasien Penderita Kanker Payudara

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

1. Catatan

- Alasan penolakan proyek penelitian
- Proyek penelitian tidak relevan dengan program studi
- Pernah ada topik sejenis
- Metode yang dipakai tidak jelas
- Masalah terlalu sempit
-

Mengetahui
Ketua Program Studi

Beny, S.Kom, M.Sc
YDB.07.84.043

*) Coret yang tidak perlu.

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal : Implementasi Metode Naïve Bayes Untuk Menentukan Pasien Penderita Kanker Payudara

Program studi : Teknik Informatika

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

- a. Nama Lengkap : Raden Tio Putra Sudewo
- b. Nim : 8020190235
- c. Jenis Kelamin : Laki-Laki
- d. Tempat/Tgl.Lahir : Kuala Tungkal, 25 Mei 2001
- e. Alamat : Jl.Peltu Simatupang, Lrg.Mawar,
Rt.26, Kel.Pal 5, Kec.Kota Baru
- f. No. Telepon : 0853-5758-9403
- g. Email : radentio88@gmail.com

1.1 LATAR BELAKANG MASALAH

Kemajuan teknologi saat ini tidak terlepas dari kehidupan masyarakat. Berbagai informasi bermunculan di berbagai belahan dunia yang kini dapat kita ketahui secara instan berkat kemajuan teknologi (globalisasi). Tentunya kemajuan teknologi tersebut telah menyebabkan perubahan yang luar biasa dalam kehidupan masyarakat dengan segala peradaban dan budayanya. Perubahan ini juga sangat mempengaruhi transformasi nilai-nilai yang ada di masyarakat.[1]

Kemajuan ilmu dan teknologi adalah salah satu hasil produktivitas dari manusia yang memiliki sebuah implementasi pengetahuan yang didapat dari pendidikan. Dimana perkembangan ilmu pengetahuan dan teknologi memiliki implikasi yang luas dalam kehidupan manusia sehingga diharapkan manusia-manusia tersebut perlu mendalami dan memahami penerapan tersebut untuk mengambil manfaatnya secara optimal dan mereduksi implikasi negatif yang ada. [2]

Hampir semua kalangan masyarakat memperoleh begitu saja teknologi digital modern yang bahkan tidak dapat diakses beberapa dekade silam. Inovasi-inovasi dan perkembangan teknologi ini, yang telah lama ditandai sebagai hal sepele, tidak penting, atau bahkan berbahaya selama masa tenang kini menjadi sangat diperlukan dalam menghadapi pandemi yang jelas akan sangat membantu dalam proses implementasinya. [3]

Dari pernyataan di atas dapat disimpulkan bahwa kemajuan teknologi sudah menjadi tolok ukur dalam kehidupan masyarakat yang mana hal tersebut tidak terlepas dalam penerapan dalam kegiatan-kegiatan masyarakat. Selain dari itu penerapan teknologi dalam kehidupan masyarakat sudah memiliki jangkauan yang luas termasuk aspek yang tertera seperti aspek pendidikan, bisnis, pertanian dan kesehatan. Tetapi dapat di pahami bahwa perkembangan teknologi tidak sepenuhnya memiliki dampak yang baik, maka dari itu di perlukan filterisasi untuk mengambil dampak ataupun manfaat-manfaat yang baik dan memiliki prospek yang akan di gunakan kedepannya.

Seperti halnya perkembangan teknologi dalam aspek kesehatan, yang mana dengan penerapan teknologi informasi mengenai kesehatan dapat memiliki *range* dan target yang jelas. Karena dengan penerapan tersebut, informasi-informasi yang terdengar tabu pun akan bisa di jelaskan dengan logika dan memiliki ketepatan informasi yang tinggi. Maka dari itu penerepan teknologi dalam bidang Kesehatan ini merupakan suatu titik balik dari sebuah peradaban yang maju. Seperti halnya salah satu penyakit mematikan yaitu kanker.

Kanker adalah penyakit yang disebabkan oleh pertumbuhan sel-sel abnormal yang tidak terkendali di dalam tubuh. Pertumbuhan sel yang tidak normal ini dapat merusak sel-sel normal di sekitarnya dan di bagian tubuh lainnya. Kanker adalah penyebab kematian nomor dua di dunia. Kanker seringkali berakibat fatal, karena seringkali tidak menimbulkan gejala pada tahap awal perkembangannya. Oleh karena itu, kondisi ini baru terdeteksi dan diobati bila sudah berkembang ke stadium yang parah. Seperti yang kita ketahui bahwa kanker memiliki beberapa jenis, secara umum diketahui seperti kanker tulang, kanker kulit, kanker otak, kanker payudara dan sebagainya. Sedangkan,

Pasien kanker payudara masih dipahami sebagai kondisi manusia yang kompleks, terdiri dari kendala fisik, serta aspek lain seperti psikologis, sosial budaya, dan spiritual. Jadi perjalanan hidup sebagai pasien kanker payudara dan melalui penderitaan selalu dikaitkan dengan faktor fisik, psikologis, sosial budaya dan spiritual. [4]

Kanker payudara adalah jenis kanker yang paling umum terjadi pada Wanita. Menurut Kementerian Kesehatan (Kemenkes) tahun 2014 jenis kanker terbanyak di Indonesia berdasarkan jenis kelamin adalah kanker payudara dan kanker serviks pada wanita, Pada tahun 2010 jumlah pasien rawat jalan dan rawat inap pasien kanker payudara sebanyak 12.014 (28,7%) dan kanker serviks 5.349 (12,8%) pasien. Data rekam medis RSUPNCM pada tahun 2013 menunjukkan kasus kanker payudara sebanyak 2.092 dan pada tahun 2014 sebanyak 1815 kasus. Kanker payudara adalah tumor ganas yang menyerang jaringan payudara. Ciri-ciri kanker payudara adalah nyeri pada payudara, benjolan pada payudara yang lebih

besar, kulit pada payudara menjadi kering seperti kulit jeruk, dan terkadang keluar cairan atau darah dari puting susu. [5]

Maka dari itu dengan menerapkan suatu pemahaman tentang penerapan metode yang tepat guna menentukan penderita kanker payudara yang mana dalam menentukan klasifikasi tersebut diperlukan sebuah metode yang tepat pula yaitu *naïve bayes*. Sehingga korelasi antara teknologi terhadap bidang Kesehatan terutama dalam penderita kanker payudara dapat di implementasikan dengan baik dengan menggunakan metode *naïve bayes*. Yang mana penerapan dari metode tersebut dapat digunakan sebagai acuan dalam mengambil sebuah keputusan terkait yaitu dalam bidang Kesehatan (penderita kanker payudara).

Pemanfaatan algoritma *naïve bayes* dalam klasifikasi kanker payudara memperoleh nilai yang sangat baik, yang mana rata-rata persentase data yang telah terklasifikasi dengan benar mencapai 96.9% sedangkan rata-rata persentase data terklasifikasi salah hanya 3.1%. [6]

Dengan bebarapa argument diatas, maka penulis menentukan judul penelitian yang telah dipahami dan didiskusikan dengan beberapa pihak. Maka dari itu penulis mengangkat judul penelitian :

“IMPLEMENTASI METODE NAÏVE BAYES UNTUK MENENTUKAN PASIEN PENDERITA KANKER PAYUDARA”

Dengan adanya penelitian ini di harapkan mampu untuk menentukan keputusan terhadap penderita kanker payudara dengan menggunakan metode *naïve bayes*.

1.2 RUMUSAN MASALAH

Masalah yang di bahas oleh penulis dalam penelitian ini meliputi :

1. Menentukan keputusan terkait terhadap penderita kanker payudara dengan akurasi yang tepat dan persentase yang benar.
2. Menerapkan suatu metode yang tepat serta terstruktur yaitu *naïve bayes* dalam menentukan penderita kanker payudara

3. Memaksimalkan hasil analisis sebagai keputusan yang dapat digunakan sebagai landasan dan dapat dipertanggung jawabkan.

1.3 BATASAN MASALAH

Batasan masalah yang digunakan dalam pembahasan penelitian ini adalah untuk tujuan membuat interpretasi atau elaborasi yang berorientasi rinci dan konsisten dengan tujuan penulis, oleh karena itu penulis menetapkan batasan masalah, yaitu :

1. Objek yang menjadi prioritas dalam penelitian ini adalah penderita kanker payudara
2. Metode yang digunakan dalam penelitian ini adalah *naïve bayes*
3. Keluaran yang dihasilkan berupa “meninggal” atau “hidup” seorang penderita dalam penelitian terkait penderita kanker payudara
4. Pengujian hasil analisis menggunakan aplikasi *WEKA*
5. Dataset dalam penelitian ini adalah dataset publik yang diperoleh melalui website *Kaggle.com* pada tahun 2022

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan penelitian

Tujuan yang akan disampaikan penulis dalam penelitian ini adalah :

1. Untuk mengetahui persentase penderita kanker payudara dengan menggunakan *naïve bayes*.
2. Memberikan suatu landasan dalam menerapkan sebuah metode untuk mengetahui penderita kanker payudara yang valid dan dapat dipertanggung jawabkan.

1.4.2 Manfaat Penelitian

Manfaat yang akan disampaikan penulis dalam penelitian ini adalah :

1. Dengan adanya pemanfaatan algoritma tersebut dapat membantu beberapa pihak dalam menentukan ataupun sebagai landasan dalam penerapan penderita kanker payudara

2. Menentukan sebuah keputusan dengan tingkat akurasi yang tinggi bagi penderita kanker payudara
3. Memperoleh akurasi yang tepat dengan menggunakan penerapan dari algoritma *naïve bayes*

1.5 SISTEMATIKA PENULISAN

Untuk mempermudah dalam memahami penulisan proposal tugas akhir ini, maka penulis menampilkan sistematika penulisan penelitian ini sebagai berikut :

BAB I : PENDAHULUAN

Dalam bab pendahuluan ini terdiri atas latar belakang, perumusan masalah, batasan masalah, tujuan dan manfaat penelitian, dan sistematika penulisan.

BAB II : LANDASAN TEORI

Dalam bab landasan teori ini memuat konsep-konsep teoritis yang digunakan sebagai kerangka atau landasan yang digunakan untuk mendukung pemahaman terhadap penelitian yang penulis lakukan.

BAB III : METODOLOGI PENELITIAN

Dalam bab metodologi penelitian ini berisi mengenai parameter penelitian, metode penelitian yang digunakan, dan teknik pengumpulan data penelitian.

BAB IV : JADWAL PENELITIAN

Bagian ini berisi rencana aktivitas dan jadwalnya selama 6 bulan pengerjaan tugas akhir.

2.1 LANDASAN TEORI

Data mining didefinisikan sebagai seperangkat prosedur yang berguna untuk menemukan dan mencari nilai dalam bentuk informasi serta hubungan kompleks yang telah disimpan dalam database hingga saat ini. Ekstraksi sampel informasi data yang berguna untuk memanipulasi data menjadi informasi baru dan lebih berguna yang ditujukan untuk mengekstrak sampel berharga atau menarik yang diperoleh dari data database. [7]

Data mining adalah analisis yang memeriksa kumpulan data untuk menemukan hubungan yang tidak terduga dan merangkum data dengan cara yang berbeda dengan cara yang berbeda dari sebelumnya, yang mudah dipahami dan berguna bagi pemilik data. [8]

Data mining adalah proses menemukan hubungan baru yang bermakna, pola dan kebiasaan dengan menyortir sebagian besar data yang disimpan dalam media penyimpanan menggunakan teknologi pengenalan pola, seperti teknik statistik dan matematika. [9]

Sehingga dapat di simpulkan bahwa data mining merupakan suatu penerapan ilmu yang ditujukan untuk memprediksi, klasifikasi ataupun digunakan sebagai landasan dalam mengambil sebuah keputusan yang mana dalam implementasiannya data mining menjadi syarat dalam sebuah pengambilan keputusan bagi sebuah instansi sehingga dari itu data mining merupakan cabang ilmu yang memiliki prospek yang panjang dan akurat.

2.2 PENGOLAHAN DATA MINING

Dalam penerapannya, data mining memiliki beberapa tahapan dalam pengolahannya yang mana dari beberapa tahapan tersebut memiliki fungsi dalam penyelesaian atau pengolahan data terkait.

Sebutkan bahwa KDD atau Pengetahuan data mining dari Data, adapun tahapan dalam pengolahan data tersebut yaitu : [10]

1. *Data Cleaning*

Proses pembersihan data dari noise dan data yang tidak konsisten sehingga dalam tahap ini pengolahan bisa menjadi optimal dan memiliki hasil yang optimal.

2. *Data Integration*

Proses tersebut menggabungkan data dari sejumlah sumber yang berbeda. Yang memiliki asumsi bahwa perbandingan atau penambahan data menjadi landasan dalam pengolahan.

3. *Data Selection*

Proses pemilihan data dari database sangat tepat untuk keperluan analisis. Yang mana dengan tahapan ini merupakan tahap yang vital karena merupakan landasan dalam penelitian ini.

4. *Data Transformation*

Proses mengubah data menjadi data yang tepat untuk proses pengolahan. Proses konversi data kedalam format tertentu sehingga data tepat sasaran untuk data mining dan pengolahannya.

5. *Data Mining*

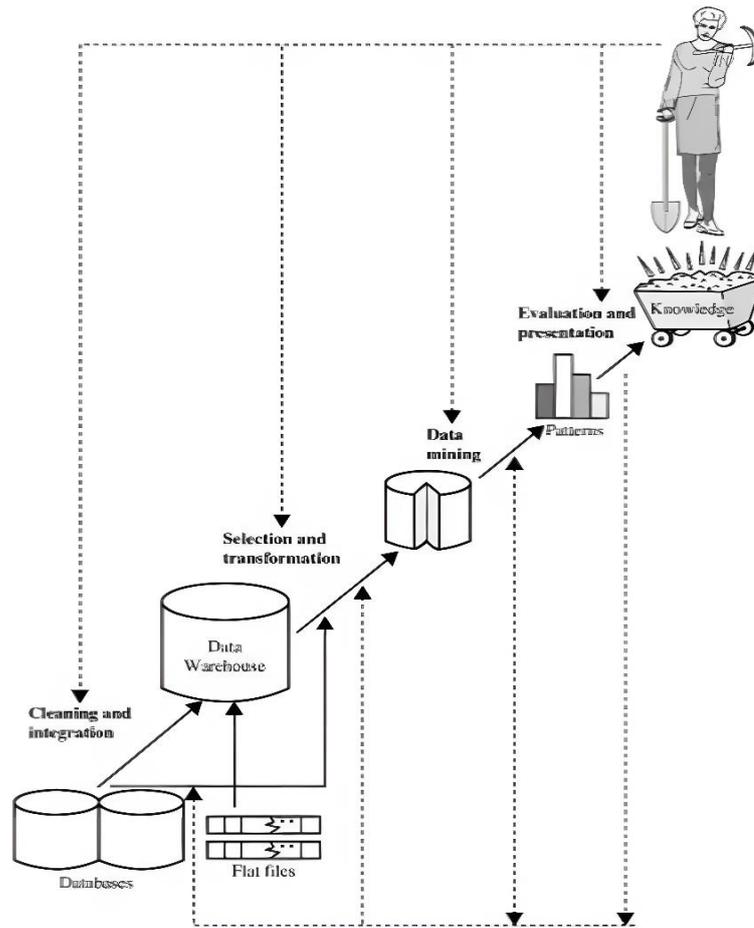
Proses utama yang menjadi dasar dalam pengolahan data dengan menggunakan sebuah metode tertentu untuk memperoleh sebuah pola dari data.

6. *Pattern Evaluation*

Proses mengidentifikasi pola. Pada tahap ini, membuat model unik untuk menilai apakah prediksi menjadi kenyataan.

7. *Knowledge presentation*

Pada tahap ini proses penelitian menjadi tujuan akhir. dimana dapat mewakili informasi yang diperlukan, prosedur dimana informasi tersebut selanjutnya digunakan oleh pemilik data.



Gambar 2. 1 Pengolahan Data Mining

2.3 KANKER PAYUDARA

Kanker payudara merupakan kanker paling mematikan setelah kanker serviks. sel kanker payudara pertama dapat tumbuh menjadi tumor 1cm dalam 8-12 tahun, Sel kanker ini berada di kelenjar susu. Sel kanker payudara ini dapat menyebar melalui aliran darah ke seluruh tubuh. Hasil penelitian menunjukkan masih banyak masyarakat yang mengetahui definisi kanker payudara sampai 68 orang (68%), faktor risiko kanker payudara yaitu 57 orang (57%), gejala mengenai kanker payudara sebanyak 64 orang (64%), pencegahan kanker payudara 74 orang (74%), dan Pemeriksaan Payudara Sendiri (SADARI) sebanyak 55 responden (55%). [11]

2.4 NAÏVE BAYES

Algoritma Naïve Bayes merupakan salah satu dari algoritma yang termasuk dalam pengklasifikasi. Naïve bayes adalah pengklasifikasi dengan probabilitas sederhana dan model statistik berdasarkan teorema Bayes dengan dengan asumsi bahwa setiap atribut adalah independent. [12]

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)}$$

Dimana persamaan Teori Bayes Tersebut adalah sebagai berikut :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

Untuk menjelaskan metode *Naive Bayes*, perlu diperhatikan bahwa proses klasifikasi memerlukan beberapa panduan untuk menentukan kelas mana yang sesuai untuk sampel yang akan dianalisis.

2.5 WEKA

WEKA adalah rangkaian perangkat lunak pembelajaran mesin yang ditulis dalam Java, dikembangkan di Universitas Waikato, Selandia Baru. Perangkat lunak ini memiliki banyak algoritma pembelajaran mesin untuk tujuan penambangan data. *WEKA* juga memiliki tools untuk pengolahan data, mulai dari preprocessing, klasifikasi, association rules, dan visualization. [13]

WEKA berisi kumpulan algoritma dan bentuk visualisasi dari analisis data dan pemodelan prediktif. *WEKA* dapat digunakan untuk menganalisis data dari

banyak bidang, terutama untuk pendidikan dan penelitian. WEKA mengimplementasikan teknik klasifikasi dan regresi.



Gambar 2. 2 Tampilan Utama WEKA

WEKA menyediakan tiga *Graphical User Interface* untuk memproses data. Layar pertama adalah *explorer* yang memberikan akses ke semua fungsi WEKA melalui menu yang mudah digunakan dan dapat diurutkan. Fungsi yang dapat digunakan melalui penjelajah diantaranya :

1. *Propocessing*, untuk memilih data yang akan diproses.
2. *Classifier*, untuk menggunakan klasifikasi atau algoritma regresi pada data yang dimasukkan dalam tabel *preprocessing*.
3. *Cluster*, menyediakan akses ke metode *Cluster* yang disediakan oleh WEKA.
4. *Associate*, menyediakan algoritma untuk *association rule* yang berguna untuk mengenali relasi antar item pada data.
5. *Select attribute*, menyediakan berbagai algoritma untuk memilih atribut dan variabel yang paling relevan.
6. *Experimenter*, adalah *GUI Weka* dirancang untuk pengujian. Tes yang dimaksud adalah membandingkan kinerja algoritma klasifikasi yang tersedia pada WEKA dengan data yang berbeda. *Interface* yang disediakan *KnowledgeFlow* mirip dengan *Explorer* menggunakan berbagai algoritma

data mining yang berbeda yang telah disediakan di *Weka* dengan data yang dimasukkan pengguna.

Tahapan atau proses penggunaan *Weka* adalah sebagai berikut :

1. Mempersiapkan file data *training* dan data *testing* dalam format *.arff*
2. Buka aplikasi *Weka* kemudian pilih menu *Explorer*
3. Setelah menu *explorer* terbuka maka langkah selanjutnya yaitu memilih file data *training* dengan klik tab menu *open file* lalu pilih data yang akan diinputkan kemudian klik *open*
4. Setelah data *training* diinputkan selanjutnya klik tab metode dan algoritma yang ingin digunakan.
5. Setelah memilih metode yang digunakan selanjutnya menginputkan data *testing* .
6. Menyimpan data hasil pengolahan menggunakan *Weka* dengan format *arff*.

2.6 PENELITIAN SEJENIS

Penulis memulai penelitian ini dengan terlebih dahulu melakukan penelusuran literatur dari penelitian dan sumber lain. Beberapa penelitian telah dilakukan dengan menggunakan teknik data mining untuk mengungkapkan informasi yang berbeda dari database terkait. Penelitian mencakup topik yang relevan dengan penelitian penulis, termasuk studi tentang algoritma yang akan penulis gunakan.

No	Judul	Masalah	Metode	Tujuan	Hasil
1.	Analisis performa metode KNN pada Dataset pasien pengidap Kanker Payudara [14]	Penelitian ini dilakukan dengan menerapkan metode knn pada dataset pasien pengidap penyakit kanker payudara. Dengan menggunakan dataset terkait sehingga bisa dianalisis dengan tepat dan akurat	KNN	Untuk mengetahui Informasi penderita kanker payudara yang dapat mempengaruhi sejumlah keputusan kurikulum menggunakan algoritma KNN	Dari perhitungan ini, dengan 569 data yang di bagi menjadi 20% training dan 80% testing dengan $K = 3,4$ dan 5 mendapat nilai tertinggi untuk akurasi adalah $0,93$ pada 20% keempat($K3$), 20% Pertama($K4$) dan 20% pertama($K5$), untuk Presisi dengan nilai $0,97$ pada 20% keempat($K3$), untuk Recall dengan nilai $0,98$ pada 20% ketiga($K3$) dan F-measure dengan nilai $0,94$ pada 20%

					keempat(K3) dan 20% ketiga(K5).
2.	Hasil Analisis Teknik Data Mining dengan Metode Naive Bayes untuk Mendiagnosa Penyakit Kanker Payudara [15]	Dalam penelitian ini menerapkan korelasi antara penderita dengan sebuah metode terkait guna menghasilkan keputusan valid dalam penyampaian informasi yang menjadi latar dalam penelitian ini.	<i>Naïve Bayes</i>	Memberikan sebuah informasi tentang mendiagnosa penderita kanker payudara	Berdasarkan uraian di atas pada pembahasan sebelumnya, dapat disimpulkan bahwa klasifikasi kekambuhan pasien kanker payudara menggunakan algoritma Naïve Bayes adalah 72,7%. Proses pembebanan berdampak baik pada algoritma Naïve Bayes yaitu menambah nilai akurasi 1,02%
3.	Analisis Komparasi Algoritma Klasifikasi Data	Penyakit jantung adalah istilah umum untuk semua jenis gangguan yang	C.45	Penelitian ini bertujuan untuk membuat suatu keputusan dalam	Hasil penelitian mendapatkan nilai akurasi sebesar tertinggi

	Mining Untuk Prediksi Penderita Penyakit Jantung [16]	mempengaruhi jantung. Penyakit jantung memiliki arti yang sama dengan penyakit jantung, tetapi bukan penyakit kardiovaskular. Penelitian ini akan membandingkan beberapa algoritma klasifikasi.		pendektesian penyakit jantung dengan menggunakan beberapa algoritma salah satunya adalah C.45	sebesar 80.38%. dalam penerapan dengan menggunakan metode terkait.
--	--	---	--	--	---

Tabel 2. 1 Penelitian Sejenis

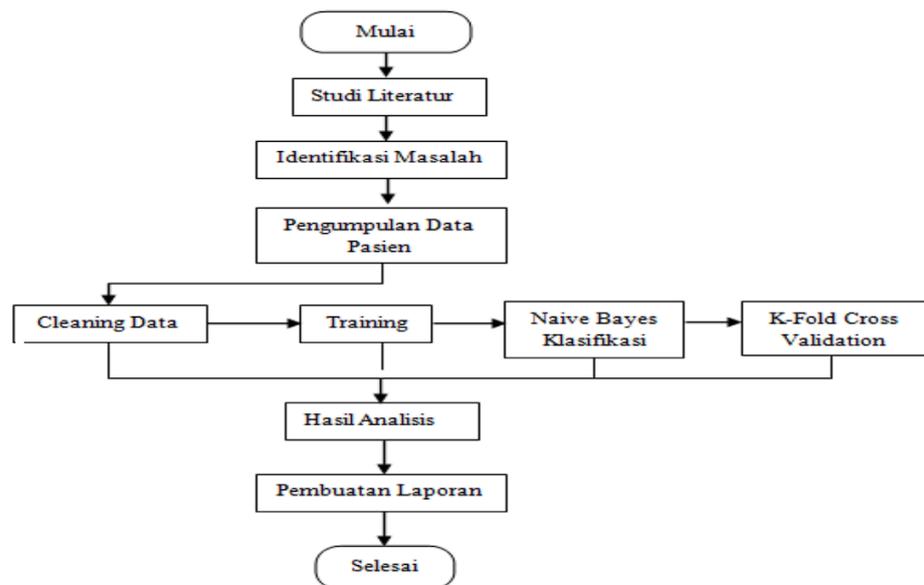
Dari beberapa penelitian di atas, dapat ditarik kesimpulan bahwa indikasi dari penggunaan data mining dalam memprediksi, klasifikasi ataupun pengambilan sebuah keputusan memiliki kekurangan dan kelebihan masing-masing. Yang mana dari beberapa metode di atas dapat diterapkan pada data yang sesuai sehingga hasil yang diperoleh pun juga berbeda dengan hasil dari metode yang lain yang dilatari dengan jumlah data, metode, dan lingkup penelitian.

3.1 METODOLOGI PENELITIAN

Metodologi penelitian berguna sebagai fungsi penting untuk mengumpulkan informasi/data yang dibutuhkan untuk memecahkan berbagai jenis masalah dan berguna untuk memberikan solusi atas masalah yang ada. [7]

Kerangka penelitian mencakup langkah-langkah yang diambil dalam pelaksanaan penelitian. Kerangka penelitian dirancang untuk memudahkan kita menerima hasil penelitian, sehingga dapat menyelesaikan penelitian tepat waktu, dan melakukan penelitian sesuai rencana. [17]

Sehingga, penulis memukakan tahapan-tahapan (dalam bentuk diagram) yang akan diterapkan dalam penelitian ini, yang mana akan dijabarkan pada gambar 3.1



Gambar 3. 1 Kerangka Kerja Penelitian

1. Studi Literatur

Di tahap ini penulis melakukan pencarian terhadap landasan-landasan teori atau materi yang telah diperoleh dari berbagai sumber seperti buku, jurnal ilmiah dan juga referensi lainnya untuk melengkapi dalam penelitian baik mengenai konsep dan teori sehingga memiliki landasan yang tepat dan relevan.

2. Identifikasi Masalah

Pada tahap ini, penulis melakukan identifikasi terkait masalah yang diangkat pada penelitian ini yaitu mengenai penderita kanker payudara sehingga peneliti mampu menerapkan sebuah metode yang tepat guna terselesaikannya masalah terkait.

3. Pengumpulan Data

Pada tahap pengumpulan data, penulis melakukan beberapa metode yang akan digunakan dalam pengumpulan data yang akan dijadikan dasar dalam penelitian ini, dan berikut beberapa metode yang akan di gunakan dalam penelitian ini, yaitu :

1. Pengamatan (*Observasi*)

Pengamatan yang dilakukan oleh penulis adalah dengan melakukan beberapa pengamatan pada data terkait yaitu mengenai data pasien penderita kanker payudara yang mana dataset diperoleh dari website (*Kaggle.com*) yang menjadikan data tersebut sebagai data publik yang akan di teliti dalam penulisan ini.

2. Penelitian Kepustakaan (*Library Research*)

Untuk mendukung penelitian yang akan dilakukan, penulis melakukan penelusuran kepustakaan yaitu dengan mencari data dari buku-buku penelitian sejenis dan jurnal yang berkaitan dengan dengan metode yang digunakan oleh penulis yang digunakan dalam penelitian ini yaitu metode *Naïve bayes*.

4. *Cleaning Data*

Pembersihan data adalah pra-pemrosesan, pada tahap ini data dibersihkan atau dipisahkan dari data yang tidak bisa di olah dan data yang tidak konsisten, seperti data pasien dengan atribut yang tidak lengkap.

5. *Training*

Pada bagian ini dilakukan proses training menggunakan algoritma klasifikasi *naïve bayes* untuk membuat sebuah bentuk *classifier* pada data pasien yang sudah ada. bentuk ini merupakan implementasi pengetahuan yang akan digunakan dalam penelitian ini.

6. *Naïve bayes* Klasifikasi

Pada tahap ini dilakukan klasifikasi data menggunakan alat *WEKA* pada data pasien yang diuji menggunakan model klasifikasi yang telah dibuat.

7. *K-fold Cross Validation*

Validasi *K-fold Cross* adalah teknik untuk mengevaluasi/mengkonfirmasi keakuratan model yang dibangun terhadap kumpulan data tertentu. Salah satu metode validasi silang yang paling populer adalah *K-fold Cross Validation*

8. Hasil Analisis

Hasil analisis akan dievaluasi keakuratan data model yang diperoleh dengan menghitung *naïve bayes* menggunakan *K-fold Cross Validation*.

9. Penulisan Laporan

Setelah semua langkah penelitian selesai, laporan studi akan didokumentasikan sehingga dapat digunakan di masa mendatang. Baik oleh peneliti itu sendiri maupun oleh peneliti lainnya

3.2 ALAT PENELITIAN

Dalam melakukan penelitian untuk menentukan penderita kanker payudara, peneliti memerlukan alat bantu penelitian (*tools*). Adapun alat bantu yang dibutuhkan adalah sebagai berikut

1. Perangkat Keras (*hardware*)
 - a. Laptop ASUS, *processor* AMD RYZEN 5 3550H-2.1 GHz, RAM 8GB HDD 1000GB.
 - b. Printer
 - c. Dan beberapa perangkat keras pendukung lainnya.
2. Perangkat Lunak (*software*)
 - a. Sistem Operasi Windows 10 Pro 64-bit
 - b. WEKA
 - c. Notepad++
 - d. Dan beberapa perangkat keras pendukung lainnya.

3.3 BAHAN PENELITIAN

Dalam penelitian ini, bahan yang digunakan adalah dataset yang diperoleh dari website (*Kaggle.com*) yang mana dataset tersebut bersifat dataset publik dan data tersebut adalah data terbaru yang diterbitkan di tahun 2022 pada website terkait. Dan penelitian ini dilakukan dengan landasan data yang memuat penderita kanker payudara sehingga dengan adanya penelitian ini diharapkan mampu dijadikan sebagai acuan dalam menentukan ataupun mengambil sebuah keputusan.

4.1 JADWAL PENELITIAN

Dan berikut ini merupakan jadwal penelitian yang akan dilaksanakan oleh penulis selama pengerjaan penelitian ini :

No	Kegiatan	September				Oktober				November				Desember			
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
1	Penyusunan Proposal																
2	Pengumpulan Data																
3	Analisis																
4	Pembuatan Data Uji																
5	Pengujian																
6	Pengumpulan Laporan																

Tabel 4. 1 Jadwal Penelitian

DAFTAR PUSTAKA

- [1] H. S. Wahyudi and M. P. Sukmasari, "Teknologi Dan Kehidupan Masyarakat," *J. Anal. Sociol.*, vol. 3, no. 1, 2018, doi: 10.20961/jas.v3i1.17444.
- [2] Nurdyansyah, "Sumber Daya dalam Teknologi Pendidikan," *Univ. Negeri Surabaya*, pp. 1–22, 2017.
- [3] R. Komalasari, "Manfaat Teknologi Informasi Dan Komunikasi Di Masa Pandemi Covid 19," *Tematik*, vol. 7, no. 1, pp. 38–50, 2020, doi: 10.38204/tematik.v7i1.369.
- [4] L. Rahayuwati, K. Ibrahim, and M. Komariah, "Pilihan Pengobatan Pasien Kanker Payudara Masa Kemoterapi: Studi Kasus," *J. Keperawatan Indones.*, vol. 20, no. 2, pp. 118–127, 2017, doi: 10.7454/jki.v20i2.478.
- [5] S. S. Utami and M. Mustikasari, "Aspek Psikososial Pada Penderita Kanker Payudara: Studi Pendahuluan," *J. Keperawatan Indones.*, vol. 20, no. 2, pp. 65–74, 2017, doi: 10.7454/jki.v20i2.503.
- [6] H. Oktavianto and R. P. Handri, "Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes," *INFORMAL Informatics J.*, vol. 4, no. 3, p. 117, 2020, doi: 10.19184/isj.v4i3.14170.
- [7] D. P. Utomo and B. Purba, "Penerapan Datamining pada Data Gempa Bumi Terhadap Potensi Tsunami di Indonesia," *Pros. Semin. Nas. Ris. Inf. Sci.*, vol. 1, no. September, p. 846, 2019, doi: 10.30645/senaris.v1i0.91.
- [8] D. P. Utomo and M. Mesran, "Analisis Komparasi Metode Klasifikasi Data Mining dan Reduksi Atribut Pada Data Set Penyakit Jantung," *J. Media Inform. Budidarma*, vol. 4, no. 2, p. 437, 2020, doi: 10.30865/mib.v4i2.2080.
- [9] Y. Mardi, "Data Mining : Klasifikasi Menggunakan Algoritma C4.5," *Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2017, doi: 10.22202/ei.2016.v2i2.1465.

- [10] H. Sulastri and A. I. Gufroni, "Penerapan Data Mining Dalam Pengelompokan Penderita Thalassaemia," *J. Nas. Teknol. dan Sist. Inf.*, vol. 3, no. 2, pp. 299–305, 2017, doi: 10.25077/teknosi.v3i2.2017.299-305.
- [11] R. Rukinah and S. Luba, "Pengetahuan Wanita Usia Subur Tentang Pencegahan Kanker Payudara," *J. Ilm. Kesehat. Sandi Husada*, vol. 10, no. 1, pp. 248–252, 2021, doi: 10.35816/jiskh.v10i1.597.
- [12] M. I. Fikri, T. S. Sabrila, and Y. Azhar, "Perbandingan Metode Naïve Bayes dan Support Vector Machine pada Analisis Sentimen Twitter," *Smatika J.*, vol. 10, no. 02, pp. 71–76, 2020, doi: 10.32664/smatika.v10i02.455.
- [13] M. Faid, M. Jasri, and T. Rahmawati, "Perbandingan Kinerja Tool Data Mining Weka dan Rapidminer Dalam Algoritma Klasifikasi," *Teknika*, vol. 8, no. 1, pp. 11–16, 2019, doi: 10.34148/teknika.v8i1.95.
- [14] D. Cahyanti, A. Rahmayani, and S. A. Husniar, "Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara," *Indones. J. Data Sci.*, vol. 1, no. 2, pp. 39–43, 2020, doi: 10.33096/ijodas.v1i2.13.
- [15] E. Tiana and S. Wahyuni, "Hasil Analisis Teknik Data Mining dengan Metode Naive Bayes untuk Mendiagnosa Penyakit Kanker Payudara," *J. Sist. Komput. dan Inform.*, vol. 1, no. 2, p. 130, 2020, doi: 10.30865/json.v1i2.1766.
- [16] R. Annisa, "Analisis Komparasi Algoritma Klasifikasi Data Mining Untuk Prediksi Penderita Penyakit Jantung," *J. Tek. Inform. Kaputama*, vol. 3, no. 1, pp. 22–28, 2019, [Online]. Available: <https://jurnal.kaputama.ac.id/index.php/JTIK/article/view/141/156>.
- [17] F. Wulandari, P. A. Jusia, and J. Jasmir, "Klasifikasi Data Mining Untuk Mendiagnosa Penyakit ISPA Menggunakan Metode Naïve Bayes Pada Puskesmas Jambi Selatan," *J. Manaj. Teknol. dan Sist. Inf.*, vol. 2, no. 3, pp. 214–227, 2020, [Online]. Available: <http://ejournal.stikom-db.ac.id/index.php/jimsi/article/view/895>.

LAMPIRAN

Pada lampiran ini penulis hanya melampirkan dataset yang akan digunakan pada penelitian ini. Yang mana penulis melampirkan tangkapan gambar awal dan akhir data yang dikarenakan jumlah data banyak sehingga apabila ditautkan langsung akan memiliki posisi yang berantakan dan dokumen menjadi error. Dan penulis melampirkan tautan ke file tersebut apabila untuk memvalidasi keaslian dan kebenaran data, dan berikut link terkait :

Link Website :

<https://www.kaggle.com/datasets/reihanenamdari/breast-cancer>

Link Drive :

<https://drive.google.com/file/d/1cTmLV6xoU3xwbhU1PdMBK7qFUGxQI1nP/vi?usp=sharing>

Atas pengertiannya, saya ucapkan terima kasih.

No	Age	Race	Marital Status	T Stage	N Stage	6th Stage	Differentiate	Grade	A Stage	Tumor Size	Estrogen Status	Progesterone Status	Regional Node Examined	Reginol Node Positive	Survival Months	Status
1	68.0	White	Married	T1	N1	IIA	Poorly differentiated	3.0	Regional	4.0	Positive	Positive	24.0	1.0	60.0	Alive
2	50.0	White	Married	T2	N2	IIIA	Moderately differentiated	2.0	Regional	35.0	Positive	Positive	14.0	5.0	62.0	Alive
3	58.0	White	Divorced	T3	N3	IIIC	Moderately differentiated	2.0	Regional	63.0	Positive	Positive	14.0	7.0	75.0	Alive
4	58.0	White	Married	T1	N1	IIA	Poorly differentiated	3.0	Regional	18.0	Positive	Positive	2.0	1.0	84.0	Alive
5	47.0	White	Married	T2	N1	IIB	Poorly differentiated	3.0	Regional	41.0	Positive	Positive	3.0	1.0	50.0	Alive
6	51.0	White	Single	T1	N1	IIA	Moderately differentiated	2.0	Regional	20.0	Positive	Positive	18.0	2.0	89.0	Alive
7	51.0	White	Married	T1	N1	IIA	Well differentiated	1.0	Regional	8.0	Positive	Positive	11.0	1.0	54.0	Alive
8	40.0	White	Married	T2	N1	IIB	Moderately differentiated	2.0	Regional	30.0	Positive	Positive	9.0	1.0	14.0	Dead
9	40.0	White	Divorced	T4	N3	IIIC	Poorly differentiated	3.0	Regional	103.0	Positive	Positive	20.0	18.0	70.0	Alive
10	69.0	White	Married	T4	N3	IIIC	Well differentiated	1.0	Distant	32.0	Positive	Positive	21.0	12.0	92.0	Alive
11	68.0	White	Widowed	T1	N1	IIA	Moderately differentiated	2.0	Regional	13.0	Positive	Positive	9.0	1.0	64.0	Dead
12	46.0	White	Married	T3	N1	IIIA	Poorly differentiated	3.0	Regional	59.0	Negative	Negative	11.0	3.0	92.0	Alive
13	65.0	White	Married	T2	N1	IIB	Poorly differentiated	3.0	Regional	35.0	Positive	Positive	13.0	3.0	56.0	Alive
14	48.0	White	Married	T1	N2	IIIA	Poorly differentiated	3.0	Regional	15.0	Positive	Positive	23.0	7.0	38.0	Alive

4007	41.0	Other	Married	T2	N1	IIB	Poorly differentiated	3.0	Regional	35.0	Negative	Negative	18.0	2.0	66.0	Alive
4008	44.0	White	Single	T4	N1	IIIB	Moderately differentiated	2.0	Regional	21.0	Positive	Positive	10.0	1.0	75.0	Alive
4009	60.0	White	Single	T1	N1	IIA	Well differentiated	1.0	Regional	18.0	Positive	Positive	19.0	2.0	52.0	Alive
4010	53.0	White	Divorced	T3	N1	IIIA	Moderately differentiated	2.0	Regional	100.0	Positive	Negative	20.0	1.0	107.0	Alive
4011	59.0	Black	Single	T2	N1	IIB	Poorly differentiated	3.0	Regional	23.0	Positive	Negative	6.0	1.0	27.0	Dead
4012	62.0	White	Widowed	T2	N2	IIIA	Moderately differentiated	2.0	Regional	25.0	Positive	Positive	24.0	4.0	100.0	Alive
4013	51.0	White	Married	T3	N3	IIIC	Poorly differentiated	3.0	Regional	80.0	Positive	Positive	20.0	12.0	79.0	Alive
4014	68.0	White	Married	T1	N1	IIA	Well differentiated	1.0	Regional	15.0	Negative	Negative	10.0	1.0	102.0	Alive
4015	40.0	White	Married	T3	N1	IIIA	Moderately differentiated	2.0	Regional	68.0	Positive	Positive	22.0	2.0	86.0	Alive
4016	65.0	White	Married	T2	N2	IIIA	Moderately differentiated	2.0	Regional	40.0	Positive	Positive	4.0	4.0	68.0	Alive
4017	54.0	White	Married	T2	N1	IIB	Well differentiated	1.0	Regional	50.0	Positive	Positive	29.0	2.0	52.0	Alive
4018	46.0	White	Married	T2	N1	IIB	Moderately differentiated	2.0	Regional	28.0	Negative	Negative	2.0	1.0	19.0	Dead
4019	64.0	White	Married	T1	N1	IIA	Moderately differentiated	2.0	Regional	10.0	Positive	Positive	11.0	1.0	70.0	Alive
4020	62.0	Other	Married	T1	N1	IIA	Moderately differentiated	2.0	Regional	9.0	Positive	Positive	1.0	1.0	49.0	Alive
4021	56.0	White	Divorced	T2	N2	IIIA	Moderately differentiated	2.0	Regional	46.0	Positive	Positive	14.0	8.0	69.0	Alive
4022	68.0	White	Married	T2	N1	IIB	Moderately differentiated	2.0	Regional	22.0	Positive	Negative	11.0	3.0	69.0	Alive
4023	58.0	Black	Divorced	T2	N1	IIB	Moderately differentiated	2.0	Regional	44.0	Positive	Positive	11.0	1.0	72.0	Alive
4024	46.0	White	Married	T2	N1	IIB	Moderately differentiated	2.0	Regional	30.0	Positive	Positive	7.0	2.0	100.0	Alive