

KLASIFIKASI PENYAKIT KANKER PAYUDARA MENGUNAKAN ALGORITMA NAÏVE BAYES

PROPOSAL TUGAS AKHIR



Untuk Memenuhi Persyaratan Penulisan

Pada Mata Kuliah Tugas Akhir

Disusun Oleh :

Fadhila

8020190034

PROGRAM STUDI TEKNIK INFORMATIKA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS DINAMIKA BANGSA

2022

IDENTITAS PROPOSAL PENELITIAN

Judul Proposal : Klasifikasi Penyakit Kanker Payudara Menggunakan
Algoritma Naïve Bayes

Program Studi : Teknik Informatika

Jenjang Pendidikan : Strata 1 (S1)

Peneliti :

- a. Nama Lengkap : Fadhila
- b. NIM : 8020190034
- c. Jenis Kelamin : Perempuan
- d. Tempat/Tgl Lahir : Sarang burung/26 juni 2001
- e. Alamat : Desa Sarang burung
- f. No. Telepon : 083175756023
- g. Email : fadhilayagituaja@gmail.com

BAB I

LANDASAN TEORI

1.1 LATAR BELAKANG

Indonesia menghadapi masalah ganda (*Double Borden*). Hal ini ditandai dengan transisi epidemiologi yang ditandai dengan perubahan pola morbiditas dan penyebab kematian di masyarakat dari penyakit menular ke berbagai jenis penyakit tidak menular [1]. Kanker adalah salah satu beragam penyakit tak menjangkit yang Insiden meningkat dari tahun ke tahun. Kanker payudara adalah jenis kanker yang paling umum pada wanita di seluruh dunia. Kanker payudara menyumbang 25,5 Republik Indonesia dari total jumlah kanker yang baru didiagnosis pada tahun 2012. Kanker adalah masalah kesehatan utama di dunia dan di Indonesia. Menurut *World Health Organization* (WHO) Kementerian Kesehatan RI (2015) pada tahun 2013, kejadian kanker meningkat dari 12,7 juta menjadi 14,2 juta kasus dari tahun 2008 hingga 2012, mencapai 11.341 pada tahun 2012. meningkat dan menurun dari tahun 2011. (19.637 kasus) [2] .

Klasifikasi data kanker bisa membantu memprediksi hasil penyakit atau menemukan genetika tumor pada pasien kanker payudara [3]. *Algoritma Naïve Bayes* memiliki tujuan untuk menghasilkan model yang akurat dari jenis-jenis data pelatihan yang diketahui dan menggunakan model tersebut untuk mengklasifikasikan data baru. Klasifikasi data kanker dapat membantu memprediksi hasil penyakit atau menemukan genetika tumor pada pasien kanker payudara. Kanker payudara adalah salah satu penyakit yang paling umum pada wanita. Di antara semua jenis kanker lainnya, kanker payudara adalah penyebab utama kematian pada wanita. Kanker adalah penyakit yang mengubah sifat sel-sel tubuh dan mendorong pembentukan sel-sel abnormal. Tumor

terbentuk ketika massa besar sel kanker yang tidak normal tumbuh. Jumlah kasus kanker payudara meningkat di seluruh dunia.

Data Mining bisa diterapkan pada bidang kesehatan, seperti mendiagnosis kanker payudara, penyakit jantung, diabetes, dll [4]. Dalam penelitian ini, penulis menggunakan *Algoritma Naïve Bayes*. *Algoritma Naïve Bayes* adalah pengklasifikasi statistik dan didasarkan pada teorema Bayes. Keuntungan dari *Algoritma Naïve Bayes* adalah dapat menghasilkan pohon keputusan yang memiliki tingkat akurasi yang dapat diterima dan dapat menangani atribut diskrit atau numerik secara efisien. Taksonomi *Naïve Bayes* digunakan untuk meningkatkan kinerja.

Berdasarkan latar belakang yang disajikan, maka judul penelitian ini adalah “Klasifikasi Penyakit Kanker Payudara Menggunakan Algoritma Naïve Bayes”.

1.2 RUMUSAN MASALAH

Berdasarkan dari latar belakang, maka didapat rumusan masalah sebagai berikut :

1. Bagaimana penerapan *Algoritma Naïve Bayes* dalam mengklasifikasi Kanker Payudara ?
2. Seberapa akurat *Algoritma Naïve Bayes* dalam mengklasifikasikan Kanker Payudara ?

1.3 BATASAN MASALAH

Berdasarkan dari latar belakang, maka adapun batasan masalah dalam penelitian ini, yaitu :

1. Data yang diperlukan adalah dataset Penyakit Kanker Payudara.
2. Variabel yang digunakan dalam penelitian ini adalah *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*.

1.4 TUJUAN DAN MANFAAT PENELITIAN

1.4.1 Tujuan Penelitian

Adapun tujuan dari penelitian yang akan dilakukan oleh penulis, yaitu :

1. Klasifikasi penyakit Kanker Payudara.
2. Menggunakan *Algoritma Naïve Bayes* untuk mendapatkan akurasi yang tepat untuk mengklasifikasi Kanker Payudara.
3. Mencari tahu seberapa akurat *Algoritma Naïve Bayes* untuk klasifikasi Kanker Payudara

1.4.2 Manfaat Penelitian

Adapun beberapa manfaat yang dapat diperoleh dari melakukan penelitian ini, yaitu :

1. Memberikan kontribusi ilmiah untuk penelitian di bidang klasifikasi *Data Mining*, khususnya klasifikasi dataset Kanker Payudara.
2. Menambah wawasan saat penulis mengklasifikasikan dataset Kanker Payudara.

BAB II

LANDASAN TEORI

2.1 DEFINISI KANKER PAYUDARA

Kanker Payudara adalah kanker yang terbentuk di jaringan payudara. Kanker payudara terjadi ketika sel-sel pada jaringan yang ada di payudara tumbuh tidak terkendali dan mengambil alih jaringan payudara yang sehat dan sekitarnya.

2.2 DATA MINING

Data mining didefinisikan sebagai pola yang mengenali informasi yang berharga dan berguna dari database yang sangat besar, atau mencoba untuk menggali informasi yang berharga dan berguna dari database yang sangat besar [5]. Menurut Gartner Group, *data mining* adalah proses menemukan data baru dengan makna, pola, dan kebiasaan dengan menyaring sebagian besar data yang tersimpan di media penyimpanan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika. *Data mining* adalah kombinasi multi disiplin yang menggabungkan pembelajaran mesin, pengenalan pola, statistik, database, dan teknik visualisasi untuk memecahkan masalah pengambilan informasi dari database besar. *Data mining*, juga dikenal sebagai penemuan pengetahuan dalam database (KDD), didefinisikan sebagai ekstraksi informasi implisit, implisit, dan tidak diketahui dari kumpulan data.

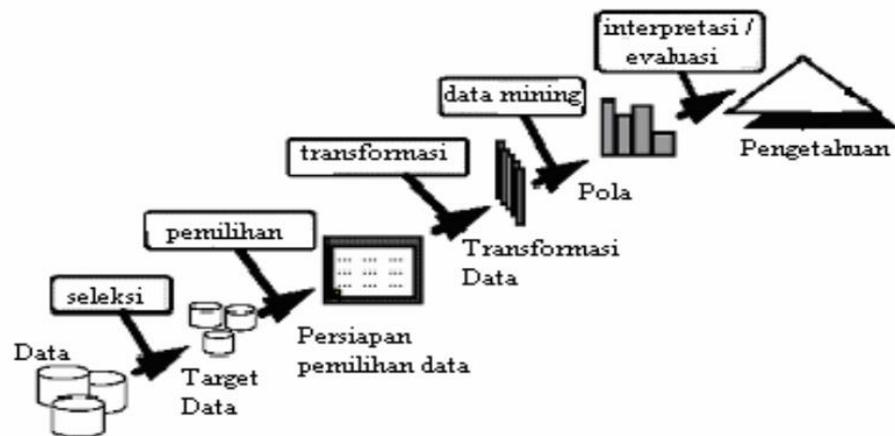
Ada beberapa metode yang diklasifikasikan sebagai KDD sebagai metode temu kembali informasi/pengetahuan, ada probabilistic seperti logika induktif, pencarian pola, dan analisis pohon keputusan. Pendekatan lain termasuk varians, analisis tren, *algoritma* genetika, jaringan saraf tiruan, dan pendekatan hibrida dari dua atau lebih pendekatan yang ada.

Pada dasarnya, ada enam elemen kunci teknologi temu kembali informasi/pengetahuan di KDD.

1. Bekerja dengan data dalam jumlah besar.
2. Efisiensi diperlukan dalam hal volume data.
3. Berikan prioritas pada penilaian dan akurasi.
4. Harus menggunakan Bahasa tingkat tinggi.
5. Gunakan beberapa bentuk pembelajaran otomatis.
6. Menghasilkan hasil yang menarik.

2.2.1 Tahapan Proses dalam *Data Mining*

Proses *data mining* terdiri dari beberapa tahap. Diagram berikut menunjukkan beberapa tahapan/proses yang terjadi pada *data mining*.



Gambar 2.1 Fase – Fase

Tahapan proses *data mining* dapat dijelaskan sebagai berikut :

1. Seleksi data

Pemilihan (seleksi) data baru sekumpulan data operasional harus terjadi sebelum tahap ekstraksi informasi *data mining* dimulai. Data terpilih yang digunakan dalam proses *data mining* disimpan dalam file terpisah dari database operasional.

2. Pre-Processing/Cleaning (pemilihan data)

Proses pembersihan termasuk menghapus data duplikat, memeriksa data yang tidak konsisten, dan memperbaiki kesalahan pada data.

3. Transformasi

Coding adalah proses transformasi data yang dipilih agar cocok untuk proses *data mining*.

4. Data Mining

Data Mining adalah proses mencari pola dan informasi yang menarik pada data terpilih dengan menggunakan teknik dan metode tertentu. Teknik, metode, atau *algoritma* dalam *data mining* sangat bervariasi. Pemilihan metode dan *algoritma* sesuai dengan kebutuhan dan tujuan.

5. Interpretasi/Evaluasi

Pola Informasi yang dihasilkan dari informasi *data mining* harus disajikan dalam format yang dapat dengan mudah dipahami oleh pihak yang berkepentingan. Fase ini memeriksa apakah pola atau informasi yang ditemukan bertentangan dengan fakta atau hipotesis yang ada.

2.2.2 Pengelompokan Data Mining

Pengelompokan *data mining* dapat dikategorikan berdasarkan fungsi yang dilakukan atau berdasarkan jenis aplikasi yang menggunakannya [6]:

1. Klasifikasi

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

2. Pengklusteran (*Clustering*)

Clustering adalah pengelompokan set data, pengamatan, atau catatan untuk membentuk kelas objek yang memiliki kesamaan. Cluster adalah kumpulan record yang memiliki kesamaan dan perbedaan dengan record di cluster lain.

3. Asosiasi

Tugas asosiasi dalam *data mining* adalah menemukan atribut-atribut yang terjadi bersama-sama. Salah satu implementasi asosiasi tersebut adalah analisis keranjang belanja.

4. Deskripsi

Peneliti dan analisis mungkin mencoba menemukan cara untuk menjelaskan pola dan tren dalam data. Deskripsi dari pola dan tren sering kali memberikan kemungkinan penjelasan untuk pola atau tren.

5. Estimasi

Estimasi hampir mirip dengan klasifikasi, kecuali bahwa variabel target untuk estimasi adalah numerik daripada kategori. Sebuah model dibangun menggunakan baris data (dataset) lengkap yang memberikan nilai-nilai variabel target sebagai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai variabel prediksi.

6. Prediksi

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil akan ada dimasa mendatang.

2.3 KLASIFIKASI

Klasifikasi didefinisikan sebagai bentuk analisis data untuk mengekstraksi model yang digunakan untuk memprediksi label kelas [7]. Kelas dalam klasifikasi merupakan atribut yang paling unik dari dataset dan variabel bebas dalam statistic.

2.4 ALGORITMA NAÏVE BAYES

Pengklasifikasi Bayesian adalah pengklasifikasi statistik dan didasarkan pada teorema Bayes. Teori keputusan Bayesian adalah pendekatan statistik yang mendasari pengenalan pola. Harus ada masalah yang dapat diamati secara statistik dengan menggunakan *algoritma* ini untuk klasifikasi. Misalkan X adalah himpunan atribut data dan h adalah variabel kelas. Jika kelas terkait dengan atribut,

kita membutuhkan X dan h sebagai variabel acak dan kita mendapatkan hubungan probabilitas $P(h|X)$. Probabilitas posterior adalah h dan sebaliknya untuk periode $P(h)$.

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

Keterangan :

X : Data dengan class yang belum diketahui

H : Hipotesis data X merupakan suatu class spesifik

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

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

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

$P(X)$: Probabilitas X

2.5 PENELITIAN SEJENIS

Tabel 2.1 Penelitian Sejenis

No .	Judul, Penulis dan Tahun	Metode	Atribut	Akurasi
1	“Analisis Klasifikasi Kanker Payudara Menggunakan Algoritma Naive Bayes” Hardian Oktavianto1 , Rahman Puji Handri2. 2019 [8]	<i>Naive Bayes</i>	<i>Id number, Clump Thickness, Uniformity of Cell Size, Uniformity of Cell Shape, Marginal Adhesion, Single Epithelial Cell Size, Bare Nuclei, Bland Chromatin, Normal Nucleoli, Mitoses, dan Class</i>	Performa algoritma naive bayes untuk klasifikasi kanker payudara menghasilkan nilai yang baik, dimana rata – rata persentase

				<p>data yang terklasifikasi dengan benar mencapai 96.9% dan rata – rata persentase data terklasifikasi salah hanya 3.1%.</p> <p>Sedangkan tingkat efektivitas klasifikasi dengan naive bayes ini termasuk tinggi, dimana rata – rata nilai precision dan recall berada di sekitar 0.96.</p> <p>Nilai precision dan recall paling tinggi yaitu ketika data uji menggunakan percentage split 40% dengan nilai masing – masing secara</p>
--	--	--	--	--

				berurutan mencapai 0.974 dan 0.973.
2	<p>“Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio Balqis Aisyah Farahdiba1 dan Yusuf Sulisty Nugroho2. 2019 [9]</p>	<p><i>Algoritma Gain Ratio</i></p>	<p>Ketebalan Rumpun, Keseragaman Ukuran Sel, Keseragaman Bentuk Sel, Adhesi Marjinal, Ukuran Sel Epitel Tunggal, Ukuran Asli Nuclei, Kromatin, Keadaan Nucleoli Normal, Mitosi, Kelas : (2 untuk Jinak, 4 untuk Ganas)</p>	<p>Hasil klasifikasi kanker payudara menggunakan algoritma gain ratio ini dapat diambil kesimpulan bahwa performa algoritma yang diukur berdasarkan tingkat recall, accuracy dan precision yang masing-masing memiliki nilai 92,55%, 95,17% dan 93,76% menunjukkan bahwa algoritma gain ratio sangat baik digunakan dalam penelitian ini.</p>

3	<p>“SISTEM PENDUKUNG KEPUTUSAN DETEKSI PENYAKIT KANKER PAYUDARA MENGGUNAKAN ALGORITMA NAIVE BAYES (Ivandari1), Erni Rahmawatie2), M. Adib Al Karomi3). 2018 [10]</p>	<p><i>Algoritma Naive Bayes</i></p>	<p>Id, clump thickness, uniformity of cell size, uniformity of cell shape, marginal adhesion, single ephitelial cell size, bare nuclei, bland chromatin, normal nucleoli, serta mitoses.</p>	<p>Tingkat akurasi algoritma naive bayes untuk klasifikasi keganasan kanker payudara adalah 97,43%. Akurasi ini didapat dengan menggunakan dataset breast cancer wisconsin dengan memanfaatkan aplikasi rapid miner. Perhitungan menggunakan x-validation serta confusion matrix.</p>
4	<p>Teknik SMOTE dan Gini Score dalam Klasifikasi Kanker Payudara *Nur Ghaniaviyanto Ramadhan 1 , Faisal Dharma</p>	<p><i>Algoritma Smooth and Gini Score</i></p>	<p><i>smoothness_se, compactness_se, concavity_se, concave points_se, symmetry_se, fractal_dimension_se, radius_worst, texture_worst, perimeter_worst</i></p>	<p>Hasil yang diperoleh menunjukkan bahwa metode yang diusulkan dalam penelitian ini mampu menurunkan</p>

	Adhinata2. 2021 [11]		, <i>area_worst</i> , <i>smoothness_worst</i> , <i>compactness_worst</i> , <i>concavity_worst</i> , <i>concave points_worst</i> , <i>symmetry_worst</i> , <i>fractal_dimension_worst</i> , <i>diagnosis</i> .	nilai FN dan meningkatkan nilai TN, namun metode yang diusulkan masih belum mampu meningkatkan nilai TP dan menurunkan nilai FP. Hal ini terjadi karena kurangnya pengujian sampel yang digunakan.
5	“Analisis performa metode Knn pada Dataset pasien pengidap Kanker Payudara Dewi Cahyantia,1, Alifah Rahmayania,2, Syafira Ainy Husniara,3. 2020 [12]	<i>Algoritma KNN</i>	<i>Crossvalidation, Akurasi, Presisi, Recall F-measure.</i>	Dari perhitungan 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%

				<p>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).</p>
6	<p>“ANALISA PREDIKSI KEKAMBUHAN KANKER PAYUDARA DENGAN MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBOR Noki Ardian Madyaningrum 1 , Sulastri 2. 2019 [13]</p>	<p><i>Algoritma KNN</i></p>	<p><i>Class, Age, Menopause, TumorSize, invoNodes. NodeCaps, DegMalig. Breast, BreastOut, Irradiat.</i></p>	<p>Berdasarkan percobaan pertama dengan komposisi data training 70% dan data testing 30% didapatkan hasil nilai K terbaik pada K= 5 dan nilai akurasi 80%. Sedangkan pada percobaan kedua dengan komposisi data training 75% dan data testing</p>

				25% didapatkan hasil nilai K terbaik pada K= 13 dan nilai akurasi 73%. Adapun pada percobaan ketiga dengan komposisi data training 80% dan data testing 20% didapatkan hasil nilai K terbaik pada K= 15 dan nilai akurasi 74%.
7	“Penerapan Algoritma C4.5 untuk Klasifikasi Tingkat Keganasan Kanker Payudara Dwi Ayu Nursela. 2020 [14]	<i>Algoritma C4.5</i>	<i>Uniformity of cell size</i>	Dari hasil Pengujian menghasilkan akurasi yang cukup besar yaitu 97,12 % Precision sebesar 93,02% %, dan Recall sebesar 100,00%
8	“Diagnosa Penderita Penyakit Kanker Payudara Menggunakan	<i>Algoritma Naïve Bayes</i>	<i>Age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin, MCP.1, Label Kelas.</i>	Untuk perhitungan manual sendiri, hasil akurasi untuk data

	<p>Metode Naïve Bayes Taufik Frissetyo 1,* , Heri Kuswara 2. 2019 [15]</p>			<p>training sebesar 81.00% sama dengan hasil pengklasifikasi an menggunakan Rapidminer sementara hasil akurasi data testing sebesar 62.50% sama persis dengan hasil perhitungan menggunakan Rapidminer</p>
9	<p>“Data Mining untuk Klasifikasi Penderita Kanker Payudara Berdasarkan Data dari University Medical Center Menggunakan Algoritma Naïve Bayes Ibnu Ramadhan, Kurniawati. 2020 [4]</p>	<p><i>Naïve Bayes</i></p>	<p><i>Class, Age, Menopause, Tumor-Size, Inv-Nodes, Node-Caps, Deg-Malig, Breast, Breast-Quad dan Irradiant.</i></p>	<p>Class pasien “Kambuh” lebih besar dibandingkan Class pasien “Tidak Kambuh”. Akurasi terhadap klasifikasi data yang sudah dilakukan sebelumnya menggunakan algoritma Naive Bayes,</p>

				diperoleh nilai akurasi terhadap pasien yang mengalami kambuh atau tidak sebesar 71,43% di mana hal ini juga bisa disebabkan oleh kekurangan kompleksan data sehingga model dapat memprediksi secara akurat.
10	<p>“Implementasi Algoritma Random Forest Untuk Mendiagnosis Kejadian Berulang (Kekambuhan) Pada Kanker Payudara Berbasis Web Runi Hari Bagus Saputra1*, Roy Mubarak2. 2022 [16]</p>	<p><i>Algoritma Random Forest</i></p>	<p><i>Class, Age, Menopause, Tumor-Size, Inv-Nodes, Node-Caps, Deg-Malig, Breast, Breast-Quad dan Irradiant.</i></p>	<p>Sistem Diagnosis Kekambuhan Kanker Payudara berbasis web bahwa kekambuhan pada kanker payudara dapat di diagnosis dengan menggunakan sistem pakar. Implementasi</p>

				Algoritma Random Forest untuk mendiagnosis kekambuhan kanker payudara dengan cara membagi setiap simpul menjadi variable berdasarkan variable terbaik diantara variable yang dipilih secara acak disetiap simpul.
--	--	--	--	--

BAB III

METODOLOGI PENELITIAN

3.1. KERANGKA KERJA PENELITIAN

Untuk membantu dalam penyusunan penelitian ini, maka perlu adanya kerangka kerja (*framework*) yang jelas tahapan-tahapannya. Kerangka kerja ini merupakan langkah- langkah yang akan dilakukan dalam penyelesaian masalah yang akan dibahas. Adapun kerangka kerja penelitian yang akan digunakan adalah sebagai berikut :



Gambar 3.1. Tahapan Penelitian

Berdasarkan kerangka kerja penelitian yang telah di gambar di atas, maka dapat di uraikan pembahasan masing masing tahapan dalam penelitian adalah sebagai berikut :

1. Perumusan Masalah

Masalah yang dirumuskan dalam penelitian ini adalah bagaimana penerapan metode *Algoritma Naïve Bayes* dalam klasifikasi penyakit diabetes.

2. Penentuan Tujuan

Tujuan yang akan dicapai dalam penelitian ini adalah ingin mengetahui berapa besar tingkat akurasi *Algoritma Naïve Bayes* dalam klasifikasi penyakit kanker payudara.

3. Mempelajari Literatur

Mempelajari literatur-literatur yang dapat mencapai tujuan penelitian, literatur-literatur bersumber dari buku-buku perpustakaan Universitas Dinamika Bangsa Jambi dan jaringan internet. Literatur-literatur yang digunakan nanti dilampirkan dalam daftar pustaka.

4. Pengumpulan Data dan Informasi

Dalam pengumpulan data, penulis mendapatkan dataset online yang terdapat pada sebuah website di internet.

5. Proses Data Mining

Data Mining adalah proses pengekstrasian *knowledge* yang tersimpan dalam dataset bervolume besar. Untuk mendapatkan *knowledge* dalam dataset digunakanlah *Algoritma Naïve Bayes*.

6. Pengujian

Pada tahap ini dilakukan pengujian dari hasil yang didapat dari tahap sebelumnya sebagai pedoman untuk mendapatkan hasil klasifikasi penyakit diabetes.

7. Pembuatan Laporan

Pada tahap ini dilakukan pembuatan laporan, membuat hasil akhir dari suatu kegiatan penelitian berdasarkan data dan fakta yang telah diamati pada saat meneliti.

3.2. ANALISA TEKNIK ALGORITMA NAÏVE BAYES

Pada bagian ini data dan informasi yang diperoleh dan diproses dengan menggunakan *Algoritma Naïve Bayes* untuk mendapatkan hasil yang sesuai.

3.3. METODE PENELITIAN

Metode pengumpulan data yang digunakan dalam penelitian adalah mencari dan mengambil *dataset* online dari sebuah website yang bernama “*www.kaggle.com*” untuk dijadikan pengujian dalam penelitian ini. Pengumpulan data dilakukan untuk memperoleh hasil dan informasi yang dibutuhkan dalam hal mencapai tujuan penelitian.

3.4. ALAT PENELITIAN

Penulis menggunakan beberapa alat/piranti yang digunakan untuk melakukan pengolahan data/bahan penelitian, yaitu:

1. Hardware, dengan spesifikasi sebagai berikut:
 - a. Laptop, CPU Intel® Core™ i3-6006U @ 2,00 GHz (4 CPU), ~2,0 GHz
 - b. RAM : 10 GB
 - c. SSD : 500 GB
2. Software, dengan keterangan sebagai berikut:
 - a. OS Windows 11 (64 bit)
 - b. Microsoft Excel & Word 2016
 - c. RapidMiner Studio Professional 7.1

3.5. JADWAL PENELITIAN

No	Jenis Kegiatan	Bulan																									
		Sept				Okt				Nov				Des				Jan				Feb					
		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4		
1	Perumusan Masalah	■	■																								
2	Penentuan Tujuan	■	■																								
3	Mempelajari Literatur		■	■	■	■																					
4	Pengumpulan Data & Informasi				■	■	■																				
5	Proses Data Mining							■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■	■		
6	Pengujian															■	■	■	■	■	■	■	■	■	■		
7	Pembuatan Laporan																							■	■		

DAFTAR PUSTAKA

- [1] Sukaca and Suryaningsih, “Analisis risiko kanker payudara berdasar riwayat pemakaian kontrasepsi hormonal dan usia,” *J. Berk. Epidemiol. Vol. 3, No. 1 Januari 2015* 12–23, vol. 3, pp. 12–23, 2009.
- [2] J. K. Masyarakat, “(1) , 2) , 2) 1,” vol. 4, 2016.
- [3] I. Solikin, “Teknik Data Mining untuk Prediksi Kanker Payudara yang Efisien,” vol. 3, no. 3, pp. 63–67, 2021.
- [4] I. Ramadhan and K. Kurniawati, “Data Mining untuk Klasifikasi Penderita Kanker Payudara Berdasarkan Data dari University Medical Center Menggunakan Algoritma Naïve Bayes,” *JURIKOM (Jurnal Ris. Komputer)*, vol. 7, no. 1, p. 21, 2020, doi: 10.30865/jurikom.v7i1.1755.
- [5] L. Hermawanti, “Penerapan Algoritma Klasifikasi C4.5 untuk Diagnosis Penyakit Kanker Payudara,” *J. Tek. Unisfat*, vol. 7, no. 1, pp. 57–64, 2012.
- [6] G. Gunadi and D. I. Sensuse, “Penerapan Metode Data Mining Market Basket Analysis Terhadap Data Penjualan Produk Buku Dengan Menggunakan Algoritma Apriori Dan Frequent Pattern Growth (Fp-Growth);,” *Telematika*, vol. 4, no. 1, pp. 118–132, 2012.
- [7] 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>
- [8] 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.
- [9] B. Aisyah and Y. Sulistyono, “Klasifikasi Kanker Payudara Menggunakan Algoritma Gain Ratio,” *J. Tek. Elektro*, vol. 8, no. 2, pp. 43–46, 2016.

- [10] Ivandari, E. Rahmawatie, and M. A. Al Karomi, “Sistem Pendukung Keputusan Deteksi Penyakit Kanker Payudara Menggunakan Algoritma Naive Bayes,” *Semin. Nasional Edusaintek FMIPA UNIMUS 2018*, no. Prasetyo 2012, pp. 149–153, 2018.
- [11] G. Ramadhan and F. D. Adhinata, “Teknik SMOTE dan Gini Score dalam Klasifikasi Kanker Payudara,” vol. 9, no. 2, pp. 125–134, 2021.
- [12] 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.
- [13] N. A. Madyaningrum and Sulastri, “Analisa Prediksi Kekambuhan Kanker Payudara Dengan,” *Proceeding SINTAK 2019*, pp. 180–185, 2019.
- [14] D. A. Nursela, T. I. S, F. I. Komputer, and J. N. I. No, “Penerapan Algoritma C4 . 5 untuk Klasifikasi Tingkat Keganasan Kanker Payudara,” pp. 1–5, 2014.
- [15] T. Frissetyo and H. Kuswara, “Diagnosa Penderita Penyakit Kanker Payudara Menggunakan Metode Naïve Bayes,” *Inf. Manag. Educ. Prof.*, vol. 4, no. 1, pp. 51–62, 2019.
- [16] R. Hari, B. Saputra, and R. Mubarok, “Implementasi Algoritma Random Forest Untuk Mendiagnosis Kejadian Berulang (Kekambuhan) Pada Kanker Payudara Berbasis Web,” vol. 1, no. 06, pp. 564–572, 2022.

LAMPIRAN DATASET

Age	Race	Marital Status	T Stage	N Stage	6th Stage	differentiate	Grade	A Stage	Tumor Size	Estrogen Status	Progesterone Status	Regional Node Examined	Reginod Node Positive	Survival Months	Status
68	White	Married	T1	N1	IIA	Poorly differentiated	3	Regional	4	Positive	Positive	24	1	60	Alive
50	White	Married	T2	N2	IIIA	Moderately differentiated	2	Regional	35	Positive	Positive	14	5	62	Alive
58	White	Divorced	T3	N3	IIIC	Moderately differentiated	2	Regional	63	Positive	Positive	14	7	75	Alive
58	White	Married	T1	N1	IIA	Poorly differentiated	3	Regional	18	Positive	Positive	2	1	84	Alive
47	White	Married	T2	N1	IIIB	Poorly differentiated	3	Regional	41	Positive	Positive	3	1	50	Alive
51	White	Single	T1	N1	IIA	Moderately differentiated	2	Regional	20	Positive	Positive	18	2	89	Alive
51	White	Married	T1	N1	IIA	Well differentiated	1	Regional	8	Positive	Positive	11	1	54	Alive
40	White	Married	T2	N1	IIIB	Moderately differentiated	2	Regional	30	Positive	Positive	9	1	14	Dead
40	White	Divorced	T4	N3	IIIC	Poorly differentiated	3	Regional	103	Positive	Positive	20	18	70	Alive
69	White	Married	T4	N3	IIIC	Well differentiated	1	Distant	32	Positive	Positive	21	12	92	Alive
68	White	Widowed	T1	N1	IIA	Moderately differentiated	2	Regional	13	Positive	Positive	9	1	64	Dead
46	White	Married	T3	N1	IIIA	Poorly differentiated	3	Regional	59	Negative	Negative	11	3	92	Alive
65	White	Married	T2	N1	IIIB	Poorly differentiated	3	Regional	35	Positive	Positive	13	3	56	Alive
48	White	Married	T1	N2	IIIA	Poorly differentiated	3	Regional	15	Positive	Positive	23	7	38	Alive
62	White	Divorced	T2	N3	IIIC	Moderately differentiated	2	Regional	35	Positive	Positive	16	14	64	Alive
61	White	Married	T1	N1	IIA	Moderately differentiated	2	Regional	19	Positive	Positive	20	1	49	Alive
56	White	Single	T2	N1	IIIB	Moderately differentiated	2	Regional	46	Positive	Positive	1	1	105	Alive
43	White	Married	T2	N1	IIIB	Moderately differentiated	2	Regional	24	Positive	Positive	22	1	62	Alive
48	Black	Divorced	T2	N1	IIIB	Moderately differentiated	2	Regional	25	Positive	Positive	16	1	107	Alive
60	White	Divorced	T2	N1	IIIB	Moderately differentiated	2	Regional	29	Negative	Negative	20	1	77	Alive
48	Other	Married	T2	N1	IIIB	Well differentiated	1	Regional	30	Positive	Positive	15	2	81	Alive
57	White	Married	T2	N2	IIIA	Moderately differentiated	2	Regional	40	Positive	Positive	15	5	50	Alive
55	White	Single	T2	N1	IIIB	Poorly differentiated	3	Regional	29	Positive	Positive	4	1	78	Alive
48	White	Married	T3	N1	IIIA	Poorly differentiated	3	Regional	70	Positive	Negative	18	1	102	Alive
62	White	Married	T1	N3	IIIC	Poorly differentiated	3	Regional	20	Positive	Positive	26	22	98	Alive
63	White	Married	T2	N3	IIIC	Moderately differentiated	2	Regional	22	Positive	Positive	31	17	70	Alive
48	White	Married	T2	N3	IIIC	Poorly differentiated	3	Regional	50	Positive	Positive	25	23	102	Alive
46	White	Married	T1	N1	IIA	Well differentiated	1	Regional	17	Positive	Positive	14	1	82	Alive
57	White	Divorced	T2	N2	IIIA	Well differentiated	1	Regional	25	Positive	Negative	14	4	64	Alive
66	White	Married	T2	N1	IIIB	Poorly differentiated	3	Regional	21	Positive	Positive	10	1	86	Alive
47	White	Separated	T2	N1	IIIB	Well differentiated	1	Regional	40	Positive	Positive	3	1	52	Alive
53	White	Married	T1	N1	IIA	Moderately differentiated	2	Regional	15	Positive	Positive	5	1	49	Alive
59	White	Single	T1	N1	IIA	Moderately differentiated	2	Regional	15	Positive	Positive	6	1	90	Alive
60	White	Married	T1	N1	IIA	Moderately differentiated	2	Regional	20	Positive	Negative	19	3	62	Alive
46	Other	Married	T2	N3	IIIC	Moderately differentiated	2	Regional	30	Positive	Positive	19	10	31	Alive
51	White	Married	T1	N1	IIA	Well differentiated	1	Regional	10	Positive	Positive	9	2	77	Alive
54	White	Married	T2	N2	IIIA	Moderately differentiated	2	Regional	27	Positive	Negative	21	6	37	Alive
51	White	Married	T1	N1	IIA	Moderately differentiated	2	Regional	18	Positive	Positive	14	2	103	Alive
49	White	Married	T2	N1	IIIB	Moderately differentiated	2	Regional	35	Positive	Positive	10	2	82	Alive
51	Other	Divorced	T2	N2	IIIA	Moderately differentiated	2	Regional	23	Positive	Positive	15	5	105	Alive
57	White	Single	T3	N3	IIIC	Moderately differentiated	2	Regional	70	Positive	Positive	12	12	42	Dead
64	White	Married	T1	N1	IIA	Moderately differentiated	2	Regional	17	Positive	Positive	12	2	61	Alive
62	Other	Married	T2	N1	IIIB	Moderately differentiated	2	Regional	21	Positive	Negative	2	1	86	Alive
53	Black	Single	T2	N2	IIIA	Poorly differentiated	3	Regional	23	Positive	Negative	15	6	84	Alive
55	White	Married	T1	N1	IIA	Moderately differentiated	2	Regional	5	Positive	Positive	11	2	63	Alive
66	White	Married	T3	N2	IIIA	Moderately differentiated	2	Regional	51	Positive	Positive	10	5	90	Alive
42	White	Married	T1	N3	IIIC	Moderately differentiated	2	Regional	9	Negative	Negative	15	2	39	Dead
62	White	Widowed	T1	N1	IIA	Moderately differentiated	2	Regional	15	Positive	Positive	12	1	59	Alive
53	White	Married	T2	N2	IIIA	Moderately differentiated	2	Regional	32	Positive	Positive	16	9	82	Alive
69	White	Widowed	T1	N1	IIA	Moderately differentiated	2	Regional	5	Positive	Positive	8	3	82	Alive
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---
60	Black	Married	T3	N1	IIIA	Poorly differentiated	3	Regional	120	Positive	Negative	7	1	86	Alive
63	White	Married	T3	N1	IIIA	Well differentiated	1	Regional	77	Positive	Negative	20	2	70	Alive
57	White	Married	T1	N1	IIA	Moderately differentiated	2	Regional	2	Positive	Negative	16	1	74	Alive