

BAB II
TINJAUAN PUSTAKA

II.1 Penelitian Terkait

Berikut ini adalah beberapa jurnal penelitian terdahulu terkait judul penelitian skripsi pada tabel di bawah ini.

Tabel II.1. Penelitian Terkait

No	Peneliti	Judul	Hasil	Perbedaan
1	Irma Apriliani (2017)	Perbandingan Metode Naive Bayes dan K-Nearest Neighbour untuk Prediksi Perceraian (Studi Kasus : Pengadilan Agama Cimahi)	Metode Naive Bayes memiliki nilai akurasi lebih tinggi dibandingkan metode K-Nearest Neighbour karena perhitungan metode Naive Bayes lebih mudah dan tidak membutuhkan waktu yang lama saat prediksi untuk diaplikasikan ke database dengan data yang besar dan perhitungan manualnya tidak rumit	Penulis hanya menggunakan metode Naive Bayes dan melakukan prediksi terhadap perceraian pasangan muda di Kota Medan
2	Ahmad Thoriq Susilo, Hendra Setiawan (2021)	Penerapan Metode Naive Bayes untuk Memprediksi Tingkat Kemenangan pada Game Mobile Legends	Dalam pengujian pada 2 (dua) tim sebanyak 20 pertandingan untuk mendapatkan hasil kemenangan di <i>games Mobile Legends</i> didapatkan hasil prediksi dengan tingkat akurasi benar	Implementasi Naive Bayes untuk prediksi perceraian pasangan muda dan menggunakan aplikasi laman web

No	Peneliti	Judul	Hasil	Perbedaan
			sebesar 75% dan akurasi salah sebesar 25%	
3	Hakam Febtadianrano Putro, Retno Tri Vulansari (2020)	Penerapan Metode Naive Bayes untuk Klasifikasi Pelanggan	Berdasarkan hasil pengujian dari 25 data uji dihasilkan sebanyak 23 data yang berhasil diklasifikasikan dengan Metode Naive Bayes dengan tingkat akurasi mencapai 92%.	Penerapan metode Naive Bayes untuk prediksi tingkat kepercayaan pasangan muda di masa mendatang
4	Heliyanti Susana, Nana Suarna, Faturrohman, Kaslani (2022)	Penerapan Model Klasifikasi Metode Naive Bayes Terhadap Penggunaan Akses Internet	Hasil prediksi menggunakan algoritma Naive Bayes menggunakan aplikasi Rapid Miner pada 270 siswa menghasilkan tingkat penggunaan internet dengan tingkat akurasi sebesar 89,83%	Implementasi metode Naive Bayes untuk memprediksi tingkat kepercayaan pasangan muda di masa mendatang dengan menggunakan laman web
5	Naisha Rahma Indraswari (2018)	Aplikasi Prediksi Usia Kelahiran Dengan Metode Naive Bayes	Kegiatan prediksi dapat dilakukan setelah memasukkan data yang dialami ibu hamil yang kemudian akan menampilkan hasil prediksi sesuai dengan perhitungan Naive Bayes. Pengujian algoritma Naive Bayes menunjukkan bahwa perhitungan	Penerapan metode Naive Bayes untuk prediksi tingkat kepercayaan pasangan muda di masa mendatang

No	Peneliti	Judul	Hasil	Perbedaan
			manual dibandingkan dengan perhitungan aplikasi memiliki hasil yang sama	

II.2 Studi Literatur

Untuk mendukung keberhasilan penelitian ini, peneliti melakukan pendekatan teoritis melalui beberapa literatur yang berhubungan dengan penelitian yang dilakukan.

Beberapa tinjauan pustaka pada penelitian ini, yaitu :

II.2.1 Data Mining

Secara sederhana, data mining mengacu pada penggalian atau pengetahuan dari sejumlah data yang besar (Irma Apriliani Dahlia, 2018). *Data mining* merupakan proses yang menggunakan statistik, matematika, *artificial intelligence* dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai *database* besar. Data mining terutama digunakan untuk mencari pengetahuan yang terdapat dalam basis data yang besar sehingga sering juga disebut dengan *Knowledge Discovery in Database* (KDD), adalah kegiatan meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar. Hasil dari proses data mining dapat digunakan sebagai evaluasi pengambilan keputusan di masa depan (Putria N.E., 2018).

Definisi publik dari data mining merupakan metode pencarian pola-pola pengetahuan yang tersembunyi yang tidak diketahui sebelumnya dari suatu sekumpulan data yang sangat besar dalam database, data warehouse, atau media penyimpanan lainnya (Rahmawati, F., 2018).

Data mining digunakan untuk menggali nilai tambah berupa informasi yang tidak diketahui secara manual dari suatu basis data. Informasi diperoleh dengan cara mengekstraksi dan mengenali pola yang penting atau menarik dari data yang terdapat dalam basis data (Sulindawati, 2017).

Terdapat beberapa metode atau cara yang digunakan pada data mining untuk mengelola atau analisis penggalian data yang ditujukan berdasarkan tujuan penggunaannya, diantara *classification* (pengklasifikasian), *association* (asosiasi), *clustering* (segmentasi), *regression* (regresi), *forecasting* (peramalan), *sequencing*, dan *descriptive*.

II.2.2 Prediksi

Prediksi adalah suatu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya dapat diperkecil. Prediksi menunjukkan apa yang akan terjadi pada suatu keadaan tertentu dan merupakan input bagi proses perencanaan dan pengambilan keputusan (Mohammad Kafil, 2019).

Fungsi prediksi merupakan sebuah fungsi bagaimana suatu proses akan menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel yang ada pada data. Ketika sudah menemukan suatu pola, maka pola yang didapat tersebut dapat digunakan untuk memprediksi variabel lain yang belum diketahui nilainya.

II.2.3 Metode Naive Bayes

Metode Naive Bayes merupakan metode yang memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya, dimana metode Naive Bayes ini merupakan metode data mining yang termasuk ke dalam metode klasifikasi data mining yang populer diantara algoritma-algoritma lainnya (Ratih Yulia, 2019). Algoritma Naive Bayes sangat berguna untuk pembelajaran algoritma, pengetahuan masa lalu dan pengamatan data yang bisa digabungkan, ini dapat membantu dalam memberikan perspektif yang bermanfaat untuk memahami dan juga mengevaluasi banyak pembelajaran. Ini membantu menentukan probabilitas yang tepat untuk hipotesis (Parveen dan Pandey, 2017).

Algoritma Naive Bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa lalu atau masa sebelumnya dengan asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian. Istilah lain secara lebih sederhana menyatakan bahwa

teorema Bayes adalah rumus matematika sederhana untuk menemukan probabilitas ketika sudah mengetahui probabilitas tertentu lainnya.

Metode Naïve Bayes menggunakan data yang sudah ada yang bernama data *training* yang digunakan untuk membentuk sebuah model pengklasifikasian. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Kemudian membutuhkan data latih yang bernama data uji atau data *testing* yang digunakan untuk mengukur sejauh mana pengklasifikasian berhasil melakukan klasifikasi dengan benar (Parveen & Pandey, 2017).

Keuntungan penggunaan Naive Bayes adalah bahwa metode ini hanya membutuhkan jumlah data pelatihan (*Training Data*) yang kecil untuk menentukan estimasi parameter yang diperlukan dalam proses pengklasifikasian. Persamaan Naïve Bayes seperti berikut ini.

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

Dimana :

X : Data dengan *class* yang belum diketahui

H : Hipotesis data merupakan suatu *class* spesifik

$P(H/X)$: Probabilitas hipotesis H berdasarkan 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

Persamaan di atas menjelaskan bahwa peluang masuknya sampel karakteristik tertentu dalam kelas H adalah peluang munculnya kelas H dikalikan dengan peluang kemunculan karakteristik-karakteristik sampel pada kelas H , kemudian dibagi dengan peluang kemunculan karakteristik sampel secara global (X).

Dari persamaan di atas, terdapat contoh kasus pada Jurnal Informatika yang ditulis oleh Ratih Yulia Hayunngtyas (2019) dengan judul ““Penerapan Algoritma Naïve Bayes untuk Menentukan Pilihan Web”. Penelitian tersebut ditentukan oleh beberapa atribut, seperti usia, jenis kelamin, dan pendidikan terakhir.

Berdasarkan hasil dan pembahasan, penelitian dilakukan terhadap 50 data seperti tabel di bawah ini.

Tabel II.2. Data Penelitian

No	Usia	Jenis Kelamin	Pendidikan Terakhir	Pilihan Web
1	Anak-anak	P	SD	Pendidikan
2	Anak-anak	P	SMP	Social media
3	Anak-anak	P	SD	Pendidikan
4	Remaja	L	SMA	Game
5	Dewasa	L	S1	Game
6	Dewasa	P	S1	Game
7	Remaja	P	SMA	Social media
8	Remaja	L	SMA	Social media
9	Remaja	P	SMA	Game

10	Anak-anak	P	SMP	Game
11	Anak-anak	P	SD	Social media
12	Anak-anak	P	SD	Social media
13	Anak-anak	P	SMP	Pendidikan
14	Anak-anak	P	SMP	Pendidikan
15	Remaja	P	SMA	Pendidikan
16	Remaja	P	SMA	Pendidikan
17	Remaja	P	SMA	Social media
18	Remaja	P	SMA	Social media
19	Remaja	P	SMA	Social media
20	Dewasa	P	S1	Social media
21	Dewasa	P	S1	Social media
22	Remaja	L	SMA	Game
23	Remaja	P	SMA	Game
24	Remaja	P	SMA	Game
25	Remaja	L	SMA	Game
26	Remaja	L	S1	Game
27	Dewasa	P	S1	Game
28	Dewasa	P	S1	Social media
29	Dewasa	L	S1	Social media
30	Dewasa	L	S1	Game
31	Dewasa	P	SMA	Social media

32	Dewasa	P	SMA	Social media
33	Dewasa	P	SMA	Social media
34	Dewasa	P	S1	Social media
35	Dewasa	L	S1	Game
36	Remaja	P	SMA	Social media
37	Remaja	L	SMA	Game
38	Remaja	L	SMA	Pendidikan
39	Remaja	P	SMA	Pendidikan
40	Dewasa	L	S1	Social media
41	Dewasa	P	S1	Social media
42	Remaja	P	SMA	Pendidikan
43	Remaja	P	SMA	Pendidikan
44	Remaja	P	SMA	Social media
45	Remaja	P	SMA	Social media
46	Remaja	L	S1	Pendidikan
47	Dewasa	L	S1	Game
48	Dewasa	L	S1	Game
49	Dewasa	L	S1	Game
50	dewasa	L	S1	Game

Berikut adalah data tes yang akan dicari hasil perhitungannya dengan metode Naïve Bayes, dapat dilihat pada tabel berikut.

Tabel II.3. Data Tes

Usia	Jenis Kelamin	Pendidikan Terakhir	Pilihan Web
Remaja	L	SMA	?

Pertama menentukan nilai $P(H)$ seperti di bawah ini.

$$P(\text{Pendidikan}) = 11/50 = \mathbf{0,22}$$

$$P(\text{Social media}) = 20/50 = \mathbf{0,40}$$

$$P(\text{Game}) = 19/50 = \mathbf{0,38}$$

Selanjutnya menghitung nilai $P(X|H) / P(X)$ seperti di bawah ini

a. $P(\text{Usia}=\text{"Remaja"} \mid \text{Pilihan web} = \text{"Pendidikan"}) = 7/11 = 0,64.$

$$P(\text{Usia}=\text{"Remaja"} \mid \text{Pilihan web} = \text{"Social Media"}) = 8/20 = 0,40.$$

$$P(\text{Usia}=\text{"Remaja"} \mid \text{Pilihan web} = \text{"Game"}) = 8/19 = 0,42.$$

b. $P(\text{Jenis Kelamin}=\text{"L"} \mid \text{Pilihan web} = \text{"Pendidikan"}) = 2/11 = 0,18.$

$$P(\text{Jenis Kelamin}=\text{"L"} \mid \text{Pilihan web} = \text{"Social Media"}) = 2/20 = 0,10.$$

$$P(\text{Jenis Kelamin}=\text{"L"} \mid \text{Pilihan web} = \text{"Game"}) = 13/19 = 0,68.$$

c. $P(\text{Pendidikan Terakhir}=\text{"SMA"} \mid \text{Pilihan web} = \text{"Pendidikan"}) = 6/11 = 0,54.$

$$P(\text{Pendidikan Terakhir}=\text{"SMA"} \mid \text{Pilihan web} = \text{"Social Media"}) = 11/20 = 0,55.$$

$$P(\text{Pendidikan Terakhir}=\text{"SMA"} \mid \text{Pilihan web} = \text{"Game"}) = 7/19 = 0,37.$$

Selanjutnya kalikan semua atribut yang memiliki pilihan web Pendidikan, Social Media, dan Game seperti di bawah ini.

$$P(X | \text{Pilihan Web} = \text{"Pendidikan"}) = 0,64 * 0,18 * 0,54 = \mathbf{0,062}$$

$$P(X | \text{Pilihan Web} = \text{"Social Media"}) = 0,40 * 0,10 * 0,55 = \mathbf{0,022}$$

$$P(X | \text{Pilihan Web} = \text{"Game"}) = 0,42 * 0,68 * 0,37 = \mathbf{0,11}$$

Kemudian lakukan perhitungan dengan rumus Naïve Bayes

$$P(H/X) = \frac{P(X/H) \cdot P(H)}{P(X)}$$

a. $P(X|\text{Pilihan Web} = \text{"Pendidikan"}) * P(\text{Pilihan Web} = \text{"Pendidikan"})$
 $= 0,062 * 0,22 = \mathbf{0,014}$

b. $P(X|\text{Pilihan Web} = \text{"Social Media"}) * P(\text{Pilihan Web} = \text{"Social Media"})$
 $= 0,022 * 0,40 = \mathbf{0,0088}$

c. $P(X|\text{Pilihan Web} = \text{"Game"}) * P(\text{Pilihan Web} = \text{"Game"}) = 0,11 * 0,38 = \mathbf{0,042}$

Berdasarkan hasil perhitungan di atas bahwa nilai probabilitas Pilihan Web = "Game" lebih besar dibandingkan dengan nilai Pilihan Web Pendidikan dan Social media, maka dapat disimpulkan bahwa data tes di atas termasuk ke dalam **Pilihan Web = "Game"**.

II.2.4. UML

"Unified Modeling Language (UML) adalah bahasa spesifikasi UML adalah salah satu standar bahasa yang banyak digunakan di dunia

industri untuk mendefinisikan requirement, membuat analisis dan desain, serta menggambarkan arsitektur dalam pemrograman berorientasi objek (Dede Wira, 2019). *Unified Modeling Language* (UML) adalah sebuah bahasa yang berdasarkan grafik atau gambar untuk memvisualisasi, menspesifikasikan, membangun, dan pendokumentasian dari sebuah sistem pengembangan software berbasis OO (*Object-Oriented*). UML digunakan dalam pengembangan sistem perangkat lunak yang menggunakan pendekatan berorientasi objek. UML menyediakan banyak sekali diagram yang diperlukan untuk menjelaskan sistem yang sedang dikembangkan, baik dari aspek statis maupun dinamisnya (OMG, 2017). UML mendefinisikan diagram-diagram sebagai *Use case* diagram, *Class* diagram, *Activity* diagram, *Sequence* diagram, dan diagram lainnya.

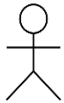




1. *Use Case Diagram*



Use case diagram merupakan pemodelan untuk kelakuan sistem informasi yang akan dibuat. *Use case* bekerja dengan mendeskripsikan tipikal interaksi antara user sebuah sistem dengan sistemnya sendiri melalui sebuah cerita bagaimana sistem itu dipakai (Dede Wira, 2019).

Setiap *use case* menyatakan perilaku yang harus dijalankan oleh sistem dalam kaitannya dengan satu atau lebih aktor (OMG, 2017).

Simbol-simbol pada diagram *use case* dapat dilihat pada tabel di bawah ini.

Tabel II.4. Simbol dan keterangan *use case diagram*

No	Simbol	Nama	Keterangan
1		<i>Actor</i>	Menspesifikasikan peran pengguna atau <i>user</i> ketika berinteraksi dengan <i>use case</i>
2		<i>Include</i>	Menspesifikasikan bahwa <i>use case</i> sumber secara eksplisit
3		<i>Generalization</i>	Hubungan dimana objek anak (<i>descendent</i>) berbagi perilaku dan struktur data dari objek yang ada di objek induk (<i>ancestor</i>)
4		<i>Extend</i>	Menspesifikasikan bahwa <i>use case</i> target memperluas perilaku dari <i>use case</i> sumber pada suatu titik yang diberikan
5		<i>Assosiation</i>	Menghubungkan antara objek satu dengan objek lainnya

No	Simbol	Nama	Keterangan
6		<i>System</i>	Menspesifikasikan paket yang menampilkan sistem secara terbatas
7		<i>Use Case</i>	Abstraksi dari interaksi antara <i>system</i> dan <i>actor</i>



(Sumber : *OMG, 2017*)

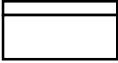


2. *Class Diagram*

Class diagram merupakan gambaran struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem. *Class diagram* terdiri dari atribut dan operasi dengan tujuan pembuat program dapat membuat hubungan antara dokumentasi perancangan dan perangkat lunak sudah sesuai (Dede Wira, 2019).

Simbol-simbol pada *class diagram* dapat dilihat pada tabel di bawah ini.

Tabel II.5. Simbol dan keterangan *Class Diagram*

No	Simbol	Nama	Keterangan
1		<i>Generalization</i>	Relasi antarkelas dengan makna umum-khusus
2		<i>Nary Association</i>	Upaya untuk menghindari asosiasi lebih dari 2 objek

3		<i>Class</i>	Kelas pada struktur sistem
4		<i>Directed Association</i>	Relasi antar kelas dengan makna kelas yang satu digunakan oleh kelas yang lain juga
5		<i>Collaboration</i>	Urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor



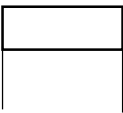
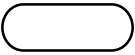


(Sumber : *OMG, 2017*)

3. *Activity Diagram*

Activity diagram merupakan diagram yang menggambarkan *workflow* atau aktivitas dari sebuah sistem yang ada pada perangkat lunak. Diagram *activity* menunjukkan aktivitas sistem dalam bentuk kumpulan aksi-aksi, bagaimana masing-masing aksi tersebut dimulai, keputusan yang mungkin terjadi hingga berakhirnya aksi. (Dede Wira, 2019).

Simbol-simbol pada *activity diagram* dapat dilihat pada tabel di bawah ini.

Tabel II.6. Simbol dan keterangan *Activity Diagram*

No	Simbol	Nama	Keterangan
1		Status Awal	Status awal aktivitas sistem
2		Status Akhir	Status akhir yang dilakukan sistem
3		<i>Swimlane</i>	Menunjukkan siapa yang bertanggung jawab dalam melakukan aktivitas pada suatu diagram
4		Aktivitas	Aktivitas yang dilakukan sistem, diawali dengan kata kerja
5		<i>Decision</i>	Kondisi percabangan dimana jika ada aktivitas lebih dari satu
6		<i>Join</i>	Penggabungan dimana lebih dari satu aktivitas digabungkan menjadi satu

(Sumber : OMG,2017)



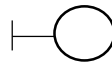


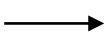
4. *Sequence Diagram*

Sequence diagram menggambarkan kelakuan objek pada *use case* dengan mendeskripsikan waktu hidup objek dan pesan yang dikirimkan dan diterima antar objek. Gambaran *sequence* diagram

dibuat minimal sebanyak pendefinisian *use case* yang memiliki proses sendiri atau yang telah didefinisikan interaksi jalannya pesan sudah dicakup pada *sequence* diagram sehingga semakin banyak *use case* yang didefinisikan, maka *sequence* diagram yang harus dibuat juga semakin banyak. (Dede Wira, 2019).

Simbol-simbol pada *sequence diagram* dapat dilihat pada tabel di bawah ini.

Tabel II.7. Simbol dan keterangan *Sequence Diagram*

No	Simbol	Nama	Keterangan
1		<i>Actor</i>	Menggambarkan orang yang sedang berinteraksi dengan sistem
2		<i>Entity Class</i>	Hubungan kegiatan yang akan dilakukan
3		<i>Boundary Class</i>	Menggambarkan sebuah penggambaran dari form
4		<i>Control class</i>	Penghubung antara <i>boundary</i> dengan tabel
5		<i>A focus of control and Lifeline</i>	Tempat mulai dan berakhirnya sebuah <i>message</i>
6		<i>Message</i>	Menggambarkan pengiriman pesan

(Sumber : OMG, 2017)