

BAB II

TINJAUAN PUSTAKA

2.1 Penelitian Terdahulu

Beberapa studi sebelumnya yang meneliti tentang prediksi BLT-DD yang dilakukan dengan algoritma *Naïve Bayes* serta *C4.5* telah menjadi referensi selama penyusunan laporan tugas akhir ini. Pada tabel 2.1 dapat ditemukan ringkasan dari hasil studi terdahulu yang digunakan untuk studi ini.

Tabel 2.1 Penelitian Terkait

No	Topik	Metode	Hasil	Referensi
1.	Penerapan Algoritma <i>C4.5</i> Dalam Memprediksi Keterlambatan Pembayaran Uang Sekolah Menggunakan Python	Pada studi ini mengimplemen tasikan algoritma Decision Tree <i>C4.5</i>	Pada studi ini menunjukkan bahwa prediksi yang dihasilkan memiliki akurasi bernilai 73%, recall bernilai 71%, dan presisi bernilai 71%	(Ginting et al., 2020)

2.	<p>Analisis Performa Algoritma Klasifikasi <i>Naive Bayes</i> dan <i>C4.5</i> untuk Prediksi Penerima Bantuan Jaminan Kesehatan</p>	<p>Dalam studi ini, menggunakan metode pengolahan data standarisasi CRISP-DM serta menggunakan algoritma <i>Naive Bayes</i> dan <i>C4.5</i>.</p>	<p>Nilai akurasi algoritma <i>Naive Bayes</i> adalah 92.97%, sedangkan algoritma klasifikasi <i>C4.5</i> memiliki nilai akurasi tertinggi sebesar 99.04%</p>	<p>(Nurfazri ah et al., 2023)</p>
3.	<p>Penerapan Metode <i>Naive Bayes</i> Dan <i>C4.5</i> Pada Penerimaan Pegawai Di Universitas Potensi Utama</p>	<p>Pada studi ini menerapkan <i>Naive Bayes</i> serta <i>C4.5</i> serta menerapkan tools Weka 3.8.</p>	<p>Hasil menyimpulkan <i>C4.5</i> menyiptakan akurasi yang lebih besar yaitu 94.4% dibandingkan <i>Naive Bayes</i> yang hanya sebesar 77.78%.</p>	<p>(Paramith a et al., 2020)</p>

4.	<p>Klasifikasi Opportunity Menggunakan Algoritma <i>C4.5</i>, <i>C4.5</i> Dan <i>Naïve Bayes</i> Berbasis Particle Swarm Optimization</p>	<p>Studi ini menggunakan algoritma <i>Naïve Bayes</i> dan <i>C4.5</i> yang berbasis PSO</p>	<p>Hasil studi menunjukkan bahwa algoritma <i>C4.5</i> dengan basis PSO mempunyai hasil akurasi, sebesar 80.90%. Algoritma <i>Naïve Bayes</i> dengan basis PSO mempunyai akurasi sebesar 83.15%.</p>	<p>(Palupi & Pahlevi, 2020)</p>
5.	<p>Perbandingan Tingkat Akurasi Klasifikasi Penerimaan Dosen Tetap Menggunakan Metode <i>Naive Bayes</i> Classifier dan <i>C4.5</i></p>	<p>Pada studi ini menerapkan metode <i>Naïve Bayes</i> dan <i>C4.5</i> serta menerapkan tools Weka 3.8.</p>	<p>Hasil studi ini mencakup perhitungan Instance Correctly Classified dan Instance Incorrectly Classified. Dari 37 data latih, metode <i>C4.5</i> membuat nilai</p>	<p>(Sadikin et al., 2020)</p>

			akurasi sebesar 91,8% dan metode <i>Naïve Bayes</i> mencapai 83,7%. Dengan demikian, metode <i>C4.5</i> dianggap lebih sesuai untuk digunakan daripada metode <i>Naïve Bayes</i> .	
6.	Analisis Perbandingan Algoritma <i>Naive Bayes</i> Dan C.45 Dalam Klasifikasi <i>Data mining</i> Untuk Memprediksi Kelulusan	Penelitian ini menggunakan literatur yang relevan dan mempelajari data saat ini. Serta menerapkan algoritma <i>Naïve Bayes</i> dan <i>C4.5</i> .	Dari studi ini memperlihatkan metode <i>Naïve Bayes</i> memiliki skor keakuratan data sebesar 94%. Sedangkan algoritma C.45 memiliki nilai pengukuran akurasi prediksi kelulusan tepat	(Tyas et al., 2021)

			waktu sebesar 92,60%.	
7.	Klasifikasi Algoritma <i>Naive Bayes</i> Dalam Memprediksi Tingkat Kelancaran Pembayaran Sewa Teras UMKM	Pada studi ini menggunakan algoritma <i>Naive Bayes</i> dan model CRISP-DM.	Pengolahan data menggunakan aplikasi Rapidminer yang menggunakan algoritma <i>Naive Bayes</i> membuat akurasi 81%, presisi 66%, dan recall sebesar 100% dengan AUC sebesar 0.800.	(Rachman & Handayani, 2021)
8.	Implementasi <i>Data mining</i> dengan Algoritma <i>Naive Bayes</i> Untuk Klasifikasi Kelayakan Penerima Bantuan Sembako	Pada studi ini memanfaatkan metode <i>Naive Bayes</i> .	Studi ini menghasilkan hasil evaluasi yang berupa akurasi senilai 86%, <i>recall</i> 85%, dan presisi 88%.	(Damuri et al., 2021)

9.	Perbandingan Algoritma <i>C4.5</i> Dan <i>Naïve Bayes</i> Untuk Menentukan Kelayakan Penerima Bantuan Program Keluarga Harapan	Studi ini memanfaatkan algoritma <i>C4.5</i> dan <i>Naïve Bayes</i> , sedangkan tools yang dipakai yaitu RapidMiner.	Pada studi ini menghasilkan nilai akurasi pada algoritma <i>C4.5</i> yaitu bernilai 91% dan <i>Naïve Bayes</i> bernilai 87%.	(Fitriani, 2020)
10.	Klasifikasi Masyarakat Penerima Bantuan Langsung Tunai Dana Desa Menggunakan <i>Naïve Bayes</i> Dan SMOTE	Metode yang digunakan pada studi ini adalah CRISP-DM serta Algoritma yang akan digunakan adalah <i>Naïve Bayes</i> dan SMOTE.	Studi ini menyimpulkan bahwa pembuatan model <i>Naïve Bayes</i> dengan menerapkan teknik SMOTE dapat menyiptakan model kinerja yang paling baik dengan menghasilkan nilai akurasi senilai 97% dan skor AUC 0,99.	(Kurniadi et al., 2022)

11.	Analisis Prediksi Masa Studi Mahasiswa Menggunakan Algoritma Naïve Bayes	Pada studi ini dilakukan dengan menerapkan algoritma <i>Naïve Bayes</i> .	Pada studi ini menghasilkan perhitungan akurasi yang bernilai 82,26%.	(Prabowo & Kodar, 2019)
12.	Perbandingan Algoritma <i>Naïve Bayes</i> Dan K- Nearest Neighbor Untuk Klasifikasi Berita Hoax Kesehatan Di Media Sosial Twitter	Metode yang akan diimplementasi kan pada studi ini diantaranya <i>Naïve Bayes</i> serta K-Nearest Neighbor	Hasil dari studi menunjukkan bahwa algoritma K- Nearest Neighbor (KNN) menciptakan skor akurasi yang tinggi daripada algoritma <i>Naïve Bayes</i> . Hasil skor akurasi yang diperoleh untuk algoritma KNN adalah 68%, sedangkan <i>Naïve Bayes</i> menghasilkan	(Krisna & Salamah, 2022)

			nilai akurasi sebesar 66%.	
13.	Implementasi Algoritma <i>Naive Bayes</i> dalam Klasifikasi Status Kesejahteraan Masyarakat Desa Gunungsari	Dalam studi ini, metode klasifikasi <i>Naive Bayes</i> digunakan untuk menghasilkan model klasifikasi yang efektif.	Dari studi yang dilakukan menunjukkan besaran akurasi yang bernilai 93% selain itu nilai akurasi error sebanyak 6,31%.	(Fauzia & Dana, 2023)
14.	Implementasi <i>Naive Bayes Classification</i> for Sentiment Analysis on Internet Movie Database	Studi ini dibangun dengan menggunakan data mentah tanpa proses lemmatization, menggunakan ukuran vektor sebesar 500, dan mengimplementasikan metode klasifikasi <i>Naive Bayes</i> .	Setelah studi ini dilakukan, ini menghasilkan sebuah model yang mencapai akurasi sebesar 78,96 persen dan f1-score sebesar 78,81 persen.	(Samsir et al., 2022)

15.	The Determination Analysis Of Telecommunications Customers Potential Cross-Selling With Classification <i>Naïve Bayes</i> And <i>C4.5</i>	Studi ini dibangun dengan mengimplementasikan <i>Naïve Bayes</i> dan <i>C4.5</i> .	Dalam studi ini, algoritma <i>C4.5</i> menyiptakan besaran akurasi yang tinggi sebesar 88,61% dengan AUC bernilai 0,870.	(Purnama sari et al., 2020)
-----	---	--	--	-----------------------------

2.2 Teori Pendukung

Berdasarkan dengan topik pada penelitian ini, berikut teori pendukung yang terkait dengan penelitian ini:

2.2.1 *Data mining*

Data Mining yaitu sebuah proses analisis dari sekelompok data yang bertujuan untuk menggali informasi yang berharga yang sebelumnya tidak diketahui. Ini merupakan pengetahuan yang mempelajari metode dan algoritma untuk memperoleh ilmu untuk menemukan sebuah pola yang belum ditemukan pada sebuah data.



Gambar 2.1 Data Mining Process

Data mining memiliki banyak aplikasi dalam mengelola sekelompok data, dengan tujuan membantu para pengambil keputusan dalam membuat keputusan terbaik. Beberapa jenis data mining termasuk *Linear Regression*, *Classification/Regression Trees*, *Neural Networks*, dan berbagai metode lainnya. Semua metode ini berfungsi untuk menganalisis data dan menghasilkan informasi yang berharga bagi pengambil keputusan.

2.2.2 Klasifikasi

Klasifikasi yaitu proses pengelompokkan beberapa kriteria menjadi sebuah *class* atau yang didasari persamaan karakteristik mereka. Metode klasifikasi ini menggunakan pendekatan *Supervised*, yang merupakan salah satu teknik dalam machine learning. Pada pendekatan *Supervised*, mesin menggunakan dataset (*data training*) yang sudah memiliki label (*labeled data*) untuk proses pembelajaran. Dengan demikian, mesin dapat mengenali label dari input dengan menggunakan fitur-fitur yang ada dan melakukan klasifikasi

berdasarkan pembelajaran yang telah dilakukan (Anggeli & Agung, 2021). Klasifikasi sendiri bertujuan untuk melakukan prediksi sebuah *class* atau *label* dari sebuah data yang tidak diketahui labelnya.

2.2.3 Bantuan Langsung Tunai Dana Desa (BLT – DD)

Sebuah penyakit epidemik yaitu Covid-19 memiliki dampak global yang signifikan, termasuk di Indonesia. Dalam konteks ekonomi, banyak masyarakat yang menghadapi PHK akibat pandemi ini, yang pada gilirannya berdampak pada kondisi ekonomi keluarga. Situasi ini berpotensi meningkatkan jumlah masyarakat yang berada dalam kondisi miskin. Untuk mengatasi masalah ini, pemerintah daerah harus menerapkan beberapa ide penting untuk menekan penyebaran COVID-19 dan mengurangi tingkat kemiskinan di Indonesia. Cara yang dapat dilakukan adalah dengan menerapkan kebijakan baru yang berkaitan dengan pandemi sebagai program jaminan sosial. Salah satu bentuk dari kebijakan ini adalah Bantuan Langsung Tunai Dana Desa, yang lebih dikenal dengan sebutan BLT-DD. Tujuan dari BLT-DD ini adalah memberikan bantuan tunai kepada masyarakat desa yang terkena dampak pandemi sebagai upaya untuk membantu mereka menghadapi situasi ekonomi yang sulit (Filki, 2022).

Bantuan tersebut diberikan kepada individu yang memenuhi beberapa kriteria tertentu, di antaranya berasal dari keluarga miskin, kehilangan pekerjaan akibat pandemi COVID-19, berusia lanjut, atau memiliki anggota keluarga yang menderita sakit parah. Bantuan sosial tersebut berupa uang tunai dan akan disalurkan setiap bulan sepanjang tahun.

2.2.4 Algoritma *Naïve Bayes*

Naïve Bayes didefinisikan sebagai salah satu metode data mining yang sangat populer dan termasuk dalam daftar sepuluh algoritma klasifikasi terpopuler. Sebagian besar orang percaya bahwa metode *Naïve Bayes* dapat mengklasifikasikan data dengan lebih baik daripada metode klasifikasi lainnya (Fahdia et al., 2020).

Metode Algoritma *Naïve Bayes* dimaksudkan dapat meringankan beban otoritas daerah pada penentuan bakal penerima BLT - Dana Desa pada Kab. Majalengka. Kelebihan dari metode *Naïve Bayes* mencakup kemampuannya untuk mengatasi data kuantitatif dan diskrit, serta tangguh dalam mengatasi titik noise yang terisolasi dengan merata-ratakan titik tersebut saat mengestimasi peluang bersyarat data. Dengan berbagai kelebihan tersebut, metode *Naïve Bayes* menjadi pilihan yang cocok untuk penelitian ini. Berikut adalah persamaan dari *Teorema Bayes* yang relevan untuk metode ini.

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

Gambar 2.2 Persamaan *Teorema Naïve Bayes*

Keterangan:

- X : Data yang belum diketahui dengan class nya
 H : Hipotesis data
 $P(H|X)$: Probabilitas hipotesis H yang mendasari kondisi X (posteriori probabilitas)
 $P(H)$: Probabilitas hipotesis H (prior probabilitas)
 $P(X|H)$: Probabilitas hipotesis X yang mendasari kondisi pada H
 $P(X)$: Probabilitas X

Gambar 2.3 Keterangan Persamaan Teorema Bayes

2.2.5 Algoritma C4.5

Algoritma C4.5 didefinisikan sebagai algoritma yang biasa diimplementasikan dalam menghasilkan pohon keputusan. Metode pohon keputusan mentransformasi data yang kompleks menjadi struktur pohon keputusan yang mewakili aturan (Alkhairi & Situmorang, 2022).

Algoritma C4.5 adalah sebuah peningkatan dari algoritma ID3. Terdapat peningkatan yang terjadi pada C4.5 memungkinkannya untuk menangani data dengan nilai yang hilang, data kontinu, dan melakukan pemangkasan (*pruning*) pada pohon keputusan (Azwanti & Elisa, 2020). Ketika membuat pohon keputusan, dalam hal memilih atribut *root* harus dilakukan berdasarkan nilai gain yang paling tinggi dari atribut yang tersedia. Dalam melakukan perhitungan gain, dapat dilakukan menggunakan perhitungan pada persamaan 1 berikut ini.

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (1)$$

Gambar 2.4 Rumus Perhitungan Gain C4.5

Keterangan:

- S : Himpunan Kasus
- A : Atribut
- N : Jumlah dari partisi atribut A
- $|S_i|$: Jumlah dari kasus partisi ke-i
- $|S|$: Jumlah kasus pada S

Gambar 2.5 Keterangan Rumus Perhitungan Gain C4.5

Jika untuk menghitung nilai entropi bisa dilakukan dengan perhitungan pada persamaan 2 berikut ini.

$$Entropy(S) = \sum_{i=1}^n - p_i * \log_2 p_i \quad (2)$$

Gambar 2.6 Rumus Perhitungan Entropy C4.5

Keterangan:

- S : Himpunan pada Kasus
- A : Fitur
- N : Jumlah dari partisi S
- p_i : Proporsi dari S_i terhadap S

Gambar 2.7 Keterangan Rumus Perhitungan Entropy C4.5