

BAB II

TINJAUAN PUSTAKA DAN LANDASAN TEORI

2.1 Penelitian terkait

Dalam penulisan proposal skripsi diperlukan beberapa referensi dari penelitian terdahulu sebagai perbandingan dengan penelitian sekarang. Penelitian terdahulu digunakan sebagai gambaran dan bahan perbandingan mengenai topik yang akan digunakan untuk penelitian sekarang. Berikut ini penelitian terdahulu yang berhubungan dengan penelitian sekarang.

Pada penelitian yang dilakukan oleh (Leonardi et al., 2021) dengan judul *Prediksi Penjualan Produk Rokok Pada PT.Indomarco Prismatama Menggunakan Algoritma C4.5* bertujuan untuk mempermudah perusahaan melakukan penggantian barang sesuai dengan permintaan. Pada penelitian ini diperoleh hasil dengan nilai akurasi tertinggi didapatkan dari pengujian dengan data 58 training dan 37 data testing, mendapatkan tingkat akurasi 92,11% tingkat kesalahan 7,894% dan nilai AUC 87,8% dan akurasi klasifikasi baik, atribut yang paling mempengaruhi penjualan tembakau adalah atribut ukuran termasuk bestseller atau underseller.

Penelitian yang telah dilakukan oleh (Yusuf et al., 2022) dengan judul *Model Prediksi Penjualan Jenis Produk Tekstil Menggunakan Algoritma K-Nearest Neighbor (KNN)*, bertujuan untuk membantu perusahaan menentukan arah dalam produksi barang serta penjualan dimasa depan. Hasil yang diperoleh pada penelitian ini menghasilkan nilai k-1 dengan akurasi 86.9% dengan RMSE sebesar 0.362 dan diperoleh nilai class recall dan class precision pada confusion matrix yang didapatkan dengan nilai cukup baik sehingga model algoritma yang digunakan dalam penelitian ini bekerja dengan baik.

Penelitian yang telah dilakukan oleh (M & Jayyidah, 2022) dengan judul *Prediksi Penjualan Barang Pada Toko Baby Shop Dengan Algoritma Single Moving Average (SMA)*, bertujuan untuk memudahkan dalam memprediksi penjualan di Toko Baby Shop dan dapat mengetahui perhitungan prediksi

dengan menerapkan metode Single Moving Average. Hasil dari penelitian yang telah dilakukan adalah dapat memprediksi setiap item penjualan barang pada bulan berikutnya dengan hasil rata-rata akurasi 90%.

Penelitian yang telah dilakukan oleh (Wulandari & Desyanti, 2022) dengan judul Implementasi Algoritma *K-Nearest Neighbor* Dalam Memprediksi Stok Sepeda Motor. Penelitian ini bertujuan untuk membantu perusahaan dalam mengetahui suatu kecenderungan jumlah penjualan sepeda motor ditahun berikutnya. Pada penelitian kali ini penulis menggunakan metode *K-Nearest Neighbor* yang mana metode ini biasa digunakan sebagai analisis klasifikasi namun pada beberapa tahun ini metode KNN juga diterapkan untuk prediksi. Hasil yang diperoleh pada penelitian ini adalah memperoleh jumlah sepeda motor setiap merek yang nantinya akan terjual ditahun 2022 dan memperoleh tambahan 5 unit sepeda motor untuk setiap merek sepeda motor. Dari hasil penelitian tersebut dengan menggunakan metode *K-Nearest Neighbor* didapatkan nilai akurasi sebesar 97%.

Penelitian yang telah dilakukan oleh (Nugroho et al., 2022) dengan judul Perancangan Sales Prediction Model Menggunakan Metode Neural Network. Penelitian ini bertujuan untuk memberikan kontribusi dalam menerapkan suatu model regresi sales untuk melakukan prediksi menggunakan metode *Neural Network* yang dapat digunakan sebagai alat bantu untuk pengambilan keputusan dalam menjual jenis produk yang diminati diberbagai macam segment di toko online. Pada penelitian ini diperoleh hasil kinerja yang diukur dengan Mean Squared Error (MSE) sebesar 0.831, Root Mean Square Error (RMSE) sebesar 0,911 dan Mean Absolute Error (MAE) sebesar 0,650.

Penelitian yang telah dilakukan oleh (Maskuri et al., 2022) dengan judul Penerapan Algoritma *K-Nearest Neighbor* (KNN) Untuk Memprediksi Penyakit Stroke. Penelitian ini bertujuan menerapkan algoritma *K-Nearest Neighbor* dalam melakukan prediksi pada penyakit stroke dan dapat mengetahui akurasi yang dihasilkan. Hasil yang diperoleh dalam penelitian ini dengan menerapkan metode KNN ini memperoleh akurasi 95% dengan nilai $k=9$.

Penelitian yang telah dilakukan oleh (Suparman et al., 2021) yang berjudul *Prediksi Barang Keluar TB. Wijaya Bangunan Menggunakan Algoritma KNN Regression dengan RStudio* bertujuan untuk membantu memudahkan TB.Wijaya Bangunan dalam mengelola barang persediaan berdasarkan data yang telah ada sebelumnya agar selisih jumlah antara barang masuk dan keluar dapat diperkecil. Pada penelitian ini penulis menggunakan algoritma *Time Series KNN Regression* untuk melakukan prediksi terhadap jumlah barang keluar sesuai data deret waktu dengan data yang telah terjadi sebelumnya. Hasil yang diperoleh dengan menerapkan algoritma *Time Series KNN Regression* yaitu hasil prediksi yang paling banyak keluar terjadi pada periode minggu ke-9 sebanyak 22.40%. kesimpulan yang diperoleh dari hasil penelitian tersebut menghasilkan nilai evaluasi *Root Mean Square Error (RMSE)* paling kecil 3.55 ini berarti memiliki hasil akurasi prediksi yang baik.

Penelitian yang telah dilakukan oleh (Wijayanto & Tengah, 2021) dengan judul *Forecasting Penjualan Gas LPG di Toko Sembako Menggunakan Metode Fuzzy Time Series*. Penelitian ini bertujuan untuk membantu pihak toko sembako dalam memprediksi penjualan produk dalam mengontrol barang masuk maupun barang keluar. peramalan data pada penelitian ini menggunakan dua teknik peramalan yaitu analisis kualitatif dan analisis kuantitatif, dan untuk perhitungan dari nilai error hasil dari peramalan penelitian ini menggunakan *MAPE* yang menghasilkan nilai error 0,3% karena menyatakan persentase kesalahan yang kecil maka hasil akurasinya lebih akurat.

Penelitian yang telah dilakukan oleh (Muttaqin & Srihartini, 2022) dengan judul *Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat Jenis Tablet*. Penelitian ini bertujuan untuk memprediksi persediaan obat tablet. Hasil perhitungan yang didapatkan dengan metode regresi linier didapatkan akurasi sebesar 98,505% yang mana metode ini bisa diimplementasikan untuk memprediksi stok obat jenis tablet pada klinik teluk.

Penelitian yang telah dilakukan oleh (Alfani et al., 2021) dengan judul *Prediksi Penjualan Produk Unilever Menggunakan Metode K-Nearest Neighbor*. Penelitian ini bertujuan untuk memudahkan pihak toko dalam

melakukan perencanaan stok produk pada masa mendatang. Pada penelitian ini, untuk mendapatkan informasi mengenai penjualan produk terlaris maka menggunakan Teknik klasifikasi data mining dengan menerapkan algoritma *K-Nearest Neighbor*. Hasil yang didapatkan dari penelitian ini berdasarkan nilai akurasi tertinggi pada klasifikasi penjualan produk sebesar 86,66% dan mendapatkan nilai akurasi terendah pada klasifikasi penjualan produk sebesar 40%. Berdasarkan hasil penelitian yang diperoleh maka dapat disimpulkan bahwa algoritma *K-Nearest Neighbor* ini dapat diimplementasikan untuk memprediksi penjualan produk unilever pada toko Rizky Barokah Nganjuk.

Tabel 2. 1Penelitian Terdahulu

NO	Penelitian	Judul	Dataset	Metode	Evaluasi
1	(Leonardi et al., 2021)	Prediksi Penjualan Produk Rokok Pada PT.Indomarco Prismatama Menggunakan Algoritma C4.5	Dataset penjualan produk PT.Indomarco Prismatama	C4.5	Accuracy 92,11% AUC 0,878%
2	(Yusuf et al., 2022)	Model Prediksi Penjualan Jenis Produk Tekstil Menggunakan Algoritma <i>K-Nearest Neighbor</i>	Perusahaan tekstil 2019 Padalarang, Bandung Barat	K-Nearest Neighbor	Accuracy 86.9% RMSE 0.362

3	(M Jayyidah, 2022)	& Prediksi Penjualan Barang Pada Toko Baby Shop Dengan Algoritma <i>Single Moving Average</i> (SMA)	Dataset penjualan juli 2021-januari 2022 Toko Baby Shop	Single Moving Average (SMA)	Accuracy 90%
4	(Wulandari & Desyanti, 2022)	Implementasi Algoritma <i>K-Nearest Neighbor</i> Dalam Memprediksi Stok Sepeda Motor	Dataset penjualan dari tahun 2019, 2020, dan 2021 PT.Dasatama Cemerlang Motor	K-Nearest Neighbor	Accuracy 97%
5	(Nugroho et al., 2022)	Perancangan Sales Prediction Model Menggunakan Metode <i>Neural Network</i>	Dataset sales conversion optimization yang diunduh dari kaggle	Neural Network	Mean Squared Error (MSE) = 0.831 Root Mean Square Error (RMSE) = 0.911 Mean Absolute Error (MAE) = 0.650
6	(Maskuri et al., 2022)	Penerapan Algoritma <i>K-Nearest</i>	Dataset Stroke Disease	K-Nearest Neighbor	Pada 100 data strok yang digunakan

Neighbor diambil dari dan dibagi
(KNN) untuk dataset public menjadi data
Memprediksi yang latih dan data
Penyakit berjumlah 100 uji dengan
Stroke data metode split
validation
menggunakan
perbandingan
80% : 20%.
Nilai k yang
digunakan
dalam
menguji
kinerja
algoritma
KNN untuk
memprediksi
penyakit
strok yaitu k-
9, sehingga
didapatkan
nilai akurasi
sebesar 95%.

7	Prediksi Barang Keluar TB.Wijaya Bangunan Menggunakan Algoritma KNN Regresion	Data transaksi TB.Wijaya Bangunan pada 1 januari sampai 31 desember 2019	KNN Regression	Root Mean Square Error (RMSE) paling kecil 3.55 yang artinya memiliki hasil akurasi
---	---	--	-------------------	--

		dengan RStudio			prediksi terbaik.
8	(Wijayanto & Tengah, 2021)	Forecasting Penjualan Gas LPG di Toko Sembako Menggunakan Metode <i>Fuzzy Time Series</i>	Data mentah penjualan gas selama 5 bulan dari bulan maret, april, mei, juni, dan juli	Fuzzy Time Series	Dalam perhitungan nilai eror dari hasil penelitian ini menggunakan MAPE maka menghasilkan nilai eror 0,3%9
9	(Muttaqin & Srihartini, 2022)	Penerapan Metode Regresi Linier Sederhana Untuk Prediksi Persediaan Obat jenis Tablet	Data transaksi penjualan yang dicatat dalam buku besar klinik Teluk	Regresi Linear	Diperoleh accuracy 98,505%
10	(Dewi & Rahayu, 2022)	Prediksi Penjualan Produk Unilever Menggunakan Metode <i>K-Nearest Neighbor</i>	Data penjualan Toko Rizky Barokah produk Unilever, P&G, Nestle, dan Indofood dari tahun 2017, 2018, dan 2019	K-Nearest Neighbor	Hasil penelitian ini berdasarkan pada nilai tertinggi pada klasifikasi penjualan produk yakni sebesar 86,66% dan nilai akurasi

terendah pada
klasifikasi
penjualan
produk yakni
sebesar 40%

Berdasarkan penelitian terdahulu dapat disimpulkan bahwa penelitian sebelumnya sama-sama melakukan prediksi namun dengan metode yang berbeda serta menghasilkan nilai akurasi yang berbeda-beda dan objek penelitian yang berbeda. Setiap penelitian memiliki kekurangan dan kelebihan masing-masing, dari penelitian sebelumnya metode yang paling cocok untuk penelitian saat ini adalah metode *K-Nearest Neighbor* karena berdasarkan penelitian terdahulu metode *K-Nearest Neighbor* dapat melakukan prediksi dengan baik dengan tipe data penjualan serta objek yang diteliti pada penelitian kali ini sehingga pada penelitian ini diharapkan mampu memberikan hasil yang lebih baik dan tingkat akurasi yang lebih tinggi dari metode sebelumnya.

2.2 Landasan Teori

2.2.1 Penjualan

Penjualan merupakan suatu kegiatan berusaha yang dapat memberikan suatu keuntungan lebih bagi sebuah usaha dagang dengan penjualan barang maupun jasa pada konsumen dalam jumlah ecer atau satuan (Dewi & Rahayu, 2022). Penjualan adalah suatu kegiatan interaksi antara individu dimana individu ini saling bertatap muka untuk menciptakan, memperbaiki, menguasai dan mempertahankan pertukaran sehingga saling menguntungkan antar pihak. Penjualan juga diartikan sebagai upaya yang dilakukan manusia untuk memberikan barang atau jasa bagi mereka yang dibutuhkan dengan imbalan uang sesuai dengan harga barang yang telah ditentukan sebagai kesepakatan kerjasama (M & Jayyidah, 2022).

2.2.2 Prediksi

Prediksi merupakan suatu kegiatan yang meramalkan suatu data penjualan dimasa yang akan datang dalam hal ini berarti menentukan pemikiran

besarnya volume suatu data penjualan, bahkan dapat menentukan potensi suatu penjualan dan luas pasar yang akan dikuasai dimasa yang akan datang. Hal ini juga dapat membantu perusahaan dalam melakukan prediksi penyediaan stok barang, karena prediksi ini dapat memberikan output yang baik bagi perusahaan sehingga perusahaan dapat meminimalisir kesalahan dalam perencanaan (Dewi & Rahayu, 2022). Prediksi merupakan suatu proses untuk memperkirakan secara sistematis mengenai sesuatu yang paling mungkin terjadi dimasa mendatang berdasarkan informasi masa lalu dan sekarang yang dimiliki, untuk meminimalisir kesalahan yang terjadi. Prediksi tidak selalu memberikan jawaban yang pasti terhadap kejadian yang akan terjadi, namun dalam prediksi ini berusaha mendapatkan jawaban yang sedekat mungkin terhadap masalah yang terjadi (M & Jayyidah, 2022).

2.2.3 Data Mining

Data mining merupakan suatu Teknik pembelajaran mesin, pengenalan pola, dataset, statistic, dan visualisasi yang dapat digunakan untuk penanganan masalah pengambilan informasi dari penyimpanan database yang besar (Alfani et al., 2021). Menurut mujiasi Data Mining atau sering disebut sebagai knowledge discovery in database (KDD) merupakan suatu kegiatan yang meliputi pengumpulan, penggunaan data historis untuk menemukan keteraturan, pola atau hubungan pada data yang berukuran besar. Dengan adanya data mining ini bisa digunakan untuk membantu dalam pengambilan keputusan dimasa mendatang (Wulandari & Desyanti, 2022).

2.2.4 Klasifikasi

Klasifikasi adalah suatu pekerjaan untuk menilai objek data untuk memasukkan kedalam kelas tertentu dari jumlah kelas yang tersedia pada data (Utomo & Mesran, 2020). Klasifikasi ini melakukan pembangunan model berdasarkan data latih yang tersedia, kemudian menggunakan data latih yang tersedia untuk mengklasifikasikan data yang baru. Klasifikasi dapat didefinisikan sebagai suatu pekerjaan yang melakukan pelatihan atau pembelajaran terhadap suatu fungsi target yang memetakan setiap set atribut (fitur) pada jumlah label kelas yang tersedia. Suatu sistem yang melakukan klasifikasi diharapkan mampu melakukan klasifikasi terhadap semua data set

dengan benar, namun tidak dapat dipungkiri bahwa kinerja suatu sistem tidak bisa 100% benar, sehingga sistem klasifikasi juga harus diukur kinerjanya. Pada umumnya pengukuran kinerja klasifikasi dilakukan dengan *confusion matrix*. Tujuan dari klasifikasi merupakan untuk memperkirakan kelas dari suatu objek yang labelnya belum diketahui (Junaedi et al., 2019).

2.2.5 K-Nearest Neighbor

K-Nearest Neighbor (KNN) merupakan algoritma yang paling sederhana dalam melakukan proses penyelesaian masalah klasifikasi (Batlayeri et al., 2022). Algoritma *K-Nearest Neighbor* merupakan suatu metode yang digunakan untuk mengklasifikasikan suatu objek berdasarkan data pembelajaran dari jarak paling dekat dengan objek tersebut (Wulandari & Desyanti, 2022). Berdasarkan pengertian *K-Nearest Neighbor* tersebut. KNN merupakan suatu metode klasifikasi pada sekumpulan data berdasarkan pembelajaran data, algoritma knn ini termasuk dalam supervised learning, yang melakukan klasifikasi berdasarkan dengan kedekatan jarak suatu data dengan data yang lain. Konsep dasar dari KNN adalah untuk mencari jarak terdekat antara data yang akan dievaluasi dengan k tetangga terdekatnya pada data pelatihan. KNN ini memiliki keunggulan seperti akurasi yang tinggi, tidak ada asumsi data, algoritma ini juga memiliki kekurangan, seperti penentuan nilai k yang optimal, memori yang tinggi, dan komputasi yang mahal.

Dibawah ini merupakan rumus Euclidean distance:

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (\text{II.1})$$

keterangan :

- 1) $d(x, y)$ merupakan jarak data x dan data y.
- 2) x_i merupakan data testing ke-i
- 3) y_i merupakan data training ke-i

2.2.6 Confusion Matrix

Confusion matrix merupakan suatu alat ukur berupa matriks digunakan untuk menentukan akurasi klasifikasi dari suatu kelas sesuai dengan algoritma yang digunakan (Qadrini et al., 2021). *Confusion matrix*

menjelaskan mengenai informasi yang digunakan untuk membandingkan hasil klasifikasi yang dibuat oleh sistem dengan hasil yang seharusnya (Astiningrum et al., 2020).

Tabel 2. 2 Confusion Matrix

<i>Confusion Matrix</i>		<i>Actual Values</i>	
		<i>True</i>	<i>False</i>
<i>Predicted Values</i>	<i>True</i>	TP	FP
	<i>False</i>	FN	TN

Keterangan :

TP : True Positive

FP : False Positive

TN : True Negative

FN : False Negative

Berdasarkan tabel Confusion Matrix, dapat diketahui jika label pada prediksi bernilai bernilai true dan label sebenarnya bernilai false maka disebut FP (False Positive). Jika label pada prediksi bernilai false dan tabel yang sebenarnya bernilai true maka disebut FN (False Negative). Dari pernyataan tersebut dapat disimpulkan bahwa nilai TP (True Positive) dan TN (True Negative) menunjukkan ketepatan klasifikasi. Semakin tinggi nilai TP dan FN maka semakin baik tingkat klasifikasi dalam hal presisi, akurasi dan recall.

Confusion matrix terdapat 3 nilai yang digunakan untuk mengukur kemampuan pada sistem klasifikasi yang dibangun yaitu presisi, akurasi, dan recall (Azhari et al., 2021).

1. Presisi

Presisi menunjukkan tingkat ketepatan atau ketelitian dalam pengklasifikasian.

$$\text{Presisi} = \frac{TP}{FP+TP} \times 100\% \quad (\text{II.2})$$

2. Akurasi

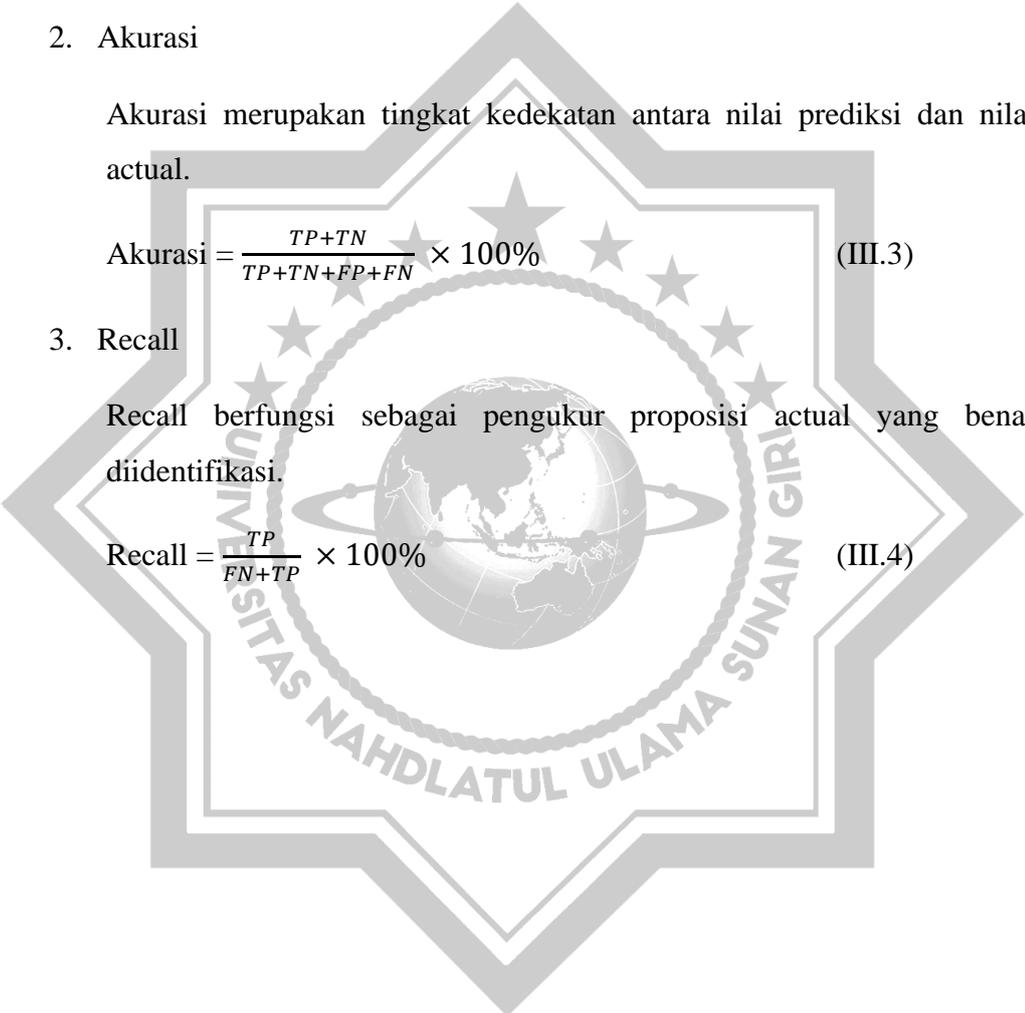
Akurasi merupakan tingkat kedekatan antara nilai prediksi dan nilai actual.

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100\% \quad (\text{III.3})$$

3. Recall

Recall berfungsi sebagai pengukur proporsi actual yang benar diidentifikasi.

$$\text{Recall} = \frac{TP}{FN+TP} \times 100\% \quad (\text{III.4})$$



UNUGIRI