

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Kajian jurnal yang pertama diambil dari jurnal yang ditulis oleh Odi Nurdiawan dan Noval Salim dari Program Studi Informatika di STMIK IKMI Cirebon pada tahun 2018 yang berjudul Penerapan Data Mining 7 Penjualan Barang Menggunakan Metode Metode Naive Bayes Classifier untuk Optimasi Strategi Pemasaran. Jurnal ini membahas tentang bagaimana penerapan metode naive bayes classifier pada data penjualan dan berbagai transaksi penjualan untuk menciptakan pengetahuan baru yang kemudian digunakan untuk optimasi strategi pemasaran produk agar bisa tercapai dengan maksimal. Tujuannya yakni sebagai salah satu strategi pemasaran untuk menghadapi pesaing bisnis lainnya dan juga bisa menambah omset penjualan perusahaan. Dari hasil penerapan metode naive bayes classifier pada data set penjualan sejumlah 120 record barang mencapai tingkat optimasi sebesar lebih dari 65%.

Kajian jurnal yang kedua diambil dari jurnal yang ditulis oleh Herry Derajad Wijaya dan Saruni Dwiasnati dari Fakultas Ilmu Universitas Mercu Buana Jakarta pada tahun 2020 yang berjudul Implementasi Data Mining dengan Algoritma Naive Bayes pada Penjualan Obat. Jurnal ini membahas tentang pengolahan data yang ada pada beberapa apotek yang ada di Indonesia masih belum dapat mengetahui dengan pasti produk apa saja yang laku atau tidak laku pada bulan sebelumnya agar mengetahui

produk mana yang harus di stok lebih banyak lagi untuk bulan selanjutnya. Tujuannya yakni mengusulkan hasil accuracy baru untuk sebuah permasalahan memprediksi Produk laku atau tidak laku pada bulan sebelumnya untuk mengetahui pola minat pelanggan terhadap pembelian vitamin pada sebuah apotek. Data yang digunakan dalam penelitian ini di ambil pada tahun 2018 bulan Januari-Mei, data yang di ambil kisaran 5.000 dan data testing untuk penelitian ini penulis hanya mengambil sampel data sebanyak 137. Dari Pengujian pada data rekapitulasi penjualan Obat dengan proses mining Algoritma Naïve Bayes menghasilkan tingkat accuracy dengan nilai 88.00%.

Kajian jurnal yang ketiga diambil dari jurnal yang ditulis oleh Ikhsan Romli, Esem Pusnawati dan Arif Siswandi dari Universitas Pelita Bangsa pada tahun 2019 yang berjudul Penentuan Tingkat Penjualan Mobil di Indonesia dengan Menggunakan Algoritma *Naive Bayes*. Jurnal ini membahas tentang penentuan tingkat penjualan mobil di Indonesia dengan melakukan klasifikasi pembeda merek mobil yang paling banyak diminati para konsumen dan kategori yang paling laris. Tujuannya ialah mengklasifikasi data dengan membentuk suatu class yaitu laris dan tidak laris, sehingga para konsumen, produsen, dan peneliti dapat mengetahui merek mobil manakah yang paling laris berdasarkan kategori maupun output-nya. Data yang dianalisis sebanyak 639 data dengan 511 data training dan 128 data testing. Tingkat akurasi klasifikasi dengan metode *naive bayes* menghasilkan nilai accuracy sebesar 92.19%, nilai

precision 98.39% dan nilai recall 87.14% sehingga metode *naive bayes* merupakan metode yang cukup baik dalam penelitian ini.

Kajian jurnal yang keempat diambil dari jurnal yang ditulis oleh Indah Purnamasari dari Sekolah Tinggi Manajemen Informatika dan Komputer Nusa Mandiri pada tahun 2018 yang berjudul *Klasifikasi Pelanggan Produk Indihome Menggunakan Naive Bayes Classifier Dengan Seleksi Fitur Algoritma Genetik*. Jurnal ini membahas tentang dilakukannya 8 pengklasifikasian terhadap pelanggan telepon dan pelanggan speedy eksisting yang berpotensi untuk menjadi pelanggan produk indihome sebagai salah satu strategi pemasaran berupa menawarkan produk baru indihome dari PT. Telkom Jakarta agar lebih efisien dan efektif dalam pemasaran produk. Tujuannya ialah sebagai salah satu strategi pemasaran produk guna mengurangi kemungkinan penolakan dari para pelanggan yang diakibatkan dari penawaran yang terlalu banyak maka dilakukan klasifikasi terhadap pelanggan telepon dan pelanggan speedy eksisting yang berpotensi untuk menjadi pelanggan produk Indihome. Dari 878 data pelanggan telepon dan akurasi yang dihasilkan model naive bayes classifier yaitu 85.08%.

Kajian jurnal yang kelima diambil dari jurnal yang ditulis oleh Pathyatus Sa'adah dari Program Studi Teknik Informatika Fakultas Teknik Universitas Pelita Bangsa pada tahun 2019 yang berjudul *Klasifikasi Data Penjualan Alat Tulis Kantor (ATK) Terlaris Untuk Optimasi Strategi Pemasaran Di Toko Citramedia Menggunakan Metode Naive Bayes*. Skripsi ini membahas tentang pengolahan data penjualan ATK untuk mengklasifikasi data ATK dan mencari nilai akurasi sebagai penentuan barang

apa saja yang paling banyak diminati *customer*. Tujuan dilakukannya penelitian ini yakni sebagai suatu strategi pemasaran agar dapat mengoptimalkan kerja toko dalam mengolah data stok barang yang tersedia sehingga dapat memenuhi kebutuhan customer dan dapat membantu meningkatkan keuntungan penjualan. Data yang ada dianalisis menggunakan metode Algoritma *Naive Bayes Classifier* dan pengujiannya diimplementasikan pada tools RapidMiner agar mendapatkan tingkat akurasi yang efektif dan efisien. Dari hasil pengujian yang dilakukan, didapatkan tingkat accuracy tertinggi sebesar 99.26% dan accuracy terendah sebesar 94.39% sedangkan evaluasi dengan kurva AUC hasil dari empat kali pengujian yang telah dilakukan termasuk dalam kategori klasifikasi sangat baik (*excellent classification*) yakni dengan evaluasi tertinggi sebesar 0.997.

Perbandingan dengan penelitian sebelumnya, yang terkait dengan penelitian yang akan dilakukan dapat dilihat pada tabel 1.

Tabel 2.1 Tinjauan Pustaka

PENULIS	METODE	OBJEK	HASIL
Odi Nurdiawan dan Noval Salim (2018)	<i>Naive Bayes Classifier</i>	Penjualan barang untuk optimasi strategi pemasaran	Dari hasil penerapan metode naive bayes classifier pada dataset penjualan barang mencapai tingkat optimasi sebesar lebih dari 65%.

Tabel Lanjutan

Herry Derajad Wijaya dan Saruni Dwiasnati (2020)	<i>Naive Bayes Classifier</i>	Penjualan obat di Apotek	Dari hasil penelitian ini menggunakan tools Rapidminer versi 8 didapat nilai ROC dan hasil accurasy 88.00%
Ikhsan Romli, Esem Pusnawati dan Arif Siswandi (2019)	<i>Naive bayes</i>	Penjualan Mobil di Indonesia	Tingkat akurasi klasifikasi dengan metode naive bayes menghasilkan nilai accuracy sebesar 92.19%, nilai precision 98.39% dan nilai recall 87.14%.
Indah Purnamasari (2018)	<i>Naive Bayes Classifier</i>	Produk Indihome	Akurasi yang dihasilkan model naive bayes classifier yaitu 85.08%.
Pathyatus Sa'adah (2019)	<i>Naive bayes</i>	Alat Tulis Kantor (ATK) di toko Citramedia	Hasil accuracy tertinggi 99.26% dan terendah 94.39% sedangkan evaluasi kurva AUC sebesar 0.997.

2.2 Dasar Teori

2.2.1 E-Commarce

E-Commerce merupakan aktivitas jual beli barang atau jasa melalui internet (Ferraro, 1998). *E-commerce* dapat dikelompokkan menjadi dua bagian, yaitu *business to business* (B2B) dan *business to customer* (B2C). B2B *e-commerce* adalah bentuk transaksi melalui internet yang dilakukan oleh dua perusahaan atau lebih, sedangkan B2C *e-commerce* adalah transaksi jual beli melalui internet antar penjual dan pembeli (*end user*) (Ustadiyanto, 2001:11).

Transaksi yang melibatkan lebih sedikit orang yaitu B2B. Yang terlibat dalam transaksi B2B biasanya adalah orang yang telah terlatih dalam menggunakan sistem informasi dan terbiasa dengan proses bisnis transaksi. Transaksi yang dilakukan pada B2B yaitu dalam bentuk *electronic data interchange* (EDI), dan transaksi ini biasanya dilakukan dengan *supplier/vendor* (Ferraro, 1998).

2.2.2 Shopee

Shopee adalah platform *e-commarce* terkemuka di Asia Tenggara dan Taiwan. Shopee memiliki beberapa layanan seperti *e-commerce* yang mempertemukan penjual dan pembeli melalui perluasan jangkauan pelanggan online dan dapat memastikan barang sampai ke tangan pelanggan dalam keadaan baik dan cepat. Shopee diluncurkan pada tahun 2015 sebagai platform yang disesuaikan untuk wilayah tersenut, memberi

pelanggan pengalaman belanja yang mudah, aman dan cepat melalui pembayaran yang kuat dan dukungan penuh (Shopee, 2019).

2.2.3 Data Mining

Data mining didefinisikan sebagai proses menemukan pola-pola dalam data. Proses ini otomatis atau seringnya semiotomatis. Pola yang ditemukan harus penuh arti dan pola tersebut memberikan keuntungan, biasanya keuntungan secara ekonomi. Data yang dibutuhkan dalam jumlah besar (Witten, 2005).

Data mining digunakan untuk mencari pengetahuan yang ada dalam *database* sehingga sering disebut *Knowledge Discovery Database (KDD)*. KDD berhubungan dengan teknik integrasi dan penemuan ilmiah, interpretasi dan visualisasi dari pola-pola sejumlah data. Serangkaian proses tersebut memiliki tahap sebagai berikut (R. A. Yulia, 2018) :

1. *Data Selection.*

Melakukan pemilihan himpunan data, menciptakan himpunan data target, atau melakukan fokus subset pada variabel (sampel data) yang penemuan (discovery). Hasil seleksi akan disimpan dalam sebuah berkas terpisah pada basis data operasional.

2. *Pre-Processing dan Cleaning Data.*

Secara umum, tugas-tugas dalam *data mining* dapat diklasifikasikan menjadi 3 bagian yaitu: *Assosiation*, *Klasifikasi* dan *Clustering*. *Clustering* atau klusterisasi adalah metode pengelompokan data.

2.2.4 Klasifikasi

Klasifikasi merupakan salah satu cara untuk menganalisis data yang diekstrak menggunakan model agar dapat menggambarkan kelas dari data *instance* berdasarkan atribut yang dimiliki. Atribut-atribut data dapat berupa data kategorikal, ordinal, integer-value, dan real-value (Supriyadi, 2017).

Klasifikasi diterapkan untuk pertama kalinya yaitu dalam tanaman untuk mengklasifikasi spesies tertentu, seperti yang dilakukan oleh Carolus von Linne yang pertama kali melakukan pengklasifikasian spesies berdasarkan karakter fisik dan dikenal sebagai bapak klasifikasi (Mardi, 2019).

2.2.5 *Naive Bayes Classifier*

Naïve Bayes Classifier merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes (aturan bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat (naif). Dengan kata lain, dalam *Naive Bayes* model yang digunakan adalah model fitur independen (Prasetyo, 2012). *Teorema Bayesian* menghitung nilai *posterior probability* $P(H|X)$ menggunakan probabilitas $P(H)$, $P(X)$, dan $P(X|H)$. Dimana nilai X adalah data testing yang kelasnya

belum diketahui. Nilai H adalah hipotesis data X yang merupakan suatu kelas yang lebih spesifik. Nilai $P(X|H)$ atau disebut *likelihood*, adalah probabilitas hipotesis X berdasarkan kondisi H. Sedangkan nilai $P(X)$ yang disebut juga dengan *predictor prior probability*, adalah probabilitas X (Suntoro, J., (2019))

$$P(H|X) = \frac{P(X|H).P(H)}{P(X)} \dots\dots\dots(2.1)$$

- X : Data dengan class yang belum diketahui
 H : Hipotesis data X yang merupakan suatu class spesifik
 $P(H|X)$: Probabilitas hipotesis H berdasarkan kondisi x (*posteriori probability*)
 $P(H)$: Probabilitas hipotesis H (*prior probability*)
 $P(X|H)$: Likelihood atau Probabilitas X berdasarkan kondisi H
 $P(X)$: Probabilitas dari X (*predictor prior probability*)

Berdasarkan persamaan teorema *Naive Bayes* diatas, dapat dilakukan proses klasifikasi dengan data kontinyu dengan menggunakan rumus Densitas Gaus :

$$g(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}.\sigma} \exp \frac{-(x-\mu)^2}{2\sigma^2} \dots\dots\dots(2.2)$$

Dimana nilai rata-rata μ dihitung terlebih dahulu menggunakan rumus

$$\mu = \frac{\sum_{i=1}^n x_i}{n} \dots\dots\dots(2.3)$$

sesuai persamaan dan standar deviasi σ

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}} \dots\dots\dots(2.4)$$

Tahapan algoritma *Naive Bayes* menurut (Suroto, J., 2019), yaitu sebagai berikut:

1. Siapkan data set
2. Menghitung jumlah kelas/label pada data training
3. Menghitung jumlah kasus yang sama dengan kelas yang sama
4. Kalikan semua hasil sesuai dengan data testing yang akan dicari kelasnya
5. Bandingkan hasil per kelas, nilai tertinggi ditetapkan sebagai kelas terbaru.

2.2.6 Evaluasi Model

Fungsi dari *confusion matrix* ialah untuk mengukur tingkat *accuracy*, *precision*, dan nilai *recall* dari suatu model algoritma yang dievaluasi. Nilai *accuracy* merupakan tingkat ketepatan presentase antara nilai prediksi dan nilai sebenarnya. Kemudian nilai *precision* adalah nilai akurasi dengan *class* yang telah diprediksi. Sedangkan nilai *recall* merupakan presentase nilai kinerja keberhasilan algoritma yang dipakai.

2.2.7 JSON (Java Script Object Notation)

JSON (*Jawa Script Object Notation*) adalah format pertukaran data yang bersifat ringan, disusun oleh Douglas Crockford. Fokus JSON adalah pada representasi data di *website*. JSON dirancang untuk memudahkan pertukaran data pada situs dan merupakan perluasan dari fungsi-fungsi *javascript*.

2.2.8 Crawling

Crawling merupakan teknik mengumpulkan data pada sebuah website dengan memasukkan *Uniform Resource Locator* (URL). URL ini menjadi acuan untuk mencari semua hyperlink yang ada pada website. Kemudian dilakukan indexing untuk mencari kata dalam dokumen pada setiap link yang ada.

Untuk penerapannya crawling menggunakan automation program dan menggunakan *Application Programming Interface* (API) sebagai jalur komunikasi dalam mendapatkan data. Dengan API kita dapat mengumpulkan data lebih spesifik sesuai dengan link URL yang ada tanpa harus mengetahui element HTML pada sebuah website.

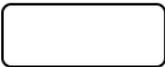
2.2.8 PHP

PHP atau yang memiliki kepanjangan PHP Hypertext Preprocessor, merupakan suatu bahasa pemrograman yang difungsikan untuk membangun suatu website dinamis. PHP menyatu dengan kode HTML. HTML digunakan sebagai pembangun atau pondasi dari kerangka layout web, sedangkan PHP difungsikan sebagai prosesnya, sehingga dengan adanya PHP tersebut, sebuah web akan sangat mudah di maintenance (Saputra, 2013). PHP berjalan pada sisi server sehingga PHP disebut juga sebagai bahasa Server Side Scripting, artinya bahwa dalam setiap / untuk menjalankan PHP, wajib membutuhkan web server dalam menjalankannya (Saputra, 2013).

2.2.9 Diagram Alur Data

Data Flow Diagram (DFD) atau disebut juga dengan Diagram Aliran Data (DAD) merupakan diagram yang menggambarkan suatu aliran data suatu sistem. Alur yang digambarkan adalah alur input – proses – output, biasa digunakan untuk menjelaskan proses – proses pada sistem informasi mencakup pula aliran informasi yang keluar dan masuk ke sistem. DFD memiliki tingkatan tertentu dimana tingkatan ini disebut dengan level. Dimulai dari DFD level 1, level 2, level 3 dan seterusnya. Proses penurunan level, misalnya dari level 1 ke level 2 disebut dengan breakdown. Semakin besar tingkatan level, berarti semakin detail suatu proses tersebut dijelaskan. DFD merupakan turunan dari Diagram Konteks, dimana Diagram Konteks juga sering disebut dengan DFD level 0 (Sulianta, 2017). Adapun simbol DAD ditunjukkan pada tabel 2.2

Tabel 2.2 Simbol DAD

Simbol	Nama simbol	Keterangan
	<i>Process / Proses</i>	Menggambarkan suatu proses atau sistem yang akan dibangun.
	<i>Entity / Entitas</i>	Menggambarkan entitas atau pengguna dari sistem atau aplikasi.

Tabel Lanjutan

	<i>Data Flow / Aliran data</i>	Aliran data yang masuk dan keluar dari sistem.
	<i>Data Store / Penyimpanan data</i>	Menggambarkan suatu tabel untuk menyimpan data, dimana nantinya <i>data store</i> ini akan menjadi salah satu tabel dalam perancangan basis data.

2.2.10 MySQL

MySQL merupakan sebuah aplikasi pengolahan *database* yang sering digunakan oleh *programmer*, disamping sudah mendukung dalam pembuaan *database* yang berbasis *Client/Server*, juga dapat mengolah *database* dalam jumlah besar. SQL adalah bahasa standar untuk *query* yang digunakan untuk manipulasi data yang dibuat dalam berbagai DBMS (*Database Management System*), salah satunya adalah MySQL. Beberapa keistimewaan yang dimiliki oleh MySQL adalah *Portability, Open Source, Multiuser, Performant Tuning, Column Tuning, Command dan Function, Security, Scalability dan Limits, Connectivity, Localization, Interface, Cliens dan Tools, Structure Table* (S. Mutiah, 2019)