

LAPORAN AKHIR

ANALISIS KEPUASAN KONSUMEN PADA MAHASISWA FTMM TERHADAP LAYANAN E-COMMERCE SHOPEE



KELOMPOK: SD-A1 - 6 :

1. SRI RANTA	162112133003
2. SUKMA SEKAR DEVITA	162112133015
3. THERESA AGNES VIRNAULI SINAGA	162112133058
4. DELLA LUKITA WARDHANI	162112133075
5. FARAH AISYAH MUNIR	162112133077

**TEAM-BASED PROJECT
MATA KULIAH DATA MINING I
PROGRAM STUDI TEKNOLOGI SAINS DATA
FAKULTAS TEKNOLOGI MAJU DAN MULTIDISIPLIN
UNIVERSITAS AIRLANGGA
2023**

DAFTAR ISI

DAFTAR ISI.....	ii
DAFTAR GAMBAR.....	iv
DAFTAR TABEL	v
BAB 1 PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Alasan Penelitian	1
1.3 Tujuan Penelitian	2
1.4 Manfaat Penelitian	2
BAB 2 TINJAUAN PUSTAKA	3
2.1 Shopee Sebagai E-Commerce Terkemuka di Indonesia.....	3
2.2 Konsep Kepuasan Konsumen	3
2.2.1 Perilaku Konsumen Mahasiswa	3
2.2.2 Faktor Pengaruh Kepuasan Konsumen Shopee	3
2.3 Penerapan Metode Klasifikasi terhadap Dataset Kepuasan Customer E-Commerce.....	4
2.3.1 Random Forest	4
2.3.2 Decision Tree	4
2.3.3 Support Vector Machine (SVM)	5
2.3.4 k-Nearest Neighbors (KNN)	5
2.3.5 Naïve Bayes.....	6
BAB 3 METODOLOGI PENELITIAN.....	7
3.1 Pengumpulan Data	7
3.1.1 Teknik Pengambilan Sampel.....	7
3.1.2 Penjelasan Variabel	7
3.2 <i>Insight Exploratory Data Analysis</i> (EDA).....	8
3.2.1 Data Shape.....	8
3.2.2 Data Information	9
3.2.3 Data Description.....	9
3.2.4 Unique Value Variabel X Categorical	9
3.2.5 Data Distribution Variabel X Numeric	11
3.2.6 Unique Value Variabel Y	13
3.3 Data Preprocessing.....	13
3.4 Implementasi Machine Learning Klasifikasi	13

3.4.1 Data Partition.....	13
3.4.2 Alur Machine Learning	14
3.4.3 Model Validation Methods.....	14
BAB 4 HASIL DAN PEMBAHASAN.....	15
4.1 Pre-Processing.....	15
4.1.1 Missing Value	15
4.1.2 Handling Outlier.....	15
4.1.3 Encoding Variabel.....	16
4.2 Hasil Confusion Matrix.....	16
4.2.1 Random Forest	16
4.2.2 Decision Tree	17
4.2.3 Support Vector Machine	17
4.2.4 k-Nearest Neighbors.....	17
4.2.5 Naïve Bayes.....	18
4.3 Matriks Evaluasi	18
4.3.1 Random Forest	18
4.3.2 Decision Tree	18
4.3.3 SVM	19
4.3.4 kNN	19
4.3.5 Naive Bayes.....	19
BAB 5 KESIMPULAN DAN SARAN.....	20
5.1 Kesimpulan.....	20
5.2 Saran.....	21
DAFTAR PUSTAKA.....	22
LAMPIRAN.....	23

DAFTAR GAMBAR

Gambar 1. Informasi Data	9
Gambar 2. Deskripsi Dataset.....	9
Gambar 3. Bar Plot Variabel Gender.....	10
Gambar 4. Bar Plot Variabel Ketersediaan Produk.....	10
Gambar 5. Bar Plot Variabel Kualitas Produk	10
Gambar 6. Bar Plot Variabel Harga.....	10
Gambar 7. Bar Plot Variabel Promosi	11
Gambar 8. Bar Plot Variabel Ketepatan Pengiriman.....	11
Gambar 9. Bar Plot Variabel Penanganan Keluhan	11
Gambar 10. Plot Variabel Usia.....	12
Gambar 11. Plot Variabel Ketersediaan Produk.....	12
Gambar 12. Plot Variabel Frekuensi Transaksi.....	12
Gambar 13. Plot Variabel Skor Kepuasan.....	12
Gambar 14. Bar Chart Variabel Target	13
Gambar 15. Flowchart Pre-Processing.....	13
Gambar 16. Dimensi Data Train dan Data Test	13
Gambar 17. Flowchart Machine Learning.....	14
Gambar 18. Hasil Pemeriksaan Missing Value	15
Gambar 19. Hasil Pemeriksaan Handling Outlier.....	15
Gambar 20. Hasil Encoding Variabel pada Variabel Kategorik.....	16
Gambar 21. Confusion Matrix Random Forest	16
Gambar 22. Confusion Matrix Decision Tree	17
Gambar 23. Confusion Matrix SVM	17
Gambar 24. Confusion Matrix KNN	17
Gambar 25. Confusion Matrix Naïve Bayes	18
Gambar 26. Classification Report Random Forest.....	18
Gambar 27. Classification Report Decision Tree	18
Gambar 28. Classification Report SVM.....	19
Gambar 29. Classification Report KNN.....	19
Gambar 30. Classification Report Naïve Bayes	19

DAFTAR TABEL

Table 1. Penjelasan Variabel Pada Datasets Kepuasan Customer Shopee.....	8
--------------------------------------------------------------------------	---

BAB 1

PENDAHULUAN

1.1 Latar Belakang

Dalam era digital ini, *e-commerce* atau perdagangan elektronik telah menjadi salah satu cara utama bagi konsumen untuk melakukan pembelian barang dan jasa secara online. Hal ini juga berlaku di kalangan mahasiswa yang sering menggunakan platform *e-commerce*. *E-commerce/electronic commerce*, dalam bahasa Indonesia dikenal dengan istilah Perdagangan elektronik merupakan sebuah aktivitas pembelian, penjualan, pemasaran barang maupun jasa melalui sistem elektronik, diantaranya dengan jaringan internet/world wide web atau jaringan komputer lainnya. Penggunaan *e-commerce* pada kalangan mahasiswa sering kali digunakan untuk membeli kebutuhan pribadi, termasuk buku kuliah, alat tulis, perangkat elektronik, pakaian, dan banyak lagi.

Dengan adanya *e-commerce* pada era digitalisasi ini memberikan banyak manfaat bagi mahasiswa seperti efisiensi biaya dan waktu, produk yang beragam, pembelian barang import, dan kemudahan akses. Dalam sebuah penelitian, didapatkan bahwa usia pengguna *e-commerce* terbanyak yaitu Generasi Z pada rentang usia 18-24 tahun. Berdasarkan hal tersebut kepuasan mahasiswa terhadap pengalaman berbelanja online menjadi hal penting untuk dipahami oleh penyedia layanan *e-commerce*. Memahami faktor-faktor yang mempengaruhi kepuasan mahasiswa dapat membantu perusahaan dalam meningkatkan kualitas layanan mereka dan memberikan pengalaman yang lebih baik kepada Konsumen.

1.2 Alasan Penelitian

Penelitian dengan judul “Analisis Kepuasan Konsumen Pada Mahasiswa FTMM Terhadap Layanan E-Commerce Shopee” memiliki beberapa alasan penting, antara lain:

1. Peningkatan pengalaman pengguna: Melalui analisis kepuasan mahasiswa, perusahaan *e-commerce* dapat mengidentifikasi aspek-aspek yang dapat ditingkatkan dalam pengalaman pengguna.
2. Meningkatkan loyalitas Konsumen: Kepuasan mahasiswa secara langsung berkaitan dengan loyalitas Konsumen. Jika mahasiswa merasa puas dengan pengalaman berbelanja online, mereka cenderung menjadi Konsumen yang loyal dan akan kembali menggunakan platform *e-commerce* yang sama untuk kebutuhan belanja mereka di masa depan.
3. Meningkatkan reputasi perusahaan: Kepuasan mahasiswa dapat berdampak pada reputasi perusahaan *e-commerce*. Jika mahasiswa merasa puas dengan layanan dan produk yang diberikan, mereka cenderung memberikan ulasan positif, merekomendasikan kepada teman dan keluarga, atau berbagi pengalaman positif mereka di media sosial.
4. Pengambilan keputusan berdasarkan data: Analisis kepuasan mahasiswa pada *e-commerce* memberikan data yang objektif dan dapat diandalkan untuk pengambilan keputusan perusahaan.

Dengan memahami pentingnya analisis kepuasan mahasiswa pada *e-commerce*, perusahaan dapat memprioritaskan pengalaman pengguna yang lebih baik, meningkatkan

loyalitas Konsumen, membangun reputasi yang positif, dan mengoptimalkan operasional bisnis mereka. Analisis ini membantu perusahaan untuk tetap kompetitif dalam industri e-commerce yang semakin pesat dan memenuhi harapan mahasiswa sebagai konsumen yang kritis dan teknologi-savvy.

1.3 Tujuan Penelitian

- a. Tujuan umum
Menganalisis tingkat kepuasan mahasiswa sebagai konsumen dalam menggunakan media dan layanan yang terdapat pada *e-commerce*.
- b. Tujuan khusus
 1. Mengidentifikasi faktor-faktor yang mempengaruhi kepuasan konsumen pada mahasiswa dalam menggunakan layanan *e-commerce*.
 2. Mengevaluasi tingkat kepuasan konsumen mahasiswa terhadap layanan *e-commerce*.
 3. Mengeksplorasi harapan dan kebutuhan khusus mahasiswa terkait layanan *e-commerce*.
 4. Memberikan rekomendasi atau saran perbaikan kepada penyedia layanan *e-commerce* untuk meningkatkan kepuasan konsumen mahasiswa.

1.4 Manfaat Penelitian

- a. Manfaat teoritis
Memperluas pengetahuan kita tentang kepuasan konsumen dalam layanan *e-commerce*, mengembangkan model konseptual, memvalidasi teori-teori yang ada, memahami faktor-faktor pengaruh, dan menjembatani kesenjangan teoritis dalam konteks khusus mahasiswa dalam layanan *e-commerce*.
- b. Manfaat praktis
 1. Memahami faktor-faktor yang mempengaruhi kepuasan konsumen pada mahasiswa dalam menggunakan layanan *e-commerce*
 2. Dapat mengidentifikasi area yang perlu ditingkatkan dalam layanan *e-commerce*
 3. Memberikan wawasan yang berharga bagi perusahaan penyedia layanan *e-commerce* dalam pengambilan keputusan strategis terkait pengembangan produk

BAB 2

TINJAUAN PUSTAKA

2.1 Shopee Sebagai E-Commerce Terkemuka di Indonesia

Shopee telah berdiri di Indonesia sejak tahun 2015. Sejak saat itu, Shopee telah menjadi pilihan utama banyak konsumen dalam melakukan pembelian secara online. Popularitas *platform* ini terus meningkat dari tahun ke tahun. Terutama, dengan munculnya pandemi Covid-19 di Indonesia pada kuartal kedua tahun 2020, belanja *online* semakin populer di kalangan masyarakat, terutama bagi mereka yang memiliki akses ke *smartphone* dan internet.

Menurut Suhendra & Krisnadi (2020), dalam penelitian mereka tentang kekuatan Shopee sebagai *e-commerce* terpopuler di Indonesia saat ini, mereka menemukan bahwa Shopee telah mengimplementasikan berbagai inovasi dan strategi pemasaran untuk memperluas pangsa pasarnya. Mereka secara aktif menggunakan iklan dengan Brand Ambassador besar, menyelenggarakan flash sale dan event khusus, serta menerapkan fitur gamifikasi pada aplikasi seluler mereka. Saat ini, *ShopeePay*, yang merupakan layanan pembayaran digital yang terkait dengan Shopee, menjadi semakin populer dan termasuk salah satu yang terdepan dalam industri ini.

2.2 Konsep Kepuasan Konsumen

2.2.1 Perilaku Konsumen Mahasiswa

Pada penelitian sebelumnya, ditemukan bahwa terdapat modifikasi perilaku konsumen pada mahasiswa dalam membeli barang (Utari, Fitri, Setyanto dan Henny, 2020). Utamanya, sejak pandemi Covid-19, pola pembelian barang di kalangan mahasiswa sudah tidak tergantung oleh kredibilitas produsen seperti pada era normal. Ketersediaan barang dan ketepatan waktu juga menjadi hal yang paling penting dalam pembelian *online*. Satu hal yang berdampak positif dari adanya keterbatasan ruang gerak secara fisik saat pandemi Covid-19, keadaan ini memunculkan kebiasaan berbelanja *online* khususnya bagi mahasiswa melalui *platform e-commerce* yang bersifat global.

2.2.2 Faktor Pengaruh Kepuasan Konsumen Shopee

Penelitian sebelumnya terkait faktor pengaruh kepuasan konsumen shopee, menemukan bahwa variabel harga dan kualitas layanan elektronik berpengaruh secara signifikan terhadap kepuasan konsumen pada pengguna *platform e-commerce* Shopee (Putri dan Iriani, 2019). Produk atau jasa yang memiliki kualitas yang baik akan memiliki spesifikasi yang sesuai dengan kebutuhan konsumen, sehingga dapat menciptakan kepuasan bagi mereka. Ini akan mendorong konsumen untuk merasa puas dengan kualitas layanan elektronik yang mereka terima. Sebagai contoh, Shopee memiliki kebijakan yang aktif dalam menyimpan dan melindungi informasi pribadi pengguna, sebagai wujud nyata dari komitmen privasi Shopee. Selain itu, Shopee juga menghargai masukan dan saran dari konsumen untuk terus meningkatkan kualitas layanan mereka ke depan.

2.3 Penerapan Metode Klasifikasi terhadap Dataset Kepuasan Customer E-Commerce

2.3.1 Random Forest

Random forest merupakan metode klasifikasi yang dilakukan dengan mengembangkan metode Decision Tree berdasarkan pemilihan atribut acak pada setiap node untuk menentukan klasifikasi. Pada proses klasifikasinya didasarkan pada suara terbanyak dari pohon keputusan yang dikembalikan. Random forest menjadi salah satu algoritma yang populer dalam machine learning karena kemampuannya dalam mengatasi masalah overfitting dan memiliki kemampuan untuk mengatasi data yang tidak seimbang (Hamidi dan Daraei, 2016).

Proses klasifikasi pada random forest dimulai dengan mengacak pembagian data sampel ke dalam decision tree. Setelah pohon-pohon terbentuk, voting dilakukan untuk setiap kelas pada data sampel. Kemudian, hasil voting dari setiap kelas digabungkan dan dipilih vote yang memiliki jumlah terbanyak. Dengan menggunakan random forest dalam klasifikasi data, akan diperoleh hasil voting yang optimal.

2.3.2 Decision Tree

Metode pengklasifikasian decision tree adalah salah satu teknik yang terkenal dalam bidang data mining dan menjadi metode populer dalam pengambilan keputusan. Keunggulan dari metode ini adalah tidak memerlukan proses pengelolaan pengetahuan sebelumnya dan dapat digunakan untuk menyelesaikan kasus-kasus dengan dimensi yang besar (Widodo, Handayanto dan Herlawati, 2013).

Atribut Decision Tree adalah atribut yang memiliki kemampuan untuk menghasilkan decision tree dengan ukuran paling kecil atau atribut yang dapat memisahkan objek berdasarkan kelasnya. Atribut yang dipilih adalah atribut yang memiliki tingkat kemurnian ("purest"). Suatu cabang dikatakan "pure" jika semua anggotanya berasal dari satu kelas. Semakin murni cabang tersebut, semakin baik. Tingkat kemurnian dinyatakan dengan ukuran impurity. Salah satu kriteria untuk mengukur impurity adalah information gain (Prasetyo, 2014).

$$entropy(S) = \sum_{i=1}^m -p(w_i|S) \cdot \log_2 p(w_i|S)$$

Source: media.neliti.com

Information gain yang berasal dari variabel dependen S dan dikelompokkan berdasarkan atribut J, sehingga dapat dinotasikan sebagai gain(S,J) dengan rumus sebagai berikut:

$$Gain(S,J) = Entropy(S) - \sum_{i=j}^n p(v_i|S) * E(S_i)$$

Source: media.neliti.com

Split information diperlukan untuk menghitung *gain ratio*. Formula *split information* sebagai berikut:

$$SplitInformation(S, J) = - \sum_{i=1}^n p(v_i|S) \log_2 p(v_i|S)$$

Source: media.neliti.com

Dengan rumus gais ratio:

$$Gainratio(S, J) = \frac{Gain(S, J)}{Split Information(S, J)}$$

2.3.3 Support Vector Machine (SVM)

Support Vector Machine (SVM) adalah salah satu teknik dalam pembelajaran mesin (Machine Learning), yang mengikuti teknik pembelajaran mesin sebelumnya yang dikenal sebagai Neural Networks (NN). Baik SVM maupun NN telah sukses digunakan dalam pengenalan pola. Proses pembelajaran dilakukan dengan menggunakan pasangan data input dan output yang diharapkan. Pendekatan pembelajaran ini dikenal sebagai *Supportvised learning* (Kerami dan Murfi, 2004).

Pada konsep SVM, awalnya terdapat masalah klasifikasi dua kelas yang memerlukan set pelatihan yang terdiri dari sampel positif dan negatif. SVM akan mencari hyperplane (pemisah) yang optimal untuk memisahkan kedua kelas tersebut dan memaksimalkan margin antara kelas-kelas tersebut. Terdapat 2 metode SVM yaitu:

a. Metode Linier

Model linier memiliki kepentingan yang signifikan dari segi komputasi dan analitis. Namun, penggunaan model linier dengan pendekatan parametrik dalam metode klasik terbatas dalam aplikasi praktis karena menghadapi kendala kutukan dimensionalitas. Dua pendekatan yang digunakan dalam metode linier ini yaitu pendekatan alternatif dan pendekatan nonparametrik.

b. Metode Kernel

Fungsi kernel merupakan fungsi k untuk semua vektor masukan x . $\phi(x) T \phi(z)$ dimana $\phi(.)$ merupakan fungsi pemetaan dari ruang masukan ke ruang fitur. Fungsi kernel digunakan untuk menerapkan model dalam dimensi yang lebih tinggi tanpa perlu menentukan fungsi pemetaan dari ruang masukan ke ruang fitur. Salah satu contoh fungsi kernel yang populer adalah Gaussian radial basis function (RBF). Fungsi ini dinyatakan sebagai $k(x, x') = \phi(\|x - x'\|) = \exp(-\|x - x'\|^2 / 2s^2)$, di mana x' merupakan titik "inti" yang dipilih dari data latihan.

2.3.4 k-Nearest Neighbors (KNN)

K-Nearest Neighbors (KNN) merupakan sebuah algoritma dalam machine learning yang digunakan untuk mengklasifikasikan dan memprediksi nilai dari data. Algoritma ini menggunakan prinsip apabila terdapat data yang memiliki atribut yang mirip, maka kemungkinan besar mereka akan memiliki label kelas atau nilai target yang sama. Dalam KNN, ketika terdapat data baru yang perlu diprediksi, maka cari data yang paling dekat secara jarak dengan data baru tersebut. Kita memilih sejumlah data terdekat yang disebut sebagai tetangga, dan kemudian kita memutuskan label kelas atau nilai target berdasarkan mayoritas suara dari tetangga tersebut. Sangat penting untuk memilih jumlah tetangga yang tepat, karena jika terlalu sedikit tetangga yang dipilih,

kita bisa terpengaruh oleh data yang tidak relevan, sedangkan jika terlalu banyak tetangga, keputusan bisa menjadi tidak jelas.

2.3.5 Naïve Bayes

Algoritma Naive Bayes adalah metode klasifikasi yang menggunakan Teorema Bayes dengan asumsi sederhana bahwa setiap fitur dalam dataset tidak saling mempengaruhi secara bersyarat terhadap label kelas. Dengan kata lain, algoritma ini menganggap setiap fitur memberikan kontribusi terpisah terhadap probabilitas dari suatu kelas. Naive Bayes menggunakan pendekatan probabilitas untuk mengklasifikasikan data baru berdasarkan kemungkinan munculnya atribut-atributnya di setiap kelas yang ada. Meskipun asumsi ini seringkali tidak sesuai dengan dunia nyata, algoritma Naive Bayes tetap efektif dalam banyak kasus dan sering digunakan dalam klasifikasi teks, deteksi spam, analisis sentimen, dan berbagai aplikasi lainnya.

BAB 3

METODOLOGI PENELITIAN

Penelitian terkait analisis kepuasan konsumen terhadap *e-commerce* shopee dengan metode klasifikasi dilakukan untuk segmentasi kepuasan konsumen terhadap *e-commerce* berdasarkan atribut prediktornya. Setelah konsumen berhasil disegmentasi, penulis dapat mengidentifikasi faktor penyebab kepuasan konsumen, sehingga bisa ditarik suatu kesimpulan dan menjadi bahan evaluasi untuk peningkatan kepuasan konsumen kedepannya. Alur penelitian yang kami lakukan adalah seperti di bawah ini.

3.1 Pengumpulan Data

Pada subbab ini penulis akan membahas terkait bagaimana data didapatkan beserta data apa saja yang dibutuhkan untuk memenuhi tujuan. Berikut adalah penjabarannya yang ditilik dari teknik pengambilan data dan penjelasan variabel.

3.1.1 Teknik Pengambilan Sampel

Untuk memberikan hasil klasifikasi yang kredibel dan tujuan dapat tercapai, penulis menggunakan data primer dimana sampel diambil langsung oleh survey menggunakan *google form*. Survey dilakukan selama kurang lebih empat hari dengan populasi target adalah mahasiswa Universitas Airlangga Fakultas Teknologi Maju dan Multidisiplin yang menggunakan aplikasi *e-commerce* secara random (*random sampling*). Adapun pertanyaan yang diajukan melalui survey merupakan variabel-variabel prediktor yang menyusun keputusan akhir di variabel respon. Variabel-variabel prediktor tersebut antara lain adalah ketersediaan produk (selalu tersedia atau sering kosong), kualitas produk (tahan lama atau mudah rusak), harga (cenderung murah atau cenderung mahal), promosi (menarik atau tidak menarik), ketepatan pengiriman (tepat waktu atau sering terlambat), penanganan keluhan (baik dan respon cepat atau buruk dan respon lambat), frekuensi akses (dalam menit), dan terakhir adalah frekuensi transaksi (per bulan). Lalu untuk variabel responnya adalah kepuasan konsumen (puas atau tidak) dengan *range* level kepuasan adalah 1- 4 (tidak puas sampai sangat puas). Dari survey yang penulis lakukan, didapatkan 316 responden dengan total responden yang puas terhadap *e-commerce* shopee sebanyak 275 data dan yang tidak puas terhadap *e-commerce* shopee sebanyak 41 data.

3.1.2 Penjelasan Variabel

Variabel Prediktor yang penulis gunakan pastinya sangat disesuaikan dengan tujuan, sehingga digunakan variabel- variabel yang berpengaruh terhadap kepuasan customer. Berikut adalah penjelasannya.

Nama Variabel	Tipe Data	Keterangan
Jenis Kelamin	Numerik	Usia responden.
Ketersediaan Produk	Kategorik	Jumlah produk tersedia dan bisa langsung proses transaksi tanpa perlu menunggu ketersediaan barang.
Kualitas Produk	Kategorik	Produk memiliki kualitas yang setara dengan harga.
Harga	Kategorik	Harga mampu disesuaikan dengan budget responden.
Promosi	Kategorik	Promosi baik terkait produk, harga, maupun pelayanan jasa diberitakan sebaik mungkin.
Ketepatan Pengiriman	Kategorik	Produk dikirim sesuai dengan waktu estimasi produk sampai.
Penanganan Keluhan	Kategorik	Keluhan ditanggapi dengan baik oleh customer services, baik keluhan terkait produk maupun pelayanan toko.
Frekuensi Akses	Numerik	Seberapa sering responden mengakses aplikasi.
Frekuensi Transaksi	Numerik	Seberapa sering responden melakukan transaksi (pembelian produk) di e-commerce.

Table 1. Penjelasan Variabel Pada Datasets Kepuasan Customer Shopee

3.2 *Insight Exploratory Data Analysis (EDA)*

Setelah data berhasil dikumpulkan, penulis mengambil *insight* terhadap data dengan menggunakan Exploratory Data Analysis (EDA). Kegunaan dari proses pembuatan EDA dalam penelitian ini adalah untuk lebih memahami data terutama terkait distribusi data, deteksi outlier, dan tentunya adalah untuk memvisualisasikan data sebelum data diolah dalam tahap *preprocessing* dan *modelling*.

3.2.1 Data Shape

Data shape digunakan untuk memahami ukuran atau dimensi pada data, sehingga peneliti dapat memilih perlakuan apa yang tepat pada proses *preprocessing*, misalnya adalah implementasi reduksi dimensi data dengan *feature selection*. Informasi terkait data shape diwakili dari dua angka, angka pertama (sebelum koma) merujuk pada jumlah baris dan angka kedua (setelah koma) merujuk pada jumlah kolom. Pada dataset yang dilakukan dalam penelitian, terdapat 316 baris dan 14 kolom.

3.2.2 Data Information

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 316 entries, 0 to 315
Data columns (total 14 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   gender                                316 non-null    object
1   prodi                                 316 non-null    object
2   angkatan                             316 non-null    int64
3   usia                                 316 non-null    int64
4   ketersediaan_produk                  316 non-null    object
5   kualitas_produk                      316 non-null    object
6   harga                                316 non-null    object
7   promosi                              316 non-null    object
8   ketepatan_pengiriman                 316 non-null    object
9   penanganan_keluhan                  316 non-null    object
10  frekuensi_akses_shopee               316 non-null    int64
11  frekuensi_transaksi                  316 non-null    int64
12  kepuasan_konsumen                    316 non-null    object
13  skor_kepuasan                        316 non-null    int64
dtypes: int64(5), object(9)
memory usage: 34.7+ KB
```

Gambar 1. Informasi Data

Data information digunakan untuk melihat informasi terkait data yang ditilik dari tiap- tiap variabelnya. Adapun informasi yang diambil antara lain adalah nama variabel, jumlah kolom tiap variabel (yang non-null maupun null), dan tipe data tiap variabel. Dari data yang penulis gunakan, terdapat 13 variabel yang tiap variabel memiliki 316 data non- null, serta tipe data variabel beragam dari yang kategorik sampai dengan numerik.

3.2.3 Data Description

	gender	prodi	angkatan	usia	ketersediaan_produk	kualitas_produk	harga	promosi	ketepatan_pengiriman	penanganan_keluhan	frekuensi_akses_shopee	frekuensi_transaksi	kepuasan_konsumen	skor_kepuasan
count	316	316	316.000000	316.000000	316	316	316	316	316	316	316.000000	316.000000	316	316.000000
unique	2	5	NaN	NaN	2	2	2	2	2	2	NaN	NaN	2	NaN
top	P	Teknologi Sains Data	NaN	NaN	Selalu Tersedia	Tahan Lama	Cenderung Murah	Menarik	Tepat Waktu	Baik dan Respon Cepat	NaN	NaN	Puas	NaN
freq	197	84	NaN	NaN	258	255	244	244	265	263	NaN	NaN	275	NaN
mean	NaN	NaN	2020.968354	19.746835	NaN	NaN	NaN	NaN	NaN	NaN	51.727848	6.620253	NaN	3.161382
std	NaN	NaN	0.595442	0.998009	NaN	NaN	NaN	NaN	NaN	NaN	46.054853	6.632395	NaN	0.810191
min	NaN	NaN	2020.000000	17.000000	NaN	NaN	NaN	NaN	NaN	NaN	0.000000	0.000000	NaN	1.000000
25%	NaN	NaN	2021.000000	19.000000	NaN	NaN	NaN	NaN	NaN	NaN	20.000000	2.000000	NaN	3.000000
50%	NaN	NaN	2021.000000	20.000000	NaN	NaN	NaN	NaN	NaN	NaN	30.000000	5.000000	NaN	3.000000
75%	NaN	NaN	2021.000000	20.000000	NaN	NaN	NaN	NaN	NaN	NaN	60.000000	10.000000	NaN	4.000000
max	NaN	NaN	2022.000000	24.000000	NaN	NaN	NaN	NaN	NaN	NaN	360.000000	60.000000	NaN	4.000000

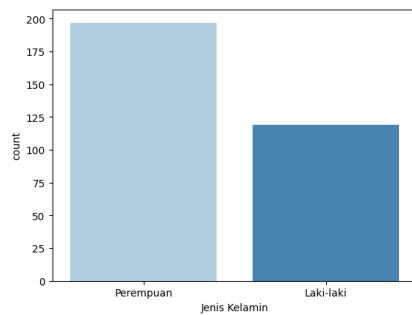
Gambar 2. Deskripsi Dataset

Data description digunakan untuk mengetahui informasi terkait nilai statistika deskriptif pada data, yang meliputi jumlah data (count), nilai unique, rata- rata, standar deviasi (std), nilai minimum (min), kuartil (25%, 50%, 75%), dan nilai maksimum (max). Dengan mengetahui informasi nilai statistika deskriptif, dapat dipilih langkah analisis yang baik dalam pemrosesan data dan pengembangan model.

3.2.4 Unique Value Variabel X Categorical

Untuk melihat bagaimana persebaran data pada variabel X_kategorik, penulis menggunakan barplot dengan output berupa dua bar yang mewakili jumlah dari kedua atribut beserta perbandingannya.

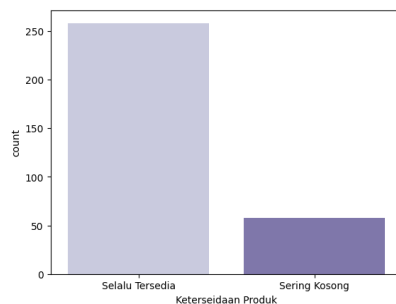
1) Gender



Gambar 3. Bar Plot Variabel Gender

Dari visualisasi bar plot, dapat disimpulkan bahwa jumlah responden perempuan lebih banyak dibandingkan jumlah responden laki-laki. Terdapat total 197 responden perempuan, dan 119 responden laki-laki.

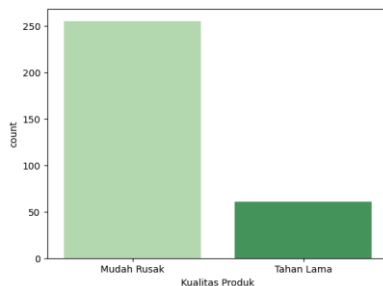
2) Ketersediaan Produk



Gambar 4. Bar Plot Variabel Ketersediaan Produk

Dari visualisasi bar plot, dapat disimpulkan bahwa jumlah responden yang memilih ketersediaan produk *e-commerce* Shopee selalu tersedia lebih banyak daripada yang memilih sering kosong. Terdapat total 258 responden yang memilih selalu tersedia, dan 58 responden yang memilih sering kosong.

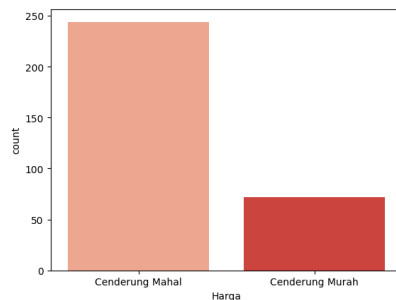
3) Kualitas Produk



Gambar 5. Bar Plot Variabel Kualitas Produk

Dari visualisasi bar plot, dapat disimpulkan bahwa jumlah responden yang memilih kualitas produk *e-commerce* Shopee mudah rusak lebih banyak daripada yang memilih tahan lama. Terdapat total 255 responden yang memilih produk mudah rusak, dan 61 responden yang memilih produk tahan lama.

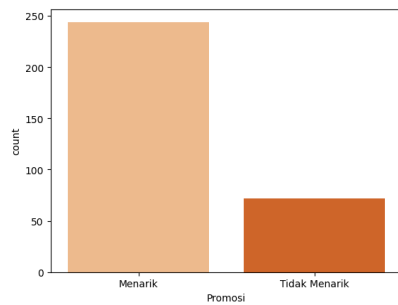
4) Harga



Gambar 6. Bar Plot Variabel Harga

Dari visualisasi bar plot, dapat disimpulkan bahwa jumlah responden yang memilih harga produk *e-commerce* Shopee cenderung mahal lebih banyak daripada yang memilih cenderung murah. Terdapat total 244 responden yang memilih produk cenderung mahal, dan 72 responden yang memilih cenderung murah.

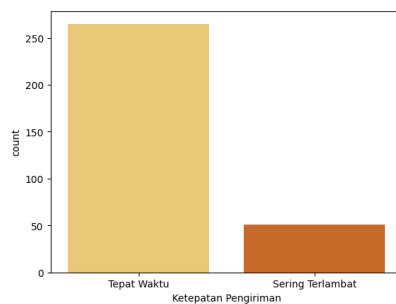
5) Promosi



Gambar 7. Bar Plot Variabel Promosi

Dari visualisasi bar plot, dapat disimpulkan bahwa jumlah responden yang memilih promosi *e-commerce* Shopee menarik lebih banyak daripada yang memilih promosi tidak menarik. Terdapat total 244 responden yang memilih produk cenderung mahal, dan 72 responden yang memilih produk cenderung murah.

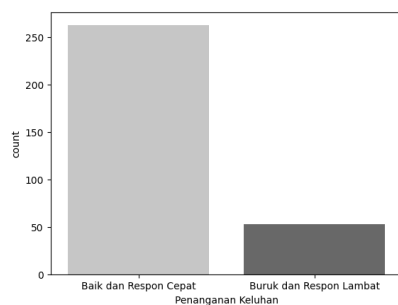
6) Ketepatan Pengiriman



Gambar 8. Bar Plot Variabel Ketepatan Pengiriman

Dari visualisasi bar plot, dapat disimpulkan bahwa jumlah responden yang memilih ketepatan pengiriman *e-commerce* Shopee tepat waktu lebih banyak daripada yang sering terlambat. Terdapat total 265 responden yang memilih produk cenderung mahal, dan 51 responden yang memilih produk cenderung murah.

7) Penanganan Keluhan



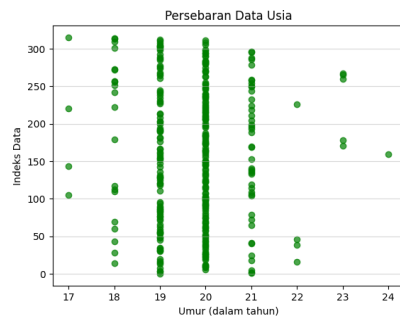
Gambar 9. Bar Plot Variabel Penanganan Keluhan

Dari visualisasi bar plot, dapat disimpulkan bahwa jumlah responden yang memilih penanganan keluhan *e-commerce* Shopee baik dan respon cepat lebih banyak daripada yang memilih buruk dan respon lambat. Terdapat total 263 responden yang memilih produk cenderung mahal, dan 53 responden yang memilih produk cenderung murah.

3.2.5 Data Distribution Variabel X Numeric

Untuk melihat bagaimana persebaran data pada variabel *X_numeric*, penulis menggunakan scatterplot dengan output berupa grafik yang menunjukkan titik-titik data pada koordinat kartesian, dimana sumbu kedua sumbu (x dan y) pada grafik merupakan perbandingan antara dua nilai yang ingin dibandingkan.

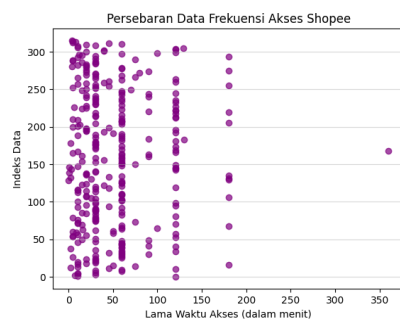
1) Usia (satuan tahun)



Gambar 10. Plot Variabel Usia

Dari visualisasi scatter plot persebaran data usia, dapat disimpulkan bahwa rata-rata usia responden yang mengisi form berada dalam rentang 19-21 tahun.

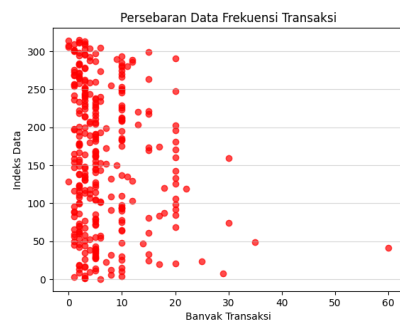
2) Ketersediaan Produk (satuan menit)



Gambar 11. Plot Variabel Ketersediaan Produk

Dari visualisasi scatter plot persebaran data frekuensi akses shopee, dapat disimpulkan bahwa rata-rata responden mengakses shopee berada pada rentang 0- 50 menit dan terdapat persebaran lain pada rentang 51- 200 menit. Lalu, terdeteksi outlier pada rentang nilai lebih dari 350 menit.

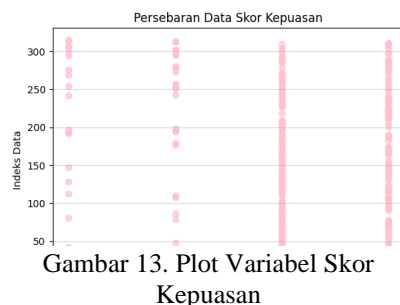
3) Frekuensi Transaksi (satuan bulan)



Gambar 12. Plot Variabel Frekuensi Transaksi

Dari visualisasi scatter plot persebaran data frekuensi transaksi, dapat disimpulkan bahwa rata-rata responden melakukan transaksi shopee berada pada rentang 0- 5 kali per bulan dan terdapat persebaran lain pada rentang 10- 35 kali perbulan. Lalu, terdeteksi outlier pada rentang nilai lebih dari 60 kali perbulan.

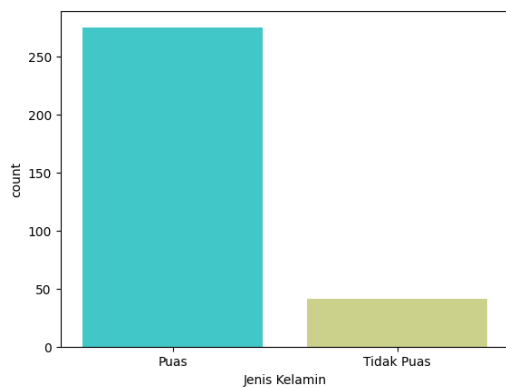
4) Skor Kepuasan



Gambar 13. Plot Variabel Skor Kepuasan

Dari visualisasi scatter plot persebaran data usia, dapat disimpulkan bahwa rata-rata usia responden memiliki skor kepuasan 3- 4 (puas- sangat puas)

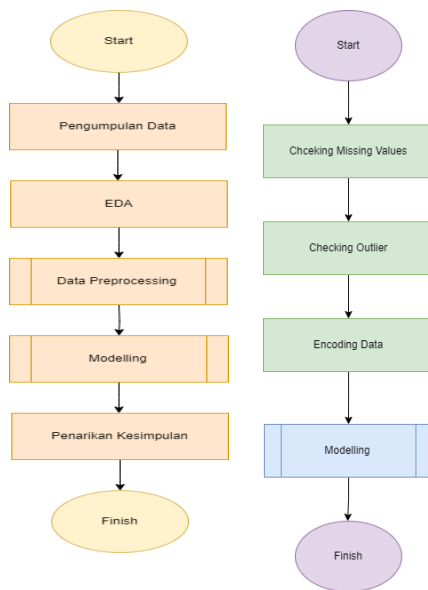
3.2.6 Unique Value Variabel Y



Gambar 14. Bar Chart Variabel Target

Untuk melihat bagaimana persebaran data pada variabel Y, penulis menggunakan barplot dengan output berupa dua bar yang mewakili jumlah dari kedua atribut beserta perbandingannya (puas dan tidak puas). Dari visualisasi bar plot, dapat disimpulkan bahwa jumlah responden yang puas terhadap e-commerce shopee lebih banyak dibandingkan jumlah responden tidak puas. Terdapat sekitar 255 responden yang puas, sedangkan jumlah responden tidak puas berkisar antara 49.

3.3 Data Preprocessing



Gambar 15. Flowchart Pre-Processing

Setelah mengetahui bagaimana bentuk, distribusi, dan visualisasi dari hasil Exploratory Data Analysis (EDA) tahap selanjutnya adalah data *preprocessing*. Tujuan dari tahap ini adalah supaya data benar- benar layak dan kredibel untuk klasifikasi, sehingga analisis dan pengambilan keputusan terhadap model dapat diimplementasikan secara optimal. Metode yang digunakan dalam *preprocessing* merupakan hal-hal dengan urgensi tinggi terhadap kredibilitas data dan disamakan dengan kebutuhan terhadap data, yaitu pemeriksaan terhadap *missing values*, penanganan terhadap *outlier*, dan *encoding* data.

3.4 Implementasi Machine Learning Klasifikasi

Dalam menentukan model klasifikasi yang menciptakan kelas- kelas dengan akurasi yang tinggi, penulis membandingkan antara lima metode klasifikasi dan mengambil dua metode dengan akurasi tertinggi.

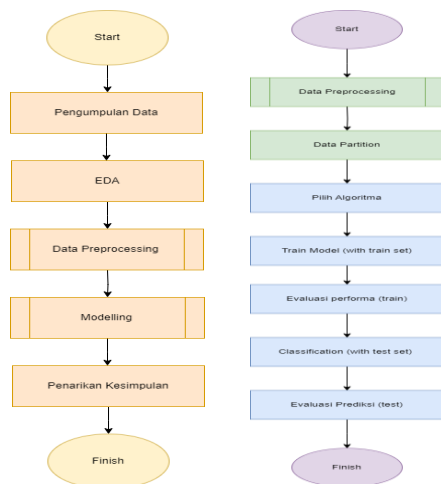
3.4.1 Data Partition

```
Dimensi X_train : (252, 11)
Dimensi X_test  : (64, 11)
Dimensi y_train : (252,)
Dimensi y_test  : (64,)
```

Gambar 16. Dimensi Data Train dan Data Test

Sebelum melakukan pemodelan, penulis melakukan partisi terhadap data terlebih dahulu yaitu memisahkan data dengan menggunakan *syntax* train test split, yaitu membagi data yang akan digunakan untuk Training (pelatihan) dan Test (pengujian) dengan proporsi 20% data test dan 80% data train. Training set digunakan untuk mempelajari pola atau hubungan antar variabel respon (X) dengan outputnya (y), sehingga model dapat mempelajari dan melakukan prediksi pada data baru. Sementara itu, testing set digunakan untuk mengukur performa model yang ditilik dari akurasi, presisi, recall, dan matriks evaluasi lainnya.

3.4.2 Alur Machine Learning



Gambar 17. Flowchart Machine Learning

Alur machine learning merujuk pada langkah secara urut dan rinci terkait proses dalam menciptakan solusi machine learning, dalam penelitian ini merujuk pada klasifikasi. Alur secara umum setelah partisi data adalah pemilihan model, pelatihan model, evaluasi model, dan penyesuaian model. Model yang penulis gunakan pada penelitian ini antara lain random forest, decision tree, SVM, kNN, dan naive bayes.

3.4.3 Model Validation Methods

Jenis metode validasi model yang penulis gunakan adalah *cross-validation*. *Cross-validation* merupakan metode evaluasi model yang memisahkan antara *train set* dan *test set*. Proses ini diulang beberapa kali dan performa model diukur berdasarkan rata-rata evaluasi skor per proses iterasi (perulangan). Lalu, untuk menilai bagaimana performa model, penulis menggunakan matrik evaluasi yang terdiri dari accuracy score, precision, recall, dan f1-score. Accuracy score merupakan nilai dari sejauh mana model dapat memprediksi dengan benar. Selanjutnya, precision merupakan hasil persentase data yang diprediksi dengan benar sebagai true positive dari keseluruhan data yang diprediksi sebagai positif. Dan terakhir f1-score merupakan ukuran rata-rata dari nilai presisi dan recall yang menjustifikasikan penilaian performa model klasifikasi.

BAB 4

HASIL DAN PEMBAHASAN

4.1 Pre-Processing

Sebelum masuk kedalam tahap modeling, data harus dipersiapkan terlebih dahulu untuk memudahkan proses analisis dan pemodelan. Data pre-processing yang kami lakukan meliputi :

4.1.1 Missing Value

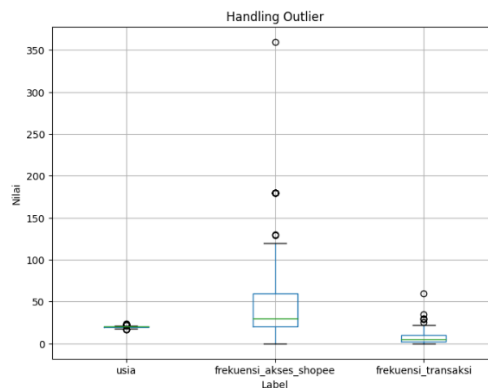
Missing value adalah informasi yang tidak tersedia untuk sebuah objek (kasus). Pada proses pengecekan missing value, didapatkan bahwa dataset yang kami gunakan tidak terdeteksi adanya missing value. Hal tersebut bisa terjadi dikarenakan dataset yang kami gunakan berasal dari data primer hasil pengumpulan kuesioner yang mengharuskan responden mengisi seluruh pertanyaan.

```
gender          0
prodi           0
angkatan       0
usia           0
ketersediaan_produk  0
kualitas_produk  0
harga          0
promosi        0
ketepatan_pengiriman  0
penanganan_keluhan  0
frekuensi_akses_shopee  0
frekuensi_transaksi  0
kepuasan_konsumen  0
skor_kepuasan   0
dtype: int64
```

Gambar 18. Hasil Pemeriksaan Missing Value

4.1.2 Handling Outlier

Outlier adalah data yang menyimpang secara ekstrim dari rata-rata sekumpulan data yang ada. Pada tahap proses *handling outlier*, kami mendeteksi bahwa terdapat beberapa variabel numerik yang memiliki nilai outlier. Variabel tersebut diantaranya adalah usia, frekuensi_akses_shopee, dan frekuensi_transaksi. Kami memilih untuk tidak menghapus nilai outlier tersebut dikarenakan nilai tersebut termasuk dalam informasi penting yang dapat mempengaruhi model.



Gambar 19. Hasil Pemeriksaan Handling Outlier

4.1.3 Encoding Variabel

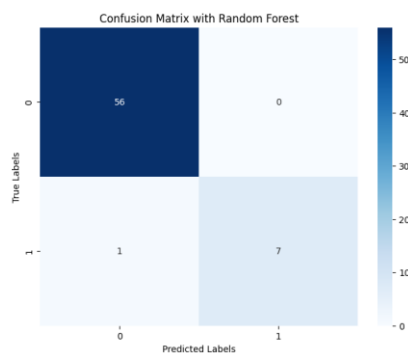
Encoding variabel adalah proses transformasi variabel kategorik menjadi bentuk numerik. Berdasarkan dataset yang kami gunakan, terdapat tujuh variabel kategorik, di antaranya : gender, ketersediaan_produk, kualitas_produk, harga, promosi, ketepatan_pengiriman, dan penanganan_keluhan. Ketujuh variabel tersebut kami encode menggunakan label encoding, sehingga bernilai 0 atau 1.

	gender	ketersediaan_produk	kualitas_produk	harga	promosi	ketepatan_pengiriman	penanganan_keluhan
0	1	0	0	0	0	1	0
1	0	0	1	1	0	1	0
2	0	0	1	1	0	1	0
3	0	0	1	1	1	1	0
4	1	0	1	1	0	1	0
...
311	0	0	1	1	1	1	0
312	0	0	0	0	1	1	0
313	0	1	1	0	0	0	1
314	0	1	0	0	1	0	1
315	1	0	0	0	0	0	1

Gambar 20. Hasil Encoding Variabel pada Variabel Kategorik

4.2 Hasil Confusion Matrix

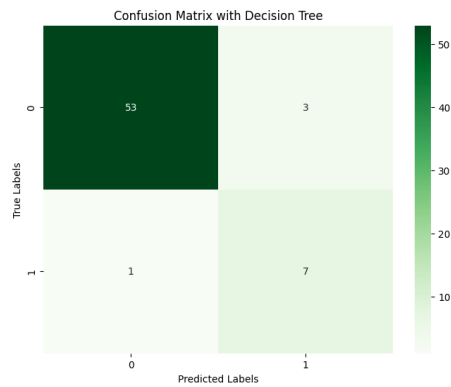
4.2.1 Random Forest



Gambar 21. Confusion Matrix Random Forest

Pada confusion matrix dari model random forest menunjukkan, data teridentifikasi True Positive berjumlah 56 data diklasifikasikan dengan benar sebagai jumlah customer puas. Lalu teridentifikasi False Negative berjumlah 1, diklasifikasikan dengan salah sebagai customer tidak puas. Kemudian teridentifikasi True Negative berjumlah 7, diklasifikasikan dengan benar sebagai customer tidak puas dan False Positive berjumlah 0, diklasifikasikan dengan salah sebagai customer puas.

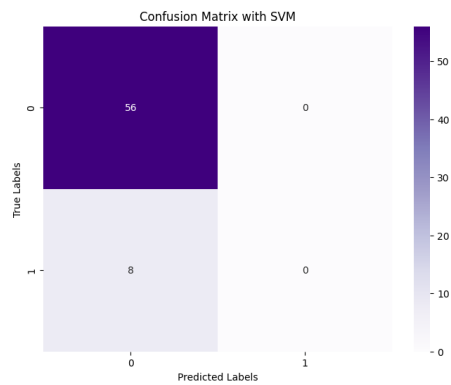
4.2.2 Decision Tree



Gambar 22. Confusion Matrix Decision Tree

Pada confusion matrix dari model Decision Tree menunjukkan, data teridentifikasi True Positive berjumlah 53 data diklasifikasikan dengan benar sebagai jumlah customer puas. Lalu teridentifikasi False Negative berjumlah 1, diklasifikasikan dengan salah sebagai customer tidak puas. Kemudian teridentifikasi True Negative berjumlah 7, diklasifikasikan dengan benar sebagai customer tidak puas dan False Positive berjumlah 3, diklasifikasikan dengan salah sebagai customer puas.

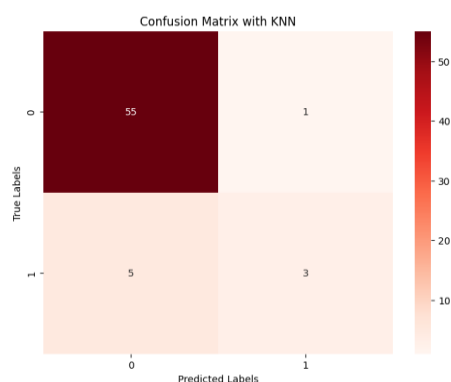
4.2.3 Support Vector Machine



Gambar 23. Confusion Matrix SVM

Pada confusion matrix dari model SVM menunjukkan, data teridentifikasi True Positive berjumlah 56 data diklasifikasikan dengan benar sebagai puas. Lalu teridentifikasi False Negative berjumlah 8, diklasifikasikan dengan salah sebagai tidak puas. Kemudian untuk data yang teridentifikasi True Negative dan False Positive berjumlah 0, artinya, tidak ada data yang diklasifikasikan dengan benar sebagai customer tidak puas dan tidak ada data yang diklasifikasikan dengan salah sebagai customer puas.

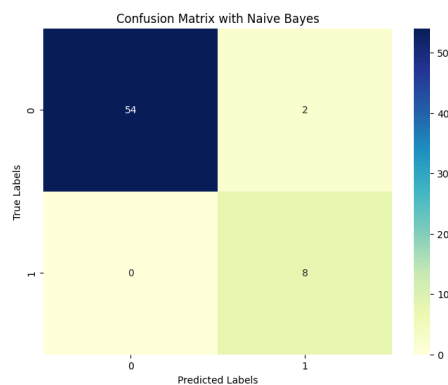
4.2.4 k-Nearest Neighbors



Gambar 24. Confusion Matrix KNN

Pada confusion matrix dari model k-Nearest Neighbors menunjukkan, data teridentifikasi True Positive berjumlah 55 data diklasifikasikan dengan benar sebagai jumlah customer puas. Lalu teridentifikasi False Negative berjumlah 5, diklasifikasikan dengan salah sebagai customer tidak puas. Kemudian teridentifikasi True Negative berjumlah 3, diklasifikasikan dengan benar sebagai customer tidak puas dan False Positive berjumlah 1, diklasifikasikan dengan salah sebagai customer puas.

4.2.5 Naïve Bayes



Gambar 25. Confusion Matrix Naïve Bayes

Pada confusion matrix dari model random forest menunjukkan, data teridentifikasi True Positive berjumlah 54 data diklasifikasikan dengan benar sebagai jumlah customer puas. Lalu teridentifikasi False Negative berjumlah 0, diklasifikasikan dengan salah sebagai customer tidak puas. Kemudian teridentifikasi True Negative berjumlah 8, diklasifikasikan dengan benar sebagai customer tidak puas dan False Positive berjumlah 2, diklasifikasikan dengan salah sebagai customer puas.

4.3 Matriks Evaluasi

4.3.1 Random Forest

Classification Report:				
	precision	recall	f1-score	support
0	0.98	1.00	0.99	56
1	1.00	0.88	0.93	8
accuracy			0.98	64
macro avg	0.99	0.94	0.96	64
weighted avg	0.98	0.98	0.98	64

Gambar 26. Classification Report Random Forest

Nilai akurasi pada model Random Forest senilai 0.984375. Berikut adalah classification report dari matriks evaluasi random forest.

Precision : 0.99 (macro average)
Recall : 0.94 (macro average)
F1-score : 0.96 (macro average)

Berdasarkan hasil pemodelan Random Forest, didapatkan nilai akurasi tertinggi dari pemodelan lainnya. Nilai akurasi pada model Random Forest senilai 0.984375. Artinya model dapat mempelajari data dan membagi ke dalam kelas-kelas dengan baik serta memiliki akurasi paling tinggi dibandingkan model yang lain. Model juga terindikasi memiliki kinerja prediksi yang baik karena nilai pada precision, recall, dan F1-score seimbang (ada pada range 0,94 - 0,99)

4.3.2 Decision Tree

Classification Report:				
	precision	recall	f1-score	support
0	0.98	0.95	0.96	56
1	0.70	0.88	0.78	8
accuracy			0.94	64
macro avg	0.84	0.91	0.87	64
weighted avg	0.95	0.94	0.94	64

Gambar 27. Classification Report Decision Tree

Nilai akurasi pada model Decision Tree senilai 0.9375. Berikut adalah classification report dari matriks evaluasi Decision Tree.

Precision : 0.84 (macro average)
Recall : 0.91 (macro average)
F1-score : 0.87 (macro average)

Berdasarkan hasil klasifikasi dengan metode decision tree didapatkan nilai akurasi sebesar 0.9375 artinya model dapat mempelajari data dan membagi ke dalam kelas-kelas dengan baik. Model juga terindikasi memiliki kinerja prediksi yang baik karena nilai pada precision, recall, dan F1-score seimbang (ada pada range 0,84 - 0,91).

4.3.3 SVM

Classification Report:				
	precision	recall	f1-score	support
0	0.88	1.00	0.93	56
1	0.00	0.00	0.00	8
accuracy			0.88	64
macro avg	0.44	0.50	0.47	64
weighted avg	0.77	0.88	0.82	64

Gambar 28. Classification Report SVM

Nilai akurasi pada model SVM senilai 0.875. Berikut adalah classification report dari matriks evaluasi SVM.

Precision : 0.44 (macro average)

Recall : 0.50 (macro average)

F1-score : 0.47 (macro average)

Dapat disimpulkan dari classification report di atas bahwa akurasi ada di angka yang cukup tinggal yaitu 0.875 artinya model dapat mempelajari data dan membagi ke dalam kelas-kelas dengan baik. Model juga terindikasi memiliki kinerja prediksi yang baik karena nilai pada precision, recall, dan F1-score seimbang (ada pada range 0,44 - 0,5).

4.3.4 kNN

Classification Report:				
	precision	recall	f1-score	support
0	0.92	0.98	0.95	56
1	0.75	0.38	0.50	8
accuracy			0.91	64
macro avg	0.83	0.68	0.72	64
weighted avg	0.90	0.91	0.89	64

Gambar 29. Classification Report KNN

Nilai akurasi pada model kNN senilai 0.90625. Berikut adalah classification report dari matriks evaluasi kNN.

Precision : 0.83 (macro average)

Recall : 0.68 (macro average)

F1-score : 0.72 (macro average)

Dapat disimpulkan dari classification report di atas bahwa akurasi ada di angka yang cukup tinggal yaitu 0.90625 artinya model dapat mempelajari data dan membagi ke dalam kelas- kelas dengan baik. Walau demikian, model tidak terindikasi memiliki kinerja prediksi yang baik karena nilai pada precision, recall, dan F1-score tidak terlalu seimbang. Terdapat rentang yang cukup jauh untuk nilai F1-score dan precision.

4.3.5 Naive Bayes

Classification Report:				
	precision	recall	f1-score	support
0	1.00	0.96	0.98	56
1	0.80	1.00	0.89	8
accuracy			0.97	64
macro avg	0.90	0.98	0.94	64
weighted avg	0.97	0.97	0.97	64

Gambar 30. Classification Report Naïve Bayes

Nilai akurasi pada model Naive Bayes senilai 0.96875. Berikut adalah classification report dari matriks evaluasi Naive Bayes.

Precision : 0.90 (macro average)

Recall : 0.98 (macro average)

F1-score : 0.94 (macro average)

Dapat disimpulkan dari classification report di atas bahwa akurasi ada di angka yang cukup tinggal yaitu 0.96875 artinya model dapat mempelajari data dan membagi ke dalam kelas- kelas dengan baik. Model juga terindikasi memiliki kinerja prediksi yang baik karena nilai pada precision, recall, dan F1-score seimbang (ada pada range 0,90 - 0,98).

BAB 5

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis yang kami lakukan pada dataset ‘Kepuasan Customer E-Shopee’ maka diperoleh beberapa kesimpulan sebagai berikut:

1. Pada tahap awal *preprocessing*, tidak terdeteksi adanya *missing value*, pada deteksi *outlier* diputuskan untuk tidak menghapus *outlier* karena data tersebut masuk kedalam informasi penting, dan untuk variabel kategorik (gender, ketersediaan_produk, kualitas_produk, harga, promosi, ketepatan_pengiriman, dan penanganan_keluhan) kami lakukan one-hot encoding untuk memudahkan pada tahap modelling selanjutnya.
2. Pada tahap modeling, diperoleh confusion matrix pada setiap model, sebagai berikut :
 - a. Random Forest: teridentifikasi True Positive berjumlah 56 data yang diklasifikasikan dengan benar sebagai jumlah customer puas. Lalu teridentifikasi False Negative berjumlah 1 artinya diklasifikasikan dengan salah sebagai customer tidak puas. Kemudian teridentifikasi True Negative berjumlah 7 yang diklasifikasikan dengan benar sebagai customer tidak puas dan False Positive berjumlah 0 artinya diklasifikasikan dengan salah sebagai customer puas.
 - b. Decision Tree: teridentifikasi True Positive berjumlah 53 data yang diklasifikasikan dengan benar sebagai jumlah customer puas. Lalu teridentifikasi False Negative berjumlah 1 artinya diklasifikasikan dengan salah sebagai customer tidak puas. Kemudian teridentifikasi True Negative berjumlah 7 yang diklasifikasikan dengan benar sebagai customer tidak puas dan False Positive berjumlah 3 artinya diklasifikasikan dengan salah sebagai customer puas.
 - c. Support Vector Machine: teridentifikasi True Positive berjumlah 56 data yang diklasifikasikan dengan benar sebagai puas. Lalu teridentifikasi False Negative berjumlah 8 yang diklasifikasikan dengan salah sebagai tidak puas. Kemudian teridentifikasi True Negative dan False Positive berjumlah 0 artinya, tidak ada data yang diklasifikasikan dengan benar sebagai customer tidak puas dan tidak ada data yang diklasifikasikan dengan salah sebagai customer puas.
 - d. KNN : teridentifikasi True Positive berjumlah 55 data yang diklasifikasikan dengan benar sebagai jumlah customer puas. Lalu teridentifikasi False Negative berjumlah 5 artinya diklasifikasikan dengan salah sebagai customer tidak puas. Kemudian teridentifikasi True Negative berjumlah 3 yang diklasifikasikan dengan benar sebagai customer tidak puas dan False Positive berjumlah 1 artinya diklasifikasikan dengan salah sebagai customer puas.
 - e. Naïve Bayes : teridentifikasi True Positive berjumlah 54 data yang diklasifikasikan dengan benar sebagai jumlah customer puas. Lalu teridentifikasi False Negative berjumlah 0 artinya diklasifikasikan dengan salah sebagai customer tidak puas. Kemudian teridentifikasi True Negative berjumlah 8 yang diklasifikasikan dengan benar sebagai customer tidak puas dan False Positive berjumlah 2 artinya diklasifikasikan dengan salah sebagai customer puas.

3. Setiap model, menghasilkan nilai matriks evaluasi, dengan nilai akurasi sebagai berikut:

- | | |
|---------------------------|------------|
| a. Random Forest | : 0.984375 |
| b. Decision Tree | : 0.9375 |
| c. Support Vector Machine | : 0.875 |
| d. K-Nearest Neighbour | : 0.96875 |
| e. Naïve Bayes | : 0.96875 |

Berdasarkan ketiga poin diatas, maka dapat diputuskan bahwa model terbaik dalam menganalisis Kepuasan Customer *E-Commerce* Shopee Pada Mahasiswa FTMM, dengan menggunakan model Random Forest yang dimana memperoleh nilai akurasi tertinggi daripada model lain. Selain tingginya nilai akurasi, terdapat beberapa alasan yang menjadikan random forest sebagai model terbaik, seperti : kemampuan handling overfitting yang lebih baik dibandingkan decision tree, random forest memiliki kemampuan yang baik dalam mengatasi masalah multikolinearitas, dan random forest mengidentifikasi fitur penting dengan memberikan peringkat berdasarkan pentingnya variabel tersebut dalam mengklasifikasikan data.

5.2 Saran

Berdasarkan analisis menggunakan kelima model machine learning diatas, berikut diperoleh beberapa saran untuk meningkatkan kepuasan konsumen pada layanan *e-commerce* Shopee :

- Meningkatkan responsivitas dan kualitas pelayanan pada konsumen, seperti menghadirkan tim *customer service* yang ramah, cepat tanggap, dan ahli dalam memecahkan masalah konsumen dengan praktis dan efisien.
- Menawarkan lebih banyak ragam produk dengan cara meningkatkan kerja sama dengan lebih banyak penjual dan memperluas variasi produk yang tersedia
- Meningkatkan kualitas mutu produk yang ditawarkan, seperti menyediakan tim khusus dalam *quality check* suatu produk.
- Meningkatkan keamanan dan privasi dengan memastikan data konsumen aman dan tidak disalahgunakan
- Membuat promosi berbasis langganan bagi konsumen loyal dengan menghadirkan berbagai macam promo maupun reward bagi konsumen yang setia untuk mempertahankan basis pengguna yang sudah ada.

DAFTAR PUSTAKA

- Binus University. (2022). "Support Vector Machine Algorithm." Retrieved from <https://sis.binus.ac.id/2022/02/14/support-vector-machine-algorithm/>
- Hamidi, H., & Daraei, A. (2016). "Analysis and Evaluation of Techniques for Myocardial Infarction Based on Genetic Algorithm and Weight by SVM."
- Kerami, D., & Murfi, H. (2004). "Kajian Kemampuan Generalisasi Support Vector Machine dalam Pengenalan Jenis Splice Site pada Barisan DNA." *Jurnal Makara Sains*, Vol. 8(3), pp. 89-95.
- Mardi, Y. (2013). "Data Mining: Klasifikasi Menggunakan Algoritma C4.5." *Jurnal Edik Informatika*, V2.i2, pp. 213-219. Diterbitkan Oleh Program Studi Pendidikan Informatika STKIP PGRI Sumbar. ISSN: 2407-0491. E-ISSN: 2541-3716.
- Prasetyo, E. (2014). "Data Mining Konsep dan Aplikasi Menggunakan MATLAB." Yogyakarta: ANDI.
- Putri, N. A. R., & Iriani, S. S. (2019). "Faktor-faktor yang mempengaruhi keputusan pembelian melalui e-commerce Shopee." *Jurnal Komunika: Jurnal Komunikasi, Media Dan Informatika*, 8(2), 69-77.
- Sabda Puta Mediti, O. C. M. (2020) "Analisis Pengaruh Harga dan Kualitas Layanan Elektronik terhadap Kepuasan Konsumen pada Pengguna Aplikasi Shopee", *Jurnal Ilmu Manajemen*, 8(4), pp. 1290–1300. doi: 10.26740/jim.v8n4.p1290-1300.
- Suhendra, G., & Krisnadi, I. (2020). "Kekuatan Shopee Sebagai E-Commerce Terpopuler Di Indonesia Saat Ini." Universitas Mercu Buana (29).
- Utari, P., Fitri, A., Setyanto, E., & Henny, C. (2020). "Belanja on-line mahasiswa di era pandemi covid-19: modifikasi perilaku konsumen." *Islamic Communication Journal*, 5(2), 143-154.
- Widodo, P. P., Handayanto, R. T., & Herlawati. (2013). "Penerapan Data Mining dengan Matlab." Bandung: Rekayasa Sains.

LAMPIRAN

Berikut dilampirkan link dataset yang digunakan pada link drive berikut ini :

<https://bit.ly/DatasetAnalisisShopee>

Berikut dilampirkan link code hasil modeling pada link berikut ini :

https://colab.research.google.com/drive/1C96QqAWUJgeLY9bd5_5McAvqZtZktPe6?usp=sharing