

***SHOP RECOMMENDATION* BERBASIS SENTIMEN  
MENGUNAKAN *FASTTEXT* DAN *XGBOOST***

**SKRIPSI**



**ANDI NURUL QALBI**

**H071191014**

**PROGRAM STUDI SISTEM INFORMASI  
DEPARTEMEN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
APRIL 2024**

***SHOP RECOMMENDATION* BERBASIS SENTIMEN  
MENGUNAKAN *FASTTEXT* DAN *XGBOOST***

**SKRIPSI**

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin Makassar

**ANDI NURUL QALBI**

**H071191014**

**PROGRAM STUDI SISTEM INFORMASI  
DEPARTEMEN MATEMATIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
APRIL 2024**

## HALAMAN PERNYATAAN KEOTENTIKAN

Yang bertanda tangan di bawah ini :

Nama : Andi Nurul Qalbi

NIM : H071191014

Program Studi : Sistem Informasi

Jenjang : S1

Menyatakan dengan ini bahwa skripsi yang saya tulis berjudul :

***SHOP RECOMMENDATION BERBASIS SENTIMEN MENGGUNAKAN  
FASTTEXT DAN XGBOOST***

Adalah benar hasil karya saya sendiri bukan merupakan pengambilan alihan tulisan orang lain dan belum pernah dipublikasikan dalam bentuk apapun.

Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa sebagian atau keseluruhan skripsi ini merupakan hasil karya orang lain, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 28 April 2024



Andi Nurul Qalbi

NIM. H071191014

**SHOP RECOMMENDATION BERBASIS SENTIMEN  
MENGUNAKAN *FASTTEXT* DAN *XGBOOST***

Disusun dan diajukan oleh:

**ANDI NURUL QALBI**

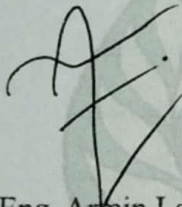
**H071191014**

Telah dipertahankan di hadapan Panitia Ujian yang dibentuk dalam rangka Penyelesaian Studi Program Sarjana pada Program Studi Sistem Informasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin dan dinyatakan telah memenuhi syarat kelulusan

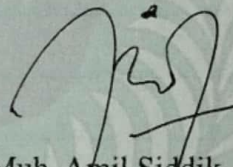
Menyetujui :

Pembimbing Utama

Pembimbing Pertama

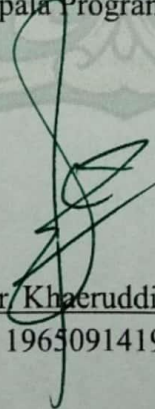


Dr. Eng. Armin Lawi, S.Si., M.Eng.  
NIP. 197204231995121001



A. Muh. Amil Siddik, S.Si., M.Si.  
NIP. 199110032019031015

Kepala Program Studi



Dr. Khaeruddin, M.Sc.  
NIP. 196509141991031003



## HALAMAN PENGESAHAN

Skripsi ini diajukan oleh :

Nama : Andi Nurul Qalbi  
NIM : H071191014  
Program Studi : Sistem Informasi  
Judul Skripsi : *SHOP RECOMMENDATION* BERBASIS  
SENTIMEN MENGGUNAKAN *FASTTEXT*  
DAN *XGBOOST*

Telah dipertahankan di depan Tim Penguji dan diterima sebagai bagian persyaratan untuk memperoleh gelar Sarjana pada Program Studi Sistem Informasi Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin.

### DEWAN PENGUJI

- |   | Tanda Tangan |
|---|--------------|
| 1. Ketua : Dr. Eng. Armin Lawi, S.Si., M.Eng.     | (.....)      |
| 2. Sekretaris : A. Muh. Amil Siddik, S.Si., M.Si. | (.....)      |
| 3. Anggota : Dr. Muhammad Hasbi, M.Sc.            | (.....)      |
| 4. Anggota : Rozalina Amran, S.T., M.Eng.         | (.....)      |

Ditetapkan di : Makassar

Tanggal : 28 April 2024



## KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah *subhanahu wa ta'ala*, karena atas limpahan rahmat dan taufiq-Nya penulis dapat menyelesaikan skripsi ini. Shalawat serta salam tidak lupa pula tercurah kepada nabi Muhammad *shallallahu `alaihi Wa Sallam* sebagai nabi terkahir sekaligus teladan bagi kehidupan dunia.

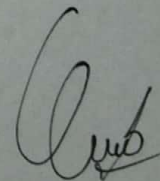
Alhamdulillah, skripsi yang berjudul "*SHOP RECOMMENDATION BERBASIS SENTIMEN MENGGUNAKAN FASTTEXT DAN XGBOOST*" yang disusun sebagai salah satu syarat untuk mencapai gelar Sarjana pada program studi Sistem Informasi fakultas Matematika dan ilmu Pengetahuan Alam Universitas Hasanuddin dapat diselesaikan dengan baik. Tidak lupa pula penulis mengucapkan terima kasih atas bantuan dan kontribusinya kepada :

1. Rektor Universitas Hasanuddin, Bapak **Prof. Dr. Ir. Jamaluddin Jompa, M. Sc**, beserta jajarannya.
2. Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam, **Dr. Eng. Amiruddin** beserta jajarannya.
3. Ketua Departemen Matematika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin, Bapak **Dr. Firman, S.Si., M.Si.**, atas seluruh ilmu dan saran-saran yang telah diberikan.
4. Ketua Program Studi Sistem Informasi Bapak **Dr. Khaeruddin, M.Sc.**, atas seluruh ilmu dan arahan bagi penulis selama masa studi penulis khususnya dalam penyusunan skripsi.
5. Pembimbing Pertama **Dr. Eng. Armin Lawi, S.Si., M.Eng.**, yang telah senantiasa membantu, membimbing, dan memberikan arahan selama masa studi penulis khususnya dalam penyusunan skripsi.
6. Bapak **A. Muh. Amil Siddik, S.Si.,M.Si.** sebagai pembimbing pendamping yang senantiasa, memberikan arahan, ide, motivavsi serta dukungan kepada penulis dalam banyak hal.

7. Bapak **Dr. Muhammad Hasbi, M.Sc.** dan ibu **Rozalina Amran, S.T., M.Eng.** sebagai tim penguji atas saran dan masukan pada penelitian yang telah dilakukan oleh penulis.
8. Bapak/Ibu Dosen Program Studi Sistem Informasi beserta seluruh tenaga pendidik yang telah memberikan ilmu dan mendidik penulis selama masa perkuliahan. Serta kepada seluruh staf dan pegawai Departemen Matematika yang telah membantu penulis terutama dalam segala proses administrasi.
9. Keluarga besar **Andi Sudirman** dan **Sukira** yang tidak pernah lelah mendoakan, memotivasi, dan memberikan dukungan moral dan materi.
10. Seluruh teman-teman program studi Sistem Informasi Angkatan 2019 yang senantiasa memberikan bantuan dan dukungan selama perkuliahan hingga selesainya penulisan skripsi penulis.

Akhir kata, penulis berharap Tuhan Yang Maha Esa melimpahkan berkat-Nya kepada semua pihak yang telah mendukung dan membantu. Peneliti menyadari bahwa pembuatan penelitian skripsi ini masih banyak kekurangan dan kelemahaannya baik dari materi maupun teknik penyajiannya. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan penulis untuk menambah kesempurnaan penelitian skripsi ini.

Makassar, 28 April 2024



Andi Nurul Qalbi  
NIM. H071191014

## PERSYARATAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

---

Sebagai civitas akademik Universitas Hasanuddin, saya yang bertanda tangan di bawah ini:

Nama : Andi Nurul Qalbi  
NIM : H071191014  
Program Studi : Sistem Informasi  
Departemen : Matematika  
Fakultas : Fakultas Matematika dan Ilmu Pengetahuan Alam  
Jenis Karya : Skripsi

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Hasanuddin **Hak Prediktor Royalti Noneklusif** (*Non-exclusive Royalty-Free Right*) atas tugas akhir saya yang berjudul :

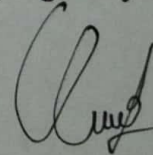
### **“SHOP RECOMMENDATION BERBASIS SENTIMEN MENGUNAKAN *FASTTEXT* DAN *XGBOOST*”**

beserta perangkat yang ada (jika diperlukan). Terkait dengan hal diatas, maka pihak Universitas Hasanuddin berhak menyimpan, mengalih-media/format-kan, mengelola dalam bentuk pangkalan data (*database*), merawat dan mempublikasikan tugas akhir pemilik Hak Cipta.

Demikian surat pernyataan ini saya buat dengan sebenarnya.

Dibuat di Makassar pada 28 April 2024

yang menyatakan



Andi Nurul Qalbi



## **ABSTRAK**

### ***SHOP RECOMMENDATION BERBASIS SENTIMEN MENGGUNAKAN FASTTEXT DAN XGBOOST***

**ANDI NURUL QALBI**

**H071191014**

Perkembangan teknologi dan telekomunikasi menjadikan E-Commerce sebagai salah satu pusat transaksi jual beli secara online yang hadir dengan banyak pilihan marketplace, beragam pelayanan dan produk berkualitas yang kian hari semakin meningkat. Dengan banyaknya produk yang disediakan, menjadikan sebagian orang kesulitan dalam memilih dan menentukan kualitas barang dari toko mana yang akan dibeli karena produk tidak dapat disentuh secara langsung. Pada penelitian ini, penulis akan membuat model yang mampu merekomendasikan toko online berdasarkan review dan rating dari produk menggunakan FastText dan XGBoost. Review akan diubah menjadi vektor dengan ukuran panjang 128 dan vektor ini akan di train menggunakan model XGBoost dengan weak model yang digunakan adalah decision tree dan dituning menggunakan RandomsearchCV dengan 3 fold. Hasil akhir akan dibuat kedalam bentuk web dengan menggunakan library Streamlit. Hasil penelitian menunjukkan model mampu bekerja dengan baik dengan score terbaik yaitu 98%. Dari model tersebut diperoleh toko dengan id 1884654 memperoleh f1-score tertinggi yaitu 99% dan menjadikan toko ini toko yang direkomendasikan pada kategori elektronik.

**Kata Kunci** : *E-Commerce, FastText, XGBoost, Streamlit*, Sentimen analisis

## ***ABSTRACT***

### ***SHOP RECOMMENDATION BASED ON SENTIMENT USING FASTTEXT AND XGBOOST***

**ANDI NURUL QALBI**

**H071191014**

The development of technology and telecommunications has made E-Commerce one of the primary platforms for online buying and selling transactions, offering numerous marketplaces, diverse services, and increasingly high-quality products. With the abundance of products available, many people find it challenging to select and determine the quality of goods from which store to purchase due to the inability to physically inspect the products. In this study, the author will create a model capable of recommending online stores based on product reviews and ratings using FastText and XGBoost. Reviews will be transformed into vectors with a length of 128, and these vectors will be trained using the XGBoost model with weak models such as decision trees, tuned using RandomsearchCV with 3 folds. The final results will be presented in the form of a web application using the Streamlit library. The research findings indicate that the model performs well with the best score being 98%. From this model, store ID 1884654 achieved the highest f1-score of 99%, making it the recommended store in the electronics category.

**Keywords** : *E-Commerce, FastText, XGBoost, Streamlit, Sentiment Analysis*

## DAFTAR ISI

<b>HALAMAN JUDUL .....</b>	<b>ii</b>
<b>HALAMAN PERNYATAAN KEOTENTIKAN ...Error! Bookmark not defined.</b>	
<b>HALAMAN PERSETUJUAN PEMBIMBING ....Error! Bookmark not defined.</b>	
<b>HALAMAN PENGESAHAN.....Error! Bookmark not defined.</b>	
<b>KATA PENGANTAR.....</b>	<b>vi</b>
<b>PERSETUJUAN PUBLIKASI KARYA ILMIAH Error! Bookmark not defined.</b>	
<b>ABSTRAK .....</b>	<b>viii</b>
<b><i>ABSTRACT</i> .....</b>	<b>x</b>
<b>DAFTAR ISI .....</b>	<b>xi</b>
<b>DAFTAR GAMBAR.....</b>	<b>xiii</b>
<b>DAFTAR TABLE .....</b>	<b>xiv</b>
<b>BAB I PENDAHULUAN.....</b>	<b>1</b>
1. 1 Latar Belakang .....	1
1. 2 Rumusan Masalah.....	3
1. 3 Batasan Masalah .....	3
1. 4 Tujuan Penelitian .....	3
1. 5 Manfaat Penelitian .....	4
<b>BAB II TINJAUAN PUSTAKA .....</b>	<b>5</b>
2. 1 <i>E-Commerce</i> .....	5
2. 2 <i>Shopee</i> .....	6
2. 3 <i>Business Intelligence (BI)</i> .....	6
2. 4 Sentimen Analisis .....	7
2. 5 <i>Python</i> .....	7
2. 6 <i>Machine Learning</i> .....	7
2. 7 <i>Natural Language Processing (NLP)</i> .....	8
2. 8 <i>FastText</i> .....	8

2. 9 <i>Decision Tree</i> .....	9
2. 10 <i>XGBoost</i> .....	9
2. 11 <i>Streamlit</i> .....	10
<b>BAB III METODE PENELITIAN .....</b>	<b>11</b>
3. 1 Waktu dan Lokasi Penelitian .....	11
3. 2 Alur Program ( <i>flowchart</i> ) .....	11
3.2.1 <i>Data Acquisition</i> .....	12
3.2.2 <i>Preprocessing</i> .....	13
3.2.3 <i>Language Modeling</i> .....	14
3.2.4 <i>Text Representation</i> .....	16
3.2.5 <i>Training Model</i> .....	17
3.2.6 <i>Evaluasi</i> .....	20
3.2.8 <i>Deployment</i> .....	22
3. 3 Rancangan Website .....	22
3. 5 Instrumen Penelitian .....	24
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>	<b>26</b>
4. 1 Evaluasi Model <i>FastText</i> .....	26
4. 2 Implementasi dan Evaluasi Model <i>XGBoost</i> .....	28
4. 2 Evaluasi Website Rekomendasi Toko <i>E-Commerce</i> .....	29
<b>BAB V PENUTUP .....</b>	<b>33</b>
5. 1 Kesimpulan .....	33
5. 2 Saran .....	33
<b>DAFTAR PUSTAKA .....</b>	<b>35</b>

## DAFTAR GAMBAR

Gambar 3. 1 Alur Penelitian.....	12
Gambar 3. 2 <i>Preprocessing</i> .....	13
Gambar 3. 3 <i>Language Modeling</i> .....	15
Gambar 3. 4 Perbedaan <i>Skip-gram</i> dan CBOW .....	16
Gambar 3. 5 <i>Text Representation</i> .....	17
Gambar 3. 6 <i>Train XGBoost Model</i> .....	18
Gambar 3. 7 Split data <i>train</i> dan data <i>test</i> .....	18
Gambar 3. 8 Parameter yang akan ditunning .....	19
Gambar 3. 9 <i>Boosting Algorithm</i> .....	19
Gambar 3. 10 <i>Decision Tree</i> .....	20
Gambar 3. 11 <i>Confusion Matrics</i> .....	20
Gambar 3. 12 Desain <i>landing page</i> .....	23
Gambar 3. 13 Desain web setelah data terinput .....	24
Gambar 4. 1 <i>Word to Vec</i> .....	26
Gambar 4. 2 <i>Text Similarity</i> .....	27
Gambar 4. 3 Hasil Fitting Model <i>XGBoost</i> .....	28
Gambar 4. 4 Score model untuk setiap toko .....	29
Gambar 4. 5 Input data .....	29
Gambar 4. 6 Preview data .....	30
Gambar 4. 7 <i>Language Modeling (FastText)</i> .....	30
Gambar 4. 8 <i>XGBoost Model &amp; Evaluation</i> .....	31
Gambar 4. 9 <i>Shop Recommendation</i> .....	32

## DAFTAR TABLE

Tabel 3. 1 waktu Penelitian .....	11
Tabel 3. 2 Informasi dataset .....	13

## **BAB I**

### **PENDAHULUAN**

#### **1. 1 Latar Belakang**

Perkembangan teknologi dan telekomunikasi yang begitu pesat memberikan kemudahan dalam kegiatan sehari-hari, salah satunya *E-Commerce*. Memberikan kemudahan untuk melakukan transaksi jual beli secara online (Putra et al., 2017), membuat hal ini dimanfaatkan oleh banyak pihak seperti produsen, Industri Kecil Menengah (IKM), Usaha Mikro Kecil Menengah (UMKM), pemasok, maupun distributor untuk menjangkau calon pembeli yang membutuhkan produknya. Begitu juga di Indonesia. Data mencatat bahwa Indonesia merupakan negara ke-20 dengan perkembangan *E-Commerce* terbesar di dunia dan negara ke-5 dengan pertumbuhan customer *E-Commerce* di dunia (World Retail Congress, 2020).

Salah satu marketplace regional yang cukup populer di Indonesia adalah shopee, dengan jumlah pengunjung per bulan sekitar 131.296.667 kunjungan pada kuartal kedua tahun 2022 (Iprice, 2022) menjadikan shopee sebagai *marketplace* dengan jumlah pengunjung terbanyak kedua di Indonesia setelah tokopedia. Shopee juga merupakan *marketplace* dengan popularitas terbanyak di Instagram dengan jumlah sekitar 8.727.742 orang, mengalahkan tokopedia dan lazada (Iprice, 2022).

Dengan kepopuleran tersebut, Shopee tentunya menyediakan banyak pilihan produk dan layanan yang berkualitas. Tentu saja terdapat kekurangan karena konsumen tidak dapat melihat dan menyentuh produk secara langsung sehingga foto dan deskripsi produk memiliki peran penting dalam hal ini (Fauzi & Lina, 2021). Penelitian menunjukkan bahwa foto yang menarik dinilai berpengaruh positif pada minat beli konsumen pada shopee (Servanda et al., 2019). Namun hal ini bertentangan dengan penelitian lainnya yang menyatakan foto bukan merupakan faktor utama untuk menilai kualitas produk (Munir et al., 2019).

Faktor lain yang digunakan untuk menilai kualitas produk adalah ulasan produk. Ulasan atau *online customer review* merupakan sumber informasi utama untuk mengetahui kualitas produk yang akan dibeli (Ardianti & Widiartanto, 2019). Review dan rating online dari pembeli sebelumnya dapat memberikan penilaian yang akurat karena diberikan antara sesama pelanggan sehingga memiliki tingkat kepercayaan yang lebih tinggi. Review merupakan salah satu bentuk *electronic word of mouth* (Sinaga & Sulistiono, 2020) yang dinilai mampu mempengaruhi minat beli konsumen. Sedangkan rating merupakan salah satu opini dari konsumen sebelumnya untuk memberikan keyakinan untuk membeli produk. Rating dinilai sebagai ukuran rekomendasi konsumen seberapa bagus produk. Rating biasanya ditampilkan sebagai skor bintang yang berjumlah 1-5 yang menunjukkan semakin banyak jumlah bintang pada produk, maka dinilai semakin baik, begitu pula sebaliknya.

Review dan rating tersebut dapat digunakan untuk melakukan sentimen analisis dan mengekstrak *insight* dari suatu produk. Dengan sentimen yang telah diekstrak tersebut, kita bisa membandingkannya dengan produk yang sama dari toko yang berbeda. Akan tetapi, jika jumlah review dalam suatu produk terlalu banyak, cara manual ternilai tidak efektif. Oleh karena itu diperlukan sebuah model *machine learning* yang mampu mengklasifikasikan review tersebut yang dalam penelitian ini algoritma yang digunakan adalah *Extrem Gradient Boosting (XGBoost)*. *XGBoost* dinilai sebagai algoritma yang cepat dan efektif dalam masalah klasifikasi teks (Handayani et al., 2017). *Machine learning* model tidak mampu memahami bahasa alami (*natural language*), sehingga data review terlebih dulu akan diubah menjadi vektor (Kusumawati & Faraby, 2021) yang disebut *word vector* menggunakan *FastText*. Diharapkan dengan menggunakan *FastText*, masalah seperti *out of vocabulary (OOV)* dapat teratasi mengingat dalam data review banyak kata yang typo (*typographical error*) dan membuat representasi vektornya lebih bermakna.



## 1. 2 Rumusan Masalah

Rumusan masalah yang akan dibahas dalam penelitian ini adalah sebagai berikut:

1. Bisakah algoritma *FastText* mengatasi masalah *out of vocabulary* (OOV) pada sentimen analisis?
2. Bagaimana mengimplementasikan algoritma *FastText* dan *XGBoost* pada sentimen analisis untuk *Shop Recommendation*?
3. Bagaimana hasil evaluasi sentimen analisis dengan *FastText* menggunakan algoritma *XGBoost*?

## 1. 3 Batasan Masalah

Pembahasan yang akan dibahas pada penelitian ini akan dibatasi pada hal-hal berikut:

1. Program ini akan berupa website yang dirancang menggunakan bahasa python,
2. Dataset yang digunakan merupakan dataset dari review produk *E-Commerce* dan setidaknya mempunyai fitur review/ komentar dan rate/sentimen,
3. Data produk yang digunakan minimal memiliki jumlah sentimen 30
4. Output dari program ini adalah website dengan menggunakan model *machine learning* dengan algoritma *XGBoost*, hasil evaluasi masing-masing produk, dan peringkat shop berdasarkan sentimennya,
5. Metric yang digunakan adalah *classification metric* yaitu *accuracy score*, *precision*, *recall*, dan *f1-score*.

## 1. 4 Tujuan Penelitian

Berdasarkan rumusan masalah, maka tujuan dari penelitian ini adalah sebagai berikut :

1. Untuk menerapkan algoritma *FastText* dapat digunakan untuk mengatasi masalah *out of vocabulary* (OOV) pada sentimen produk,

2. Untuk mengimplementasikan algoritma *FastText* dan *XGBoost* agar dapat digunakan untuk merekomendasikan toko berdasarkan sentimen,
3. Untuk merancang dan membangun website yang dapat melakukan prediksi sentimen dan evaluasinya berdasarkan algoritma *XGBoost*.

### **1.5 Manfaat Penelitian**

Berdasarkan rumusan masalah dan tujuan penelitian maka dapat diharapkan penelitian ini dapat memberi manfaat yaitu :

1. Memberikan kemudahan kepada konsumen untuk memilih dan membandingkan satu toko dengan toko lain berdasarkan sentimen produk dan kategorinya,
2. Membantu produsen untuk mengatur strategi penjualan dengan melihat sentimen positif dan negatif dari konsumen.

## BAB II

### TINJAUAN PUSTAKA

#### 2.1 *E-Commerce*

*E-Commerce* adalah istilah yang digunakan untuk menunjukkan kegiatan perdagangan dan pertukaran berupa barang, informasi atau jasa yang dilakukan antara pembeli dan penjual untuk melakukan proses penjualan dan pembelian melalui media internet (Rizqullah et al., 2022). Di Indonesia sendiri, ada beberapa model *E-Commerce* yang digunakan, seperti:

- a. B2B (*Business to Business*), transaksi ini dilakukan oleh dua pihak yang sama-sama memiliki kepentingan bisnis. Contohnya transaksi online antara produsen dan supplier.
- b. B2C (*Business to Consumer*), terjadi layaknya transaksi jual beli biasa antara pelaku bisnis dan konsumen. Konsumen membeli produk yang ditawarkan oleh pelaku bisnis.
- c. C2C (*Consumer to Consumer*), adalah model bisnis dimana konsumen menjual atau membeli produk dan layanan kepada konsumen lain melalui platform atau pasar online.
- d. C2B (*Consumer to Business*), adalah model bisnis dimana konsumen menawarkan produk atau layanan mereka kepada perusahaan atau bisnis. Contohnya seorang influencer atau *freelancer* menawarkan jasa mereka kepada perusahaan untuk dipromosikan atau digunakan dalam pemasaran.

Layanan *E-Commerce* makin banyak yang bermunculan dan populer dengan cepat. Di antara layanan tersebut seperti shopee, tokopedia, bukalapak, bli-bli.com dan lain-lain. Dengan pangsa pasar yang sangat besar, banyak penyedia layanan *E-Commerce* ini berlomba-lomba untuk menjadi yang terdepan. Persaingan diantara penyedia layanan *E-Commerce* pun secara tidak langsung semakin ketat (Felita & Oktivera, 2019).

## 2. 2 Shopee

Shopee merupakan sebuah platform belanja online yang diluncurkan pada tahun 2015 di Singapura oleh perusahaan *sea limited* (sebelumnya dikenal dengan nama Garena). Shopee pertama kali meluncur sebagai marketplace dengan konsep *consumer to consumer* (C2C). Namun semenjak meluncurkan Shopee Mall, yang merupakan platform online shop untuk brand ternama, marketplace ini beralih ke model hibrid *consumer to consumer* (C2C) dan *business to consumer* (B2C).

Shopee merupakan *E-Commerce* yang banyak memiliki cabang yang tersebar di kawasan Taiwan dan Asia Tenggara. Shopee sebagai salah satu platform *E-Commerce* telah menyajikan kemudahan dan keamanan bagi pelanggan yang menginginkan belanja secara aman, cepat, dan mudah secara online melalui layanan logistik yang terintegrasi. Dengan menyediakan kemudahan bagi penjual untuk memasarkan barang serta menghadirkan berbagai macam kategori produk untuk pembeli Shopee bertujuan untuk menghadirkan pengalaman berbelanja di berbagai platform menjadi mudah dan efisien bagi penggunanya sehingga dapat menjadi salah satu platform *E-Commerce* terpopuler di era ini (Rizqullah et al., 2022).

## 2. 3 Business Intelligence (BI)

*Business Intelligence* (BI) adalah suatu pendekatan yang mengintegrasikan teknologi, proses, dan aplikasi untuk mengumpulkan, menganalisis, dan menyajikan informasi bisnis. Organisasi menggunakan BI untuk mengoptimalkan pengambilan keputusan dengan memahami data mereka secara lebih mendalam. Proses BI melibatkan ekstraksi data dari berbagai sumber, pengolahan untuk menghasilkan wawasan yang berguna, dan penyajian informasi melalui laporan, dashboard, atau visualisasi data. Dengan BI, perusahaan dapat merespons perubahan pasar dengan lebih cepat, meningkatkan efisiensi operasional, dan mendapatkan pemahaman yang lebih baik terhadap pelanggan dan tren bisnis.

## 2. 4 Sentimen Analisis

Analisis sentimen adalah proses mengidentifikasi, mengklasifikasikan, dan mengevaluasi sentimen atau opini yang terkandung dalam teks, baik itu dalam bentuk teks berita, ulasan produk, posting media sosial, atau komentar pelanggan. Tujuannya adalah untuk memahami perasaan atau sikap positif, negatif, atau netral yang terkait dengan topik yang sedang dibahas. Sentimen analisis termasuk salah satu teks klasifikasi dan algoritma yang umumnya digunakan adalah Naïve Bayes (NB), Support Vector Machine (SVM) dan Entropi Maksimum (EM) (Li & Qiu, 2017).

## 2. 5 Python

Python adalah salah satu bahasa pemrograman tingkat tinggi dengan sintaksis yang sederhana dan mudah dipahami serta memiliki banyak dukungan library dan modul yang dapat digunakan. Python mendukung paradigma *object oriented programming* (oop) dan *funksional programming*, dan bahasa ini sering digunakan dalam pengembangan *software*, analisis data, kecerdasan buatan, dan pengembangan web. Python memiliki lingkungan pengembangan yang kuat seperti Jupyter Notebook dan banyak pustaka populer seperti NumPy, Pandas, dan TensorFlow. Platformnya dapat dijalankan di berbagai sistem operasi, dan tersedia banyak sumber daya pembelajaran untuk mendukung pengguna Python.

## 2. 6 Machine Learning

*Machine learning* merupakan suatu bidang ilmu yang menggabungkan statistik, matematika, dan ilmu komputer untuk mengajarkan komputer bagaimana belajar dan mengenali pola dari data dan melakukan task tertentu seperti klasifikasi, regresi, klusterisasi, dan lain lain tanpa harus diprogram secara eksplisit.

*Machine learning* umumnya melibatkan beberapa tahapan dalam prosesnya, diantaranya pengumpulan data (*data acquisition*), pemrosesan dan persiapan data

(*data preprocessing*), pemilihan model atau algoritma yang sesuai, training model menggunakan data yang diketahui, evaluasi performa model, dan penggunaan model yang telah di-training untuk melakukan prediksi atau mengambil keputusan pada data baru.

*Machine learning* digunakan dalam berbagai bidang, seperti *face detection*, deteksi penipuan, rekomendasi produk, pengenalan suara, analisis citra digital, dan *natural language processing* (NLP). Ada beberapa jenis pendekatan *machine learning* yang biasa digunakan, seperti *supervise learning*, *unsupervise learning*, dan lain lain, yang masing-masing pendekatan tersebut memiliki tujuan dan teknik yang berbeda.

### **2.7 Natural Language Processing (NLP)**

*Natural Language Processing* (NLP) adalah cabang ilmu komputer yang berkaitan dengan kemampuan komputer untuk berinteraksi dan memahami bahasa manusia. NLP melibatkan penggunaan algoritma, model statistik, dan teknik *machine learning* untuk memproses, menganalisis, dan menghasilkan teks atau ucapan yang memiliki makna. NLP digunakan dalam berbagai aplikasi, seperti mesin pencari, *chatbot*, analisis sentimen media sosial, penerjemah otomatis, dan banyak lagi. Tujuannya adalah untuk memungkinkan komputer untuk bekerja dengan teks dan bahasa manusia dengan cara yang semirip mungkin dengan pemahaman dan komunikasi manusia.

### **2.8 FastText**

*FastText* adalah suatu library yang digunakan untuk *natural language processing* (NLP) dan *machine learning*. Dikembangkan oleh Facebook AI Research, *FastText* memberikan metode yang efisien untuk membangun model *word embedding* (Kusumawati & Faraby, 2021) dan mengklasifikasikan teks berdasarkan konteks kata-kata.

Kelebihan *FastText* dari embedder lain adalah kemampuannya dalam memahami kata-kata yang jarang muncul dalam korpus ataupun kata kata yang

vahkan tidak ada di dalam korpus dengan mengurai mereka menjadi sub-kata atau karakter. Dengan menggunakan teknik ini, *FastText* dapat mempelajari representasi kata dengan lebih baik dan memberikan hasil yang lebih akurat dalam task seperti *text classification*, *sentimen analysis*, dan *machine translation*.

## **2.9 Decision Tree**

*Decision tree* adalah salah satu algoritma yang cukup populer dalam *machine learning* yang digunakan untuk melakukan klasifikasi dan regresi. *Decision tree* menggambarkan keputusan dan konsekuensi dari suatu keputusan dalam bentuk struktur berhingga yang menyerupai pohon. Setiap simpul dalam pohon tersebut mewakili suatu keputusan berdasarkan fitur-fitur data, sedangkan cabang-cabangnya merepresentasikan kemungkinan hasil keputusan tersebut. *Decision tree* dibuat dengan cara membagi dataset menjadi subset-subset yang lebih kecil berdasarkan fitur-fitur yang paling penting dan signifikan. Tujuan dari *decision tree* adalah untuk membuat prediksi yang akurat dan mudah diinterpretasikan berdasarkan aturan keputusan yang dihasilkan.

Keuntungan dari *decision tree* adalah kemampuannya dalam menangani data kategorikal dan numerikal, serta kemampuannya untuk mengatasi masalah yang kompleks dengan sederhana. *Decision tree* juga dapat menangani interaksi antara variabel dan mengidentifikasi variabel yang paling penting dalam membuat prediksi. Namun, *decision tree* juga rentan terhadap *overfitting*, terutama saat digunakan dalam dataset yang kompleks. Dalam praktiknya, *decision tree* sering digunakan sebagai komponen utama dalam model *ensemble* seperti *random forest* dan *XGBoost* untuk meningkatkan performa dan ketahanan terhadap *overfitting*.

## **2.10 XGBoost**

*Extreme Gradient Boosting (XGBoost)* adalah salah satu library yang cukup populer dalam *machine learning* yang dirancang untuk meningkatkan performa model berbasis pohon (*tree-based models*) seperti *Decision Trees* dan *Random*

*Forests*. *XGBoost* menggunakan teknik *ensemble learning* yang menggabungkan banyak pohon keputusan (*decision tree*) untuk meningkatkan akurasi prediksi.

*XGBoost* juga merupakan algoritma *machine learning* yang populer di platform Kaggle dan sering digunakan dalam kompetisi karena kemampuannya menghasilkan prediksi yang akurat dan performa yang tinggi. Keunggulan utama *XGBoost* adalah kemampuannya dalam mengoptimalkan fungsi objektif yang beragam, termasuk *loss regressin function* dan *clasification*. Algoritma ini menggunakan pendekatan yang disebut *gradient boosting*, di mana setiap pohon berikutnya berusaha memperbaiki kesalahan yang dibuat oleh pohon sebelumnya. Hal ini menghasilkan model yang kuat dan mampu menangani data yang kompleks dengan baik.

## **2. 11 Streamlit**

Streamlit adalah suatu *framework open-source* untuk membuat webapp interaktif dengan menggunakan bahasa pemrograman Python. Dengan Streamlit, script Python dapat diubah menjadi aplikasi web yang menarik dan mudah digunakan, tanpa memerlukan pengetahuan web development yang mendalam. Dengan menggunakan streamlit, data dapat ditampilkan secara interaktif, membuat visualisasi data, membangun model *machine learning*, dan lainnya. Streamlit secara otomatis menyediakan widget dan tata letak yang dapat diatur dengan mudah, sehingga pengguna dapat fokus pada analisis data dan visualisasi tanpa harus repot memikirkan detail tampilan web. Hal ini membuat Streamlit menjadi pilihan populer untuk *fast prototyping* dan deploy aplikasi sederhana berbasis web menggunakan Python.