

# ANALISIS MASALAH PADA DATA REVIEW APLIKASI TERHADAP LAYANAN E-COMMERCE MENGUNAKAN METODE TEXT CLASSIFICATION

*by* Fathoni Fathoni

---

**Submission date:** 11-May-2023 10:33AM (UTC+0700)

**Submission ID:** 2090066448

**File name:** 1448-5378-1-PB.pdf (375.73K)

**Word count:** 5484

**Character count:** 34877

# ANALISIS MASALAH PADA DATA REVIEW APLIKASI TERHADAP LAYANAN E-COMMERCE MENGGUNAKAN METODE TEXT CLASSIFICATION

Ahmad Rafid Rizqullah<sup>1</sup>, Ari Wedhasmara<sup>2</sup>, Rahmat Izwan Heroza<sup>3</sup>, Apriansyah Putra<sup>4</sup>, Fathoni<sup>5</sup>, Pacu Putra<sup>6</sup>

<sup>1,2,3,4,5,6</sup>Fakultas Ilmu Komputer, Universitas Sriwijaya

Email: <sup>1</sup>ahmadrafidrizqullah@gmail.com, <sup>2</sup>a\_wedhasmara@unsri.ac.id, <sup>3</sup>rahmatheroza@unsri.ac.id\*, <sup>4</sup>apriansyah@unsri.ac.id, <sup>5</sup>fathoni@unsri.ac.id, <sup>6</sup>pacuputra@unsri.ac.id

**Abstrak**—Shopee merupakan salah satu e-commerce model e-commerce Customer-To Customer (C2C) menggunakan aplikasi sebagai tempat untuk melakukan bisnis. Shopee memberikan beberapa pilihan akses kepada customer diantaranya melalui website dan mobile platform demi kemudahan kegiatan berbelanja online. Untuk mencari informasi tentang peluang yang dapat ditingkatkan dari layanan yang ada Shopee dapat memanfaatkan ulasan dari pelanggan untuk mengetahui masalah yang terjadi pada layanan yang diberikan kepada customer. Review dari customer e-commerce berbasis aplikasi tersebut dapat menjadi solusi yang lebih menjelaskan apa yang disukai dan tidak disukai customer sehingga bisa digunakan sebagai alat untuk menilai aspek-aspek pada layanan e-commerce yang dinilai bermasalah. Dilakukan klasifikasi review ke dalam beberapa aspek peningkatan layanan e-commerce menggunakan teknik Support Vector Machine (SVM) dengan kernel sigmoid, linear dan RBF. Dari penelitian ini didapatkan bahwa SVM dengan kernel sigmoid mendapatkan tingkat akurasi tertinggi yaitu 80%. Selain itu, juga diperoleh rangkuman masalah berdasarkan aspek yang bisa diperbaiki oleh pihak e-commerce untuk tetap meningkatkan layanan.

**Kata kunci** — E-Commerce, Klasifikasi Teks, Shopee, Ulasan

**Abstract**—Shopee is one of the e-commerce customer-to-customer (C2C) e-commerce models using applications as a place to do business. Shopee provides several access options for customers, including through the website and mobile platform for the convenience of online shopping activities. To find information about opportunities that can be improved from existing services, Shopee can take advantage of reviews from customers to find out problems that occur in the services provided to customers. Reviews from application-based e-commerce customers can be a solution that better explains what customers like and don't like so that it can be used as a tool to assess aspects of e-commerce services that are considered problematic. The review classification was carried out into several aspects of improving e-commerce services using the Support Vector Machine (SVM) technique with sigmoid, linear and RBF kernels. From this research, it was found that SVM with sigmoid kernel got the highest level of accuracy, that is 80%. In addition, a summary of problems is also obtained based on aspects that can be improved by e-commerce parties to continue to improve services.

**Keywords**— E-Commerce, Review, Shopee, Text Classification

## 1. PENDAHULUAN

Perkembangan dibidang teknologi dan telekomunikasi telah mendukung kemajuan teknologi internet. Hal tersebut membuat *customer e-commerce* berkembang karena semakin banyak orang menggunakan internet untuk mempermudah aktivitas mereka. Biro riset World Retail Congress telah merilis data yang menyatakan bahwa negara ke-20 dengan perkembangan *e-commerce* terbesar di dunia adalah Indonesia. Selain itu pertumbuhan *customer e-commerce* di Indonesia juga menempati peringkat ke-5 di dunia [1].

Berdasarkan data Iprice pada tahun 2020 Persentase lalu lintas *website e-commerce* secara keseluruhan di Indonesia menempatkan Shopee di urutan teratas sebagai *e-commerce* yang paling banyak dikunjungi. Berdasarkan kunjungan ke situs Shopee rata-rata setiap bulan mencapai jumlah 96.532.300 kunjungan [2]. Sebagai salah satu *e-commerce* model *e-commerce Customer-To Customer* (C2C) Shopee menggunakan aplikasi sebagai tempat untuk melakukan bisnis. Untuk menunjang hal tersebut, Shopee memberikan beberapa pilihan akses kepada *customer* diantaranya melalui *website* dan *mobile platform* demi kemudahan kegiatan berbelanja online [3].

Salah satu aplikasi mobile Shopee, terutama yang berbasis *android* dapat diunduh melalui Google Play. Google Play adalah *marketplace* aplikasi berbasis *android* yang menyediakan berbagai macam aplikasi seperti *games*, *e-commerce*, *e-wallet*, sosial media, dan pemutar musik. Fitur *review* dari para pengguna merupakan salah satu fitur yang dimiliki Google Play yang dapat digunakan untuk melihat *review* dari pengguna-pengguna lainnya [4].

Selain itu Google Play memiliki fitur untuk menganalisis *review* dari pengguna aplikasi sehingga memudahkan pembuat aplikasi memantau keadaan aplikasinya berdasarkan kata-kata atau topik yang sering muncul di *review* berkaitan dengan desain, privasi, profil, penggunaan sumber daya, kecepatan, stabilitas, *update*, *uninstall* dan kegunaan dari aplikasi tersebut [5].

*E-commerce* berbasis aplikasi dapat memanfaatkan *review* dari *customer* untuk mengetahui masalah yang terjadi pada layanan yang diberikannya kepada *customer*. Google Play memiliki fitur *review* yang dapat digunakan sebagai tempat mendapatkan *review* dari *customer* untuk mengetahui masalah pada beberapa aspek layanan yang belum diketahui dan harus segera diperbaiki dari Shopee. *Review* dari *customer e-commerce* berbasis aplikasi tersebut dapat menjadi solusi yang lebih menjelaskan apa yang disukai dan tidak disukai *customer* sehingga bisa digunakan sebagai alat untuk menilai aspek-aspek pada layanan *e-commerce* yang dinilai bermasalah [6].

*Review* dari *customer* telah mendapat banyak perhatian dari perusahaan IT terkemuka karena memahami opini publik memungkinkan mereka untuk meningkatkan kepuasan *customer* mereka [7]. Selain itu, *Review* dapat juga membantu *customer* untuk memutuskan apakah mereka akan menggunakan layanan dari suatu *e-commerce*. Hal tersebut terjadi disebabkan oleh *review* terbaru dari *customer* lain tersebut dipengaruhi dengan perspektif *customer-customer* sebelumnya yang sudah pernah menggunakannya [6].

Penelitian sebelumnya berkaitan dengan Text Classification kategori pada berita dilakukan oleh [8] dengan menggunakan metode *Support Vector Machine* (SVM) dan fitur ekstraksi menggunakan TF-IDF. Penelitian lebih lanjut oleh [9] Text Classification pada *review* pengguna aplikasi mandiri *online* di Google Play dengan menggunakan *Information Gain* dan *Naive Bayes*.

Penelitian ini bertujuan untuk mencari informasi tentang peluang yang dapat ditingkatkan pada layanan *e-commerce* dari komplain atau keluhan *review customer* pada Google Play. Meskipun Google Play juga memiliki analisis *review* yang mengkategorikan *review* berdasarkan masalah teknis aplikasi, namun pada penelitian ini pengkategorian akan lebih di titik beratkan kedalam aspek-aspek bisnis pada aplikasi. Hal tersebut dilakukan dengan mengklasifikasi *review* kedalam beberapa aspek dengan menggunakan *Text Classification* untuk mengklasifikasikan aspek ke dalam 8 komponen dari *integrated manajemen service* yang digunakan untuk membuat strategi kegiatan pemasaran, operasi, dan sumber daya manusia dari suatu perusahaan, di antaranya adalah *Physical Evidence, Productivity & Quality, Process, Distribution, Promotion, People, Price, and Product* [10]. Pengklasifikasian menggunakan 7 Komponen *integrated service management* tersebut pernah dilakukan oleh [11] dalam penelitian pada *online review* yang bertujuan untuk mengevaluasi kinerja layanan Airbnb. Setelah melakukan klasifikasi, penelitian dilanjutkan dengan mengekstrak dan mengeksplorasi informasi untuk menemukan kata-kata yang sering dibicarakan oleh *customer* pada setiap *review* yang terkelompok berdasarkan aspek. Dari hasil tersebut didapat lah rangkuman masalah berdasarkan aspek yang bisa diperbaiki oleh pihak *e-commerce* untuk tetap meningkatkan layanan.

## 2. TINJAUAN PUSTAKA

### 2.1. E-Commerce

Pengertian *E-commerce* merupakan kegiatan perdagangan pertukaran berupa barang, informasi atau jasa yang dilakukan antara pembeli dan penjual untuk melakukan proses penjualan dan pembelian melalui media internet [12]. Sedangkan berdasarkan sumber yang lain menjelaskan bahwa *e-commerce* adalah aktivitas-aktivitas kegiatan bisnis yang melibatkan pedagang perantara, konsumen, penyedia jasa, dan manufaktur melalui internet. Model-model bisnis yang dipakai oleh *e-commerce* di Indonesia diantaranya adalah sebagai berikut:

#### a. Customer-To-Customer (C2C)

*Customer-To-Customer* (C2C) mengandalkan *website* untuk tempat berdagang, promosi, dan melakukan transaksi uang secara *online* dalam model bisnisnya. Berikut ini beberapa indikator utama bagi sebuah *website marketplace*:

1. *website* sudah dapat mendukung Seluruh transaksi *online*
2. Adanya *Shopping mall*
3. Penjual yang ada bersifat individual

Komis yang didapat dari penjual menjadi salah satu keuntungan yang didapat dengan menerapkan bisnis model ini. Model ini sangat mirip sekali dengan *marketplace*, tetapi dalam model ini penjual yang menjual barang harus melalui proses verifikasi yang ketat agar dapat membentuk suasana kenyamanan bagi pelanggan.

#### b. Business-To-Consumer (B2C)

Dengan bermodalkan sebuah toko *online* dengan menggunakan *website* sendiri sebagai tempat untuk menjual produk kepada pembeli secara *online*. Dengan menggunakan model bisnis tersebut keunggulan yang dimiliki untuk pemilik toko *online* adalah kebebasan untuk mengatur *website* dalam hal tampilan sesuai keinginan sang pemilik serta dapat membuat blog untuk meningkatkan SEO dari *website*. Keuntungan yang didapat dari model tersebut adalah pemasangan iklan dan penjualan dari produk sehingga sangat relevan bagi mereka yang menginginkan kebebasan mandiri dalam mengalokasikan sumber daya yang ada.

## 2.2. Shopee

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 perangkat seluler menjadi menyenangkan dan lancar bagi semua penggunanya sehingga dapat menjadi salah satu platform e-commerce terpopuler di era ini

## 2.3. Online Review

Ada Beberapa media yang digunakan pada *website-website e-commerce* untuk mengukur kepuasan pelanggan terhadap suatu produk, layanan atau jasa. Salah satu dari media tersebut adalah *online review* sebagai sistem saran dan keluhan pada *website* [13]. Arti kata *online review* dalam Kamus Besar Bahasa Indonesia (KBBI), memiliki pengertian kupasan atau tafsiran untuk *review*, sedangkan untuk kata *online* mempunyai makna yang berarti terkoneksi atau dikontrol pada sebuah komputer.

## 2.4. Online Customer Rating

Rating adalah penilaian secara emosional dan psikologi yang dilakukan pada pengalaman yang didapat saat menggunakan suatu produk, layanan atau jasa. Biasanya rating disimbolkan dengan gambar bintang untuk mengekspresikan pendapat yang dirasakan oleh pelanggan. Keputusan Pelanggan akan untuk membeli atau memakai suatu produk atau jasa tergantung seberapa baik rating yang diberikan oleh pelanggan lain terhadap produk atau jasa tersebut sehingga sangat penting untuk menjaga rating tetap baik di mata pelanggan.

## 3.5. Google Play

Google Play ialah layanan konten digital Google, terhitung toko online untuk produk seperti musik/lagu, komik, aplikasi, games, maupun pemutar media berbasis cloud. layanan ini bisa diakses lewat web, aplikasi Android (Play Store), serta Google Televisi.

Google Play memiliki fitur untuk melakukan analisis topik pada ulasan yang relevan dengan aplikasi berdasarkan kutipan sering muncul dalam ulasan serta terdapat juga beberapa metrik pengukuran yang dikategorikan kedalam aspek seperti desain, privasi, profil, penggunaan sumber daya, kecepatan, stabilitas, update, uninstall, dan kegunaan dari aplikasi tersebut. Ini sangat sesuai buat analisis sentimen dari review pengguna pada aplikasi di Google Play, menolong pengguna buat menginstal aplikasi serta membantu developer menganalisis aplikasi.

## 2.6. Komponen dari *Integrated Service Management*

Komponen-komponen dari *integrated service management* dibahas saat menerapkan strategi untuk memasarkan barang. *Integrated service management* adalah perencanaan dan pelaksanaan terkoordinasi dari kegiatan pemasaran, operasi, dan sumber daya manusia yang penting untuk kesuksesan perusahaan jasa [10].

Komponen ini dikelompokkan atas 8 komponen, yaitu: *product, distribution, promotion, price, people, physical evidence, process* dan *productivity & quality*.

1. *Process* merupakan suatu metode penyajian produk dan layanan yang baik kepada pelanggan melalui pengoperasian atau serangkaian tindakan yang diperlukan.
2. *Product* merupakan semua hal memberikan manfaat, memenuhi kebutuhan, dan dapat memuaskan konsumen dari komponen-komponen kinerja layanan yang menciptakan nilai bagi pelanggan.
3. *People* adalah semua pihak yang terlibat dalam alur memproduksi produk dan layanan.
4. *Distribution* adalah keputusan melibatkan tentang tempat dan waktu pengiriman dan mungkin melibatkan fisik atau elektronik saluran distribusi (atau keduanya), tergantung pada sifat layanan yang disediakan apakah sudah memenuhi kebutuhan pelanggan.
5. *Price* merupakan suatu hal yang pelanggan korbankan berupa pengeluaran uang, waktu, dan usaha yang dalam membeli dan mengkonsumsi produk dan layanan yang diberikan perusahaan.
6. *Promotion* merupakan sebuah cara komunikasi dan perancangan layanan yang diberikan untuk pembentukan persepsi pelanggan yang dikehendaki oleh perusahaan.
7. *Physical Evidence* merupakan perangkat-perangkat nyata atau terlihat secara fisik yang diperlukan dalam menyajikan kualitas produk dan layanan.
8. *Productivity & Quality* adalah sejauh mana efisiensi layanan ditransformasikan ke dalam hasil-hasil layanan yang dapat menambah nilai bagi pelanggan serta derajat suatu layanan tersebut dapat memenuhi harapan pelanggan.

### 2.7. Text Classification

Kita semua memeriksa *email* setiap hari, mungkin beberapa kali. Fitur yang berguna dari sebagian besar penyedia layanan *email* adalah kemampuan untuk secara otomatis memisahkan *email spam* dari *email* biasa. Ini adalah kasus penggunaan tugas NLP yang populer dikenal sebagai *Text Classification*. Natural Language Processing (NLP) Ini adalah cabang dari kecerdasan buatan (artificial intelligence) yang berfokus pada pemrosesan bahasa alami. Bahasa alami adalah bahasa yang sering digunakan manusia dalam komunikasi timbal balik. Bahasa yang diterima oleh komputer perlu diproses dan dipahami terlebih dahulu agar komputer dapat memahami maksud pengguna dengan benar. Natural Language Processing digunakan untuk menganalisis teks tidak terstruktur karena hal tersebut lebih sulit untuk dikonsumsi langsung oleh manusia, sehingga diperlukan bantuan mesin untuk memahami struktur teks yang kompleks. *Text Classification* sendiri adalah tugas menetapkan satu atau lebih banyak kategori ke bagian teks tertentu dari kumpulan kemungkinan kategori yang lebih besar [14]. *Text classification* kadang-kadang juga disebut sebagai *topic classification*, *text categorization*, atau *document categorization*.

*Supervised classification*, termasuk *text classification*, dapat lebih lanjut dibedakan menjadi tiga jenis berdasarkan jumlah kategori yang terlibat: *binary classification*, *multiclass classification*, dan *multilabel classification*. Jika jumlah kelas adalah dua, klasifikasi itu disebut *binary classification*. Jika jumlah kelas lebih dari dua, itu disebut sebagai *multiclass classification*. Dengan demikian, mengklasifikasikan *email* sebagai *spam* atau bukan *spam* adalah contoh pengaturan *binary classification*. Mengklasifikasikan sentimen *review* pelanggan sebagai negatif, netral, atau positif adalah contoh *multiclass classification*. dalam *multilabel classification*, sebuah dokumen dapat memiliki satu atau lebih label/kelas yang melekat padanya. Misalnya, artikel berita tentang sepak bola pertandingan mungkin termasuk dalam lebih dari satu kategori, seperti "olahraga" dan "sepak bola", secara serentak.

### 2.8. Text Preprocessing

Untuk melakukan proses natural language processing, terlebih dahulu dokumen atau teks diolah dan dipersiapkan. Proses pengolahan data ini disebut juga dengan nama *text preprocessing*. *Text preprocessing* dilakukan dengan cara mengubah data teks mentah menjadi bersih dari noise dan membuat bentuknya lebih terstruktur. Berikut adalah beberapa tahap pada *text preprocessing*:

#### 1. Spelling Normalization

*Spelling Normalization* adalah tahap untuk mengubah kata-kata yang tidak baku menjadi bentuk baku. Hal tersebut dilakukan untuk mengurangi jumlah perhitungan fitur kata. Fitur kata akan bertambah jika terdapat kata tidak baku atau salah pengejaan karena kata tersebut meski memiliki arti yang sama tapi mempunyai fitur yang berbeda jika diubah kedalam bentuk matriks.

#### 2. Case Folding

*Case folding* adalah tahap untuk menyetarakan bentuk dalam suatu teks atau dokumen. Hal tersebut dilakukan karena pada umumnya teks tidak konsisten dalam penggunaan huruf kapital sehingga *case folding* memiliki peran untuk mengkonversi teks ke dalam format yang sama.

#### 3. Tokenizing

Untuk membuat fitur dari kata-kata terlebih dahulu teks diubah menjadi potongan kata-kata dengan *Tokenizing* sehingga bisa didapat kata yang lebih memiliki makna. *Tokenizing* dilakukan dengan memotong kata berdasarkan spasi dari suatu teks yang pada akhirnya akan menghasilkan potongan kata-kata yang terpisah-pisah dari teksnya.

#### 4. Filtering

Tahap *filtering* adalah tahap mengambil kata-kata penting dari suatu teks dan membuang kata seperti "di", "dan", "yang", "dari" yang termasuk dalam golongan kata *Stopword* yang tidak memiliki makna dari suatu dokumen sehingga dapat dihapus.

### 2.9. Wordcloud

*Wordcloud* adalah salah satu cara yang digunakan untuk menggambarkan kata-kata populer kedalam bentuk menyerupai awan [15]. *Wordcloud* merupakan metode yang dapat menjelaskan pertanyaan penelitian secara singkat dan mudah serta dapat melakukan analisis yang komprehensif. Kata yang frekuensi munculnya paling banyak di dalam data teks akan memiliki bentuk yang paling besar, begitu pula sebaliknya. Contoh bentuk *wordcloud* dapat dilihat pada Gambar 1.



Gambar 1: Contoh Wordcloud

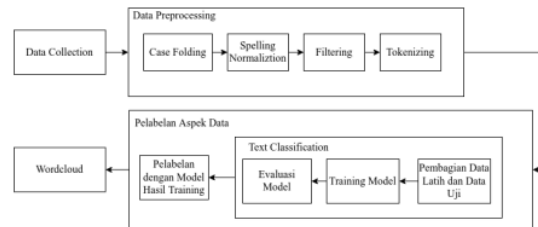
### 3. METODE PENELITIAN

#### 3.1. Data

Penelitian ini menggunakan data *review* aplikasi Shopee yang ada di website Google Play untuk regional daerah Indonesia yang berbahasa Indonesia. Rentang waktu data diambil dimulai 01 Desember 2020 sampai 01 Februari 2021. Data *review* yang akan diambil untuk penelitian ini sebanyak 1.000 *review*. Sumber data yang dibutuhkan dalam penelitian ini diperoleh melalui teknik *scraping* pada halaman *review* aplikasi Shopee di website Google Play.

#### 3.2. Metode Penelitian

Metode yang digunakan dalam penelitian ini terdiri dari *Data Collection*, *Data Preprocessing*, Pelabelan Aspek dan *Wordcloud* seperti yang ditunjukkan pada Gambar 2.



Gambar 2 Metode Penelitian

##### 1) Data Collection

Data *review* diperoleh dari situs *website* Google Play. Data yang diambil hanya data *review* pada aplikasi Shopee untuk regional Indonesia. Proses *scraping* dilakukan dengan menggunakan *library* dari bahasa pemrograman *python*. *Review-review* hasil *scraping* akan diseleksi Atribut yang akan diambil diantaranya adalah *score*, *text* dan *date* sehingga memudahkan proses pengolahan data.

##### 2) Data Preprocessing

Data ulasan yang diperoleh belum sepenuhnya siap digunakan untuk diproses secara langsung karena data masih tidak terstruktur dengan baik dan terdapat banyak *noise*. Data masih memuat angka, tanda baca, *emoticon*, serta kata-kata lain yang kurang bermakna untuk dijadikan fitur. Maka dari itu, perlu dilakukan preprocessing yang bertujuan untuk menyeragamkan bentuk kata, menghilangkan karakter-karakter selain huruf, dan mengurangi volume kosakata sehingga data akan lebih terstruktur. Beberapa tahap yang akan dilakukan diantaranya adalah *spelling normalization*, *case folding*, *tokenizing*, dan *filtering*.

- a) *Case Folding* dilakukan dengan cara melakukan penyamaan case dalam sebuah *review*. Hal tersebut dilakukan dengan cara mengonversi keseluruhan teks dalam *review* menjadi suatu bentuk standar.

- b) *Spelling Normalization* dilakukan untuk memperbaiki atau substitusi kata-kata yang salah eja atau disingkat dalam bentuk tertentu.
- c) *Filtering* dilakukan untuk menghilangkan kata-kata *Stopword* yang tidak deskriptif dan bukan merupakan kata penting dari suatu dokumen sehingga dapat dibuang.
- d) *Tokenization* merupakan pemotongan kata berdasarkan tiap kata yang menyusunnya menjadi potongan tunggal. Kata dalam dokumen yang dimaksud adalah kata yang dipisah oleh spasi. Sehingga hasil dari proses ini merupakan kata tunggal. Proses tokenisasi diawali dengan menghilangkan delimiter-delimiter yaitu simbol dan tanda baca yang ada pada teks tersebut seperti @, \$, &, tanda titik (.), koma (,) tanda tanya (?), tanda seru (!).

### 3) Pelabelan Aspek Data

Pelabelan aspek pada data dilakukan dengan cara melabelkan *review* ke dalam 8 komponen *Integrated Service Management* yaitu *process, product, promotion, people, distribution, price, physical evidence, dan productivity and quality* secara manual dan divalidasi oleh Muizzuddin, Dr., S.E., M.M. selaku dosen Program Studi Manajemen Fakultas Ekonomi Universitas Sriwijaya. Percobaan dimulai menggunakan 429 data *review* yang perlahan-lahan akan ditambah tergantung dari hasil *testing* dari model yang di dapat sudah menunjukkan tingkat akurasi yang baik. Model yang dihasilkan nantinya akan digunakan untuk memprediksi label aspek dari sisa data yang belum memiliki label.

#### a) *Text Classification*

Klasifikasi teks adalah contoh khusus dari masalah klasifikasi, di mana titik data *input* adalah teks dan tujuannya adalah untuk mengkategorikan potongan teks ke dalam satu atau lebih kelas. Berikut ini beberapa tahap untuk melakukan *Text classification* :

##### 1. Pembagian Data Latih dan Data Uji

Data perlu dibagi menjadi 2 bagian yaitu data latih dan data uji. Hal tersebut dilakukan untuk menguji model yang di *training* dengan data uji untuk mengetahui seberapa baik model mengklasifikasi data. Oleh karena itu dalam penelitian ini data akan dibagi dengan pembagian rasio 80% data latih dan data uji 20% sehingga diharapkan dapat membentuk model yang bisa mengklasifikasi data yang belum pernah terlihat pola nya oleh model yang terbentuk oleh data latih.

Sebelum dilakukannya pembagian data, perlu dilakukan transformasi data yang berupa teks menjadi vektor agar algoritma *machine learning* dapat membaca data. *Word embedding* adalah salah satu teknik Mengkonversi kata menjadi bentuk vektor. Dalam penelitian ini *tool word embedding* yang digunakan adalah *Word2vec*. Dasar untuk menghasilkan nilai vektor pada *Word2vec* adalah jaringan saraf. Ada dua arsitektur neural network untuk melatih model *Word2vec*, yaitu *continuous bag of words (CBOW)* dan *skip-gram*. Penelitian ini menggunakan arsitektur *skip-gram* dengan ukuran vektor 200 dimensi. Nilai vektor yang dihasilkan oleh *Word2vec* dapat mewakili arti sebuah kata. Sebelum *Word2vec* digunakan untuk merepresentasikan nilai vektor terlebih dahulu dilakukan pembuatan model distribusi kata selama pelatihan. Karena pembuatan model *Word2Vec* memerlukan banyak teks berbahasa Indonesia oleh sebab itu maka Pembuatan model akan menggunakan *pre-trained Word2Vec* dari data *corpus* Wikipedia Indonesia.

##### 2. Training Model

Proses *training* dimulai dengan mengambil data *review* yang telah melalui tahap *word embedding* agar teks dapat dibaca oleh algoritma *machine learning*, lalu model di *train* menggunakan data latih dengan algoritma *Support Vector Machine (SVM)* diuji dengan beberapa kernel diantaranya adalah *sigmoid, linear dan RBF (radial basis function)*. Label aspek yang digunakan berdasarkan 8 komponen dari *Integrated Service Management*. Masing-masing *review* dapat memiliki satu aspek sehingga proses klasifikasi dalam hal ini penelitian dapat dilihat sebagai masalah klasifikasi *multi-class*.

##### 3. Evaluasi Model

Untuk mendapatkan model yang baik perlu dilakukan pengujian terhadap model tersebut. Pengujian dilakukan dengan menggunakan *K-fold Cross Validation* dengan menggunakan  $k = 5$ . Data akan dibagi menjadi 5 partisi yang jumlahnya sama, terdiri dari 4 data latih dan 1 data uji. Lebih lanjut perhitungan untuk mengukur kinerja klasifikasi dengan *Confusion Matrix* dan menampilkan hasil tes dengan metric nilai *accuracy, precision, f1-score* dan *recall* pada model untuk melihat apakah model sudah melakukan klasifikasi dengan baik.

#### b) Pelabelan dengan Model hasil Training

Langka selanjutnya adalah melabelkan sisa data *review* yang belum memiliki label aspek dengan model yang terbentuk dari hasil *training* data *review* yang di label secara manual. Hal tersebut dilakukan dengan memprediksi label menggunakan model tersebut.

### 4) Pembuatan *Wordcloud*

Penggunaan *wordcloud* bertujuan untuk mendapatkan informasi tentang kata-kata/topik yang paling sering muncul di dalam *review*. Sehingga memberikan gambaran yang lebih jelas tentang topik/kata kata yang sering digunakan oleh customer dalam memberikan *review*. Semakin besar ukuran kata pada *wordcloud* menggambarkan semakin tinggi pula frekuensi kata tersebut, artinya semakin sering customer menggunakan kata tersebut sebagai topik pembicaraan atau penilaian dalam *review*.

#### 4. HASIL DAN PEMBAHASAN

##### 4.1. Data Collection

Pada Tabel 1 menunjukkan data *review* aplikasi Shopee pada website Google Play yang telah dikumpulkan menggunakan cara *scraping* dengan total data yang berjumlah sebanyak 1000 *review* yang telah terkumpul dari tanggal 01 Desember 2020 sampai tanggal 01 Februari 2021.

**Tabel 1** Data Review Hasil Scraping

<i>Text</i>	<i>Date</i>
Tolong lapak ini di perhatikan seenaknya melakukan penipuan produk. FR_SIDOARJO shopee.co.id/fr_sidoarjo Jangan sampe toko2 penipu seperti ini ada di shopee dan jangan sampe saudara2 yang lain bernasib sama kaya saya.	12-12-2020
Sekarang Shopee makin pelit voucher, biasanya voucher tiap hari ada untuk di klaim tapi sekarang udah gak ada, dan setiap ada jambore voucher sudah di klaim tapi ketika voucher mau di pake waktu nya gak ada 1 jam cuma 2-3 menit aja udh gak bisa di pake karena sudah hrs, ongkir pun jadi naikk.. gak ngerti lagi sama Shopee kenapa jadi gini	11-12-2020
Aplikasinya udah bagus sih, tapi sedih banget deh ekspedisi pengirimannya di batasi tidak sebanyak dulu, sekarang cuma ada 2 sampai 3 pengiriman kurir dulu ada sampe 5. Jadi bisa cari yg lebih murah. Sekarang nggak bisa, sedih banget. Semoga bisa kembali normal lagi shopee. Cashback koin shopee aku min, beneran kecewa gak dapet sampai sekarang	01-01-2021
belanja disini gak bakal nyesel, produk nya bagus dengan harga yg pas dikantong, selalu dapat voucher gratis terus loh disini, untuk kalian yang doyan belanja, cussss ini aplikasi yg saya rekomendasikan (SHOPEE)	24-12-2020

##### 4.2. Training Model

*Machine learning* menggunakan data latih untuk membentuk suatu *model classifier*. Model yang terbentuk nantinya akan digunakan untuk memprediksi kelas aspek pada data baru yang belum pernah ada. Dengan bertambahnya jumlah data yang digunakan untuk pembentukan model akan berpengaruh terhadap kemampuan model untuk melihat pola pada data. Untuk menguji sebegas apa model yang sudah terbuat maka diperlukan data uji yang digunakan untuk menilai tingkat akurasi model. Pada penelitian ini akan menggunakan data latih dan data uji dengan komposisi jumlah 80% : 20% seperti yang terlihat pada Tabel 2 dibawah ini.

**Tabel 2** Jumlah Komposisi Data Latih dan Data Uji Berdasarkan Aspek

Klasifikasi	Jumlah	Data Latih	Data Uji
<i>People</i>	49	39	10
<i>Price</i>	56	45	11
<i>Process</i>	56	45	11
<i>Promotion</i>	49	39	10
<i>Physical Evidence</i>	49	39	10
<i>Distribution</i>	59	47	12
<i>Product</i>	58	46	12
<i>Productivity &amp; Quality</i>	53	43	10
Total	429	343	86

Data latih berdasarkan aspek digunakan oleh algoritma SVM dalam mempelajari pola data yang ada pada masing-masing kelas aspek untuk membentuk model. Model terlebih dahulu diuji untuk mengetahui tingkat *accuracy* yang didapat dari model sudah baik atau tidak dalam mengklasifikasikan data uji. Beberapa *kernel* dicoba untuk dibandingkan hasilnya, diantaranya adalah Radial Basis Function (RBF), Sigmoid, dan Linear pada SVM menggunakan metode *Word2Vec* untuk melakukan klasifikasi dengan menggunakan *Cross Validation* dengan 5-fold demi mendapatkan hasil *accuracy* yang terbaik. Berikut ini perbandingan *accuracy* yang didapat antara fungsi



RBF, Sigmoid dan Linear pada SVM. Berikut perbandingan *accuracy* dari jenis *kernel* SVM yang dapat dilihat pada Tabel 3.

**Tabel 3** Perbandingan Jenis Kernel Pada Klasifikasi SVM

Jenis <i>Kernel</i>	<i>Accuracy</i>
Sigmoid	0,80
RBF	0,77
Linear	0,78

Pada Tabel 3 diketahui dari data hasil beberapa *kernel* yang dilakukan percobaan memakai 5-fold *Cross validation* dengan 5 kali percobaan, kombinasi *kernel* sigmoid dengan *word2vec* memiliki tingkat *accuracy* yang lebih tinggi dari pada *kernel* lain, oleh sebab itu *kernel* sigmoid dengan *word2vec* akan digunakan untuk melakukan klasifikasi pada penelitian ini.

#### 4.3. Evaluasi Model

Evaluasi model perlu dilakukan untuk mengetahui kemampuan model dalam melakukan klasifikasi. Pada penelitian akan menggunakan *Confusion matrix* untuk melakukan kalkulasi pada *precision*, *recall* dan *f1-score*.

**Tabel 4** Rangkuman Nilai *Precision*, *Recall*, *F1-Score* Klasifikasi Aspek

Klasifikasi	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>
<i>People</i>	0,67	1,00	0,80
<i>Physical Evidence</i>	1,00	0,60	0,75
<i>Distribution</i>	1,00	1,00	1,00
<i>Price</i>	0,80	0,73	0,76
<i>Process</i>	0,79	1,00	0,88
<i>Product</i>	1,00	0,92	0,96
<i>Productivity &amp; Quality</i>	0,86	0,60	0,71
<i>Promotion</i>	0,73	0,80	0,76

Pada Tabel 4 dapat diketahui bahwa klasifikasi pada aspek *productivity & quality* memiliki nilai *f1-score* paling kecil diantara yang lain nya dan klasifikasi aspek *distribution* memiliki nilai paling besar dari pada yang lain. Hal tersebut berarti klasifikasi aspek *productivity & quality* kurang baik dibandingkan dengan yang lain nya sedangkan klasifikasi aspek *distribution* lebih baik dalam mengklasifikasi data dari pada aspek-aspek yang lain.

#### 4.4. Pelabelan Dengan Model Hasil *Training*

Langka selanjutnya adalah melabelkan sisa data *review* yang belum memiliki label aspek dengan model yang terbentuk dari hasil *training* data *review* yang di label secara manual. Hal tersebut dilakukan dengan memprediksi label menggunakan model tersebut. Berikut jumlah *review* berdasarkan aspek pada Tabel 5.

**Tabel 5** Jumlah Data *Review* Berdasarkan Aspek

Aspek	Jumlah
<i>Process</i>	56
<i>Product</i>	91
<i>Promotion</i>	247
<i>Distribution</i>	149
<i>People</i>	49
<i>Physical Evidence</i>	49
<i>Price</i>	157
<i>Productivity &amp; Quality</i>	202
<b>Total</b>	1000

#### 4.5. Pembuatan *Wordcloud*

*Wordcloud* kerap digunakan buat menyoroti sebutan terkenal ataupun trend bersumber pada frekuensi kemunculan kata yang muncul pada kumpulan *review*. Semakin besar dimensi kata pada *wordcloud* menggambarkan semakin

besar pula frekuensi kata tersebut muncul dalam topik *review*. Berikut ini beberapa *wordcloud* yang terbentuk berdasarkan aspek dari data *review* :

a) *People*



**Gambar 2** Wordcloud Review Aspek *People*

Dari **Gambar 2** terdapat Kata-kata “customer”, “service”, “selesai”, “chat”, “komplain”, dan “solusi” memberikan informasi bahwa solusi yang diberikan dari pelayanan *chat customer service* dari Shopee tidak menyelesaikan masalah *komplain* dari *customer* dan terkadang *customer service* bahkan mematikan *chat* sebelum masalah yang dimiliki oleh pelanggan terselesaikan.

b) *Price*



**Gambar 3** Wordcloud Review Aspek *Price*

Dari **Gambar 3** terdapat Kata-kata “biaya”, “admin”, “penanganan”, “harga”, “ongkos”, “kirim”, dan “mahal” memberikan informasi bahwa *customer* Shopee protes dengan adanya biaya admin dan penanganan serta mahal nya harga ongkos kirim yang lumayan membebani *customer*.

c) *Promotion*



**Gambar 4** Wordcloud Review Aspek *Promotion*

Dari **Gambar 4** terdapat Kata-kata “voucher”, “gratis”, “memakai”, “belanja”, “ongkos”, dan “kirim” memberikan informasi bahwa banyak *customer* yang berbelanja di Shopee tidak bisa menggunakan dan meminta ditamhkannya voucher promo terutama voucher untuk gratis ongkos kirim.

d) *Distribution*



**Gambar 5** Wordcloud Review Aspek Distribution

Dari **Gambar 5** terdapat Kata-kata “ekspedisi”, ”jasa”, “pengiriman”, “pilih”, dan “memilih” memberikan informasi bahwa *customer* tidak suka dengan adanya pemilihan ekspedisi berdasarkan pilihan pihak Shopee dan menginginkan adanya pemilihan jasa ekspedisi sesuai keinginan *customer*.

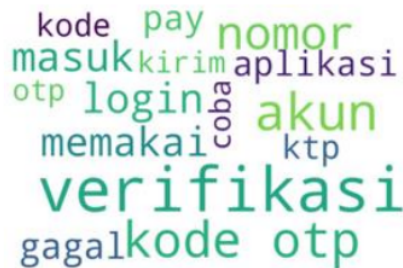
e) *Product*



**Gambar 6** Wordcloud Review Aspek Product

Dari **Gambar 6** terdapat Kata-kata “negeri”, ”barang”, “belanja”, “pilih”, dan ’produk’ memberikan informasi bahwa *customer* di Shopee tidak menginginkan kebijakan penghapusan berbelanja produk/barang dari luar negeri.

f) *Process*



**Gambar 7** Wordcloud Review Aspek Process

Dari **Gambar 7** terdapat Kata-kata “verifikasi”, ”kode”, “otp”, “nomor”, “ktp”, dan “akun” memberikan informasi bahwa *customer* Shopee memiliki masalah saat ingin melakukan verifikasi dikarenakan aplikasi gagal untuk mendeteksi foto ktp yang dikirim serta *customer* yang mempunyai masalah pada nomor yang terdaftar pada akun tidak bisa mendapatkan kode OTP untuk verifikasi akun dikarenakan kode OTP hanya dikirim melalui nomor tersebut.

g) *Productivity & Quality*



**Gambar 8** Wordcloud Review Aspek Productivity & Quality

Dari **Gambar 8** terdapat Kata-kata “barang”, “belanja”, “penjual”, “tipu”, dan “penipuan” memberikan informasi bahwa *customer* Shopee banyak terkena aksi penipuan oleh oknum yang mengatasnamakan pihak Shopee serta ada beberapa penjual yang melakukan penipuan kepada *customer*.

h) *Physical Evidence*



**Gambar 9** Wordcloud Review Aspek Physical Evidence

Dari **Gambar 9** terdapat Kata-kata “user”, “interface”, “dark”, “mode”, “aplikasi”, dan “fitur” memberikan informasi bahwa *user interface* aplikasi Shopee perlu di tambah fitur *dark mode* serta ditingkatkan lagi kemudahannya bagi *customer* untuk menggunakannya.

Rangkuman masalah-masalah yang berhasil ditemukan dari review aplikasi Shopee berdasarkan ke-8 aspek review bisa dilihat pada Tabel 6.

**Tabel 6** Rangkuman Masalah *wordcloud* aspek Review Pada E-commerce Shopee

Aspek	Permasalahan
<i>Process</i>	Verifikasi gagal untuk nomor telepon yang terblokir Verifikasi Shopee Pay yang gagal dalam mendeteksi foto ktp
<i>People</i>	Pelayanan <i>chat customer service</i> kurang membantu menyelesaikan masalah komplain <i>customer customer service</i> terkadang mematikan chat sebelum masalah komplain terselesaikan
<i>Price</i>	Ongkos kirim yang mahal Adanya biaya admin dan penanganan yang lumayan membebani <i>customer</i>
<i>Physical Evidence</i>	Tidak adanya fitur mengubah tampilan menjadi <i>dark mode</i> <i>User Interface</i> aplikasi Shopee yang tidak mudah untuk digunakan
<i>Promotion</i>	diperbanyak nya <i>Voucher</i> gratis ongkos kirim Banyak <i>Voucher</i> promo yang tidak bisa digunakan

<i>Distribution</i>	Dihilangkan fitur untuk memilih jasa ekspedisi
<i>Productivity &amp; Quality</i>	Adanya akun penjual penipu Adanya oknum penipuan mengatasnamakan Shopee memberikan promosi
<i>Product</i>	Barang dari luar negeri diblokir

## 5. SARAN DAN KESIMPULAN

### 5.1. Kesimpulan

Penelitian ini telah melakukan klasifikasi ulasan aplikasi e-commerce Shopee ke dalam beberapa aspek peningkatan layanan e-commerce. Dari tiga jenis kernel Support Vector Machine (SVM) yaitu Sigmoid, RBF dan Linear, didapatkan bahwa kernel sigmoid memiliki tingkat akurasi tertinggi yaitu 80%, yang digunakan untuk mengkategorikan ulasan sehingga dapat diketahui apa yang dibicarakan pelanggan untuk setiap aspek layanan yaitu *distribution*, *promotion*, *people*, *price*, *physical evidence*, *product*, *process*, dan *productivity & quality*. Beberapa masalah yang berhasil ditemukan diantaranya yaitu proses verifikasi, *customer service*, pembelian barang, jasa ekspedisi, promosi, *user interface*, pengiriman, *voucher*, dan ongkos kirim.

### 5.2. Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa poin yang dapat diberikan sebagai saran untuk penelitian selanjutnya diantaranya adalah:

- Untuk pelabelan aspek yang diterapkan pada penelitian ini terbatas hanya pada tingkat dokumen sehingga belum dapat mengidentifikasi makna teks dari tingkat antar kata-kata, sebaiknya metode *part of speech* (POS) *tagging* digunakan untuk penelitian selanjutnya supaya dapat menandai kata-kata dalam teks kedalam kata benda, kata sifat, dan kata keterangan sehingga dapat mendeteksi dan mengekstrak kata-kata secara lebih detail dan spesifik.
- Data *review* Google Play yang digunakan untuk penelitian ini diambil mulai dari tanggal 01 Desember 2020 sampai 01 Februari 2021 bertepatan dengan *event* 12.12 Shopee *Birthday Sale*, sehingga akan lebih baik jika jumlah data yang diambil memiliki jumlah yang lebih banyak dengan rentang waktu yang lebih panjang.

### REFERENSI

- [1] World Retail Congress, (2019), Global Ecommerce Market Ranking 2019, Retrieved November 16, 2020, from [https://www.worldretailcongress.com/\\_media/Global\\_ecommerce\\_Market\\_Ranking\\_2019\\_001.pdf](https://www.worldretailcongress.com/_media/Global_ecommerce_Market_Ranking_2019_001.pdf).
- [2] Iprice, (2020), Peta E-commerce Indonesia 2020, Retrieved November 15, 2020, from <https://iprice.co.id/insights/mapofecommerce/>.
- [3] Shopee, (2020), Tentang Shopee, Retrieved November 15, 2020, from <https://careers.shopee.co.id/about/>
- [4] Google Play. (2020). Tentang Google Play, Retrieved November 16, 2020, from [https://play.google.com/intl/id\\_id/about](https://play.google.com/intl/id_id/about)
- [5] Google Support, (2021), Tentang Panduan Analisis Ulasan Aplikasi Bagi Developer Android, Retrieved Juni 25, 2021, from <https://support.google.com/googleplay/android-developer/answer/138230?hl=id>
- [6] Ekawati, D., & Khodra, M. L. (2017). Aspect-based sentiment analysis for Indonesian restaurant reviews. Proceedings - 2017 International Conference on Advanced Informatics: Concepts, Theory and Applications, ICAICTA 2017.
- [7] Amarouche, K., Benbrahim, H., & Kassou, I. (2015). Product Opinion Mining for Competitive Intelligence. Procedia Computer Science, 73(Awict), 358–365.
- [8] Dadgar, S. M. H., Araghi, M. S., & Farahani, M. M. (2016). A novel text mining approach based on TF-IDF and support vector machine for news classification. Proceedings of 2nd IEEE International Conference on Engineering and Technology, ICETECH 2016, March, 112–116.
- [9] Sari, A. E., Widowati, S., & Lhaksana, K. M. (2019). Klasifikasi Ulasan Pengguna Aplikasi Mandiri Online di Google Play Store dengan Menggunakan Metode Information Gain dan Naive Bayes Classifier. E-Proceeding of Engineering, 6(2), 9143–9157.
- [10] Thompson, C. (2000). Book Review: Principles of Service Marketing and Management. Journal of Vacation Marketing, 6(3), 281–283.
- [11] Kwok, L., Tang, Y., & Yu, B. (2020). The 7 Ps marketing mix of home-sharing services: Mining travelers' online reviews on Airbnb. International Journal of Hospitality Management, 90(August 2019), 102616.
- [12] Indrajit, R., E. (2011). Manajemen Sistem Informasi dan Teknologi Informasi. Jakarta, DKI Jakarta, Indonesia: Elex Media Komputindo.
- [13] Kotler, P., & Armstrong, G. (1996). Dasar-Dasar Pemasaran. Jakarta, DKI Jakarta, Indonesia: Intermedia.

- [14] Lane, H., Howard, C., & Hapke, H. M. (2019). Natural Language Processing in Action(Understanding,analyzing, and generating text with python).
- [15] Arkhamsiagustinah. (2015). Perbandingan Metode Second-Order Fuzzy Time Series Dari HSU dan CHEN Dalam Peramalan Produk Domestik Regional Bruto (PDRB). Universitas Islam Indonesia, Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam. Yogyakarta: Program Studi Statistika.

# ANALISIS MASALAH PADA DATA REVIEW APLIKASI TERHADAP LAYANAN E-COMMERCE MENGGUNAKAN METODE TEXT CLASSIFICATION

## ORIGINALITY REPORT

12%

SIMILARITY INDEX

12%

INTERNET SOURCES

0%

PUBLICATIONS

0%

STUDENT PAPERS

## PRIMARY SOURCES

1

[garuda.kemdikbud.go.id](http://garuda.kemdikbud.go.id)

Internet Source

5%

2

[repository.itelkom-pwt.ac.id](http://repository.itelkom-pwt.ac.id)

Internet Source

4%

3

[dokumen.tips](http://dokumen.tips)

Internet Source

3%

Exclude quotes  On

Exclude matches  < 3%

Exclude bibliography  On