

**TUGAS AKHIR**

**ANALISIS SENTIMEN BERBASIS ASPEK PADA PERILAKU  
PENGGUNA *FINTECH* MENGGUNAKAN ALGORITMA  
LDA DAN *SUPPORT VECTOR MACHINE***

**Oleh:**

**Satria Adhy Nayoga**

**1202194226**



**PROGRAM STUDI STRATA 1 SISTEM INFORMASI**

**FAKULTAS REKAYASA INDUSTRI**

**UNIVERSITAS TELKOM**

**2023**

## ABSTRAK

Pesatnya pertumbuhan teknologi, terutama pada layanan *Financial Technology* (*fintech*), khususnya *e-wallet*, telah membawa tantangan dan stressor baru bagi pengguna di Indonesia. Pada tahun 2021, OVO menjadi salah satu dompet digital yang paling populer dan banyak digunakan. Namun, pada tahun 2022, OVO mengalami penurunan *rating* yang diduga dipengaruhi oleh adanya *technostress*. Dampak dari *technostress* ini berpengaruh pada perilaku pengguna. Penelitian ini bertujuan untuk menganalisis *technostress* dalam konteks *fintech e-wallet* di OVO menggunakan *Aspect-Based Sentiment Analysis* dan algoritma *Support Vector Machine* (SVM). Pada penelitian ini, *Aspect-Based Sentiment Analysis* digunakan untuk mendapatkan informasi sentimen dan aspek dari aplikasi OVO. Ulasan pengguna dari Google Play Store dikumpulkan dan dianalisis menggunakan metode *scraping* untuk mengidentifikasi aspek dan sentimen utama yang diungkapkan oleh pengguna. Dalam penentuan aspek, peneliti menggunakan pemodelan topik LDA untuk mendapatkan hasil sebanyak 4 topik, yaitu fitur, akses, pelayanan, dan keamanan. Algoritma SVM kemudian digunakan untuk mengklasifikasikan sentimen pada masing-masing aspek. Dengan menggunakan pembagian *dataset* optimal 75:25, penelitian ini mencapai tingkat akurasi yang tinggi untuk setiap aspek. Hasilnya menunjukkan bahwa akurasi klasifikasi mencapai 97,08% untuk aspek Fitur, 94,81% untuk aspek Akses, 91,34% untuk aspek Pelayanan, dan 96,36% untuk aspek Keamanan. Rata-rata akurasi keseluruhan juga mencapai tingkat yang memuaskan. Hasil penelitian menunjukkan bahwa aspek *technostress* yang paling banyak diperbincangkan dalam penggunaan *e-wallet* adalah terjadinya *error* pada fitur aplikasi, yang dominan membawa sentimen negatif. Selain itu, aspek lain seperti akses dan kegagalan layanan juga memunculkan sentimen negatif. Temuan ini menekankan pentingnya perbaikan terus-menerus dalam sistem dan fitur aplikasi *e-wallet* untuk meringankan *technostress* dan meningkatkan kepuasan pengguna.

Kata Kunci — *Fintech, Dompet Digital, Technostress, Aspect-Based Sentiment Analysis, LDA, Support Vector Machine*

## **ABSTRACT**

*The rapid growth of technology, especially in financial technology (fintech) services, especially e-wallets, has brought new challenges and stressors for users in Indonesia. In 2021, OVO became one of the most popular and widely used digital wallets. However, in 2022, OVO experienced a decline in rating which was thought to be influenced by technostress. The impact of technostress affects user behavior. This research aims to analyze technostress in the context of fintech e-wallet at OVO using Aspect-Based Sentiment Analysis and support vector machine (SVM) algorithm. In this research, Aspect-Based Sentiment Analysis is used to obtain sentiment and aspect information from the OVO application. User reviews from Google Play Store were collected and analyzed using scraping method to identify key aspects and sentiments expressed by users. In determining the aspects, the researcher used LDA topic modeling to get the results of 4 topics, namely features, access, service, and security. The SVM algorithm was then used to classify the sentiment on each aspect. By using an optimal dataset split of 75:25, this research achieved a high level of accuracy for each aspect. The results show that the classification accuracy reached 97.58% for Feature aspect, 95.23% for Access aspect, 92.52% for Service aspect, and 96.72% for Security aspect. The overall average accuracy also reached a satisfactory level. The results showed that the most discussed aspect of technostress in using e-wallets is the occurrence of errors in application features, which dominantly brings negative sentiments. In addition, other aspects such as access and service failure also generate negative sentiments. The findings emphasize the importance of continuous improvement in e-wallet application systems and features to alleviate technostress and increase user satisfaction.*

**Keywords — Fintech, E-Wallet, Technostress, Aspect-Based Sentiment Analysis, LDA, Support Vector Machine**

**LEMBAR PENGESAHAN**

Tugas Akhir dengan judul :

**ANALISIS SENTIMEN BERBASIS ASPEK PADA PERILAKU  
PENGGUNA *FINTECH* MENGGUNAKAN ALGORITMA LDA DAN  
*SUPPORT VECTOR MACHINE***

Telah disetujui dan disahkan pada Sidang Tugas Akhir

Program Studi Strata 1 Sistem Informasi

Fakultas Rekayasa Industri Universitas Telkom

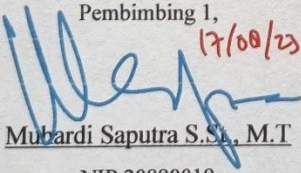
**Oleh :**

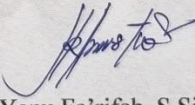
**Satria Adhy Nayoga**

**1202194226**

Bandung, 9 Agustus 2023

Disetujui oleh,

Pembimbing 1,  
  
**Mubardi Saputra S.Si., M.T**  
NIP 20890019

Pembimbing 2,  
  
**Riska Yanu Fa'rifah, S.Si., M.Si.**  
NIP 20900032

## LEMBAR PERNYATAAN ORISINALITAS



Nama : Satria Adhy Nayoga

NIM : 12022194226

Alamat : Komp Nata Endah Blok H.251,  
Kel. Margahayu Tengah, Kec.  
Margahayu, Kab. Bandung

No. Tlp : 081322913186

Menyatakan bahwa Tugas Akhir ini merupakan karya orisinal saya sendiri. Atas pernyataan ini, saya siap menanggung risiko atau sanksi yang dijatuhkan kepada saya apabila kemudian ditemukan adanya pelanggaran terhadap kejujuran akademik atau etika keilmuan dalam karya ini, atau ditemukan bukti yang menunjukkan ketidakaslian karya ini.

Bandung, Agustus 2023

Satria Adhy Nayoga

## KATA PENGANTAR

Puji dan syukur penulis panjatkan ke hadirat Allah yang telah melimpahkan rahmat serta karunia-Nya kepada kita. Shalawat serta salam selalu tercurah kepada Rasulullah SAW yang senantiasa menjadi pedoman dan tauladan bagi penulis serta seluruh umat manusia. Sehingga penulis bisa menyelesaikan Tugas Akhir dengan tepat waktu yang penulis beri Judul “ANALISIS SENTIMEN BERBASIS ASPEK PADA PERILAKU PENGGUNA *FINTECH* MENGGUNAKAN ALGORITMA LDA DAN *SUPPORT VECTOR MACHINE*”

Selama proses pembuatan Tugas Akhir ini, penulis merasa sangat bersyukur dan berterimakasih karena mendapat banyak bimbingan, bantuan, serta dukungan doa dari berbagai pihak. Oleh karena itu, penulis menyampaikan terima kasih yang sebanyak-banyaknya kepada:

1. Diri sendiri, Orang tua tercinta yaitu Bapak Agus Sukarsono serta Ibu Aprilia Kartika dan Ibu Alm Rima Aries Setiarini dan keluarga yang selalu memberikan dukungan dan doa kepada penulis dalam menyelesaikan Tugas Akhir ini.
2. Dosen wali saya, yaitu Bapak Ir. Ari Fajar Santoso, M.T. yang telah membimbing dan mendukung dalam pengerjaan tugas akhir dan perkuliahan.
3. Pembimbing Tugas Akhir saya, yaitu Bapak Muhardi Saputra S.St., M.T. dan Ibu Riska Yanu Fa'rifah, S.Si., M.Si. Selain itu, semua dosen S1 Sistem Informasi lainnya juga turut berkontribusi dengan memberikan dukungan, bimbingan, serta arahan yang berharga kepada penulis selama proses pengerjaan Tugas Akhir ini.
4. Teman satu kelompok tugas akhir, Yanuar Taruna Lutfi dan Sahra Bilqis Fauziyyah yang telah banyak membantu dan berkontribusi dalam proses penyelesaian tugas akhir ini serta banyak ilmu yang telah dibagikan kepada penulis.
5. Semua teman-teman jurusan Sistem Informasi terutama kelas SI-43-09 dan sahabat yang telah memberikan semangat, dukungan serta masukan yang membangun kepada penulis dalam menyusun proyek ini.

Dengan segala kerendahan hati, penulis menyadari bahwa masih terdapat kesalahan dan kekurangan serta masih jauh dari sempurna dalam penulisan dan pengerjaan Tugas Akhir ini. Besar harapan penulis, proyek ini bermanfaat bagi teman dan saudara di masa mendatang.

## DAFTAR ISI

ABSTRAK .....	ii
ABSTRACT .....	iii
LEMBAR PENGESAHAN .....	iv
LEMBAR PERNYATAAN ORISINALITAS .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI.....	viii
DAFTAR GAMBAR .....	xii
DAFTAR TABEL.....	xiii
DAFTAR ISTILAH .....	xiv
BAB I PENDAHULUAN.....	1
I.1 Latar Belakang .....	1
I.2 Perumusan Masalah .....	6
I.3 Tujuan Penelitian .....	7
I.4 Batasan Penelitian .....	7
I.5 Manfaat Penelitian .....	7
I.6 Sistematika Penulisan.....	8
BAB II TINJAUAN PUSTAKA.....	10
II.1 <i>Financial Technology</i> .....	10
II.2 OVO.....	12
II.3 <i>Technostress</i> .....	13
II.4 Google Play Store .....	14
II.5 Analisis Sentimen .....	14
II.5.1 Analisis Sentimen Berbasis Aspek .....	15
II.5.2 <i>Machine Learning</i> .....	15



II.5.3 <i>Text Preprocessing</i> .....	16
II.5.4 <i>Support Vector Machine</i> .....	17
II.6 Perbandingan Algoritma <i>Support Vector Machine, Naïve Bayes</i> dan <i>Decision Tree</i> .....	18
II.7 LDA .....	19
II.8 <i>Topic Coherence</i> .....	19
II.9 <i>Bag Of Words</i> .....	21
II.10 TF IDF .....	21
II.11 <i>Confusion Matrix</i> .....	22
II.12 <i>SentiStrength</i> .....	24
II.13 Penelitian Terdahulu .....	25
II.13.1 <i>Effective Comparison of LDA with LSA for Topic Modelling</i> .....	25
II.13.2 <i>Aspect-Based Sentiment Analysis in Beauty Product Reviews Using TF-IDF and SVM Algorithm</i> .....	26
II.13.3 <i>Impact of Digital Technostress and Digital Technology Self-Efficacy on Fintech Usage Intention of Chinese Gen Z Consumers</i> .....	26
II.13.4 <i>A Comparative Analysis of Social Communication Applications Using Aspect Based Sentiment Analysis 2022</i> .....	27
II.13.5 <i>Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM</i> .....	27
II.13.6 Analisis Sentimen Berbasis Aspek Pada Aplikasi Tokopedia Menggunakan LDA Dan <i>Naïve Bayes</i> .....	28
BAB III METODOLOGI PENELITIAN.....	34
III.1 Kerangka Berpikir .....	34
III.2 Sistematika Penyelesaian Masalah.....	35
III.2.1 Tahap Identifikasi Masalah.....	35
III.2.2 Tahap Implementasi.....	35

III.2.3 Tahap Penutup .....	36
III.3 Pengumpulan Data .....	36
III.4 Pengolahan Data.....	37
III.5 Metode Evaluasi .....	37
<b>BAB IV ANALISIS DAN PERANCANGAN .....</b>	<b>38</b>
IV.1 <i>Text Preprocessing</i> .....	38
IV.1.1 Pemisahan Kalimat.....	38
IV.1.2 <i>Case Folding</i> .....	38
IV.1.3 <i>Removal of Punctuation</i> .....	39
IV.1.4 <i>Removal of Number</i> .....	39
IV.1.5 <i>Spelling Correction</i> .....	40
IV.1.6 <i>Stemming</i> .....	40
IV.1.7 <i>Stopword Removal</i> .....	41
IV.1.8 <i>Null Value Removal</i> .....	41
IV.2 Pemodelan Topik .....	41
IV.2.1 Pembobotan Kata.....	41
IV.2.2 Pembuatan Model LDA.....	42
IV.2.3 Evaluasi .....	43
IV.3 Analisis Sentimen Berbasis Aspek.....	46
IV.3.1 <i>Aspect Extraction</i> .....	46
IV.3.2 <i>Aspect Sentiment Classification</i> .....	47
IV.4 <i>Split Dataset</i> .....	49
IV.5 TF-IDF .....	50
IV.6 Implementasi <i>Support Vector Machine</i> .....	51
<b>BAB V HASIL DAN PEMBAHASAN.....</b>	<b>53</b>
V.1 Evaluasi dan Validasi .....	53

V.2 Visualisasi <i>Word Cloud</i> .....	59
V.3 Grafik.....	63
BAB VI KESIMPULAN DAN SARAN .....	67
VI.1 Kesimpulan.....	67
VI.2 Saran.....	68
DAFTAR PUSTAKA .....	70

## DAFTAR GAMBAR

Gambar I.1 Metode Pembayaran yang Paling Sering Digunakan, sumber: katadata.co.id .....	3
Gambar I.2 <i>E-Wallet</i> yang Paling Sering Dipakai di Indonesia, sumber: boku.com.....	4
Gambar I.3 <i>E-wallet</i> yang paling banyak digunakan dalam 3 bulan terakhir sumber: Insight Asia .....	4
Gambar II.1 Evolusi <i>Fintech</i> .....	11
Gambar II.2 Logo OVO, sumber: berita99.co .....	12
Gambar II.3 <i>Hyperplane Support Vector Machine</i> .....	17
Gambar II.4 Perhitungan Coherence Score, sumber: João Pedro (2022) .....	20
Gambar III.1 Model Konseptual .....	34
Gambar III.2 Alur Sistematisa Penyelesaian Masalah .....	35
Gambar III.3 Hasil <i>Scraping</i> .....	36
Gambar IV.1 Grafik nilai <i>coherence</i> .....	43
Gambar IV.2 Visualisasi LDA.....	45
Gambar IV.3 Jumlah Ulasan Setiap Aspek.....	47
Gambar IV.4 Presentase sentimen .....	48
Gambar IV.5 Jumlah sentimen pada setiap aspek.....	48
Gambar V.1 <i>Confusion Matrix</i> Rasio 75:25 .....	54
Gambar V.2 <i>Confusion Matrix</i> Aspek Fitur.....	56
Gambar V.3 <i>Confusion Matrix</i> Aspek Akses.....	56
Gambar V.4 <i>Confusion Matrix</i> Aspek Pelayanan .....	57
Gambar V.5 <i>Confusion Matrix</i> Aspek Keamanan .....	58
Gambar V.6 Persentase Akurasi tiap Aspek .....	58
Gambar V.7 <i>Wordcloud</i> Ulasan Aspek Fitur Positif (a) dan Negatif (b).....	59
Gambar V.8 <i>Wordcloud</i> Ulasan Aspek Akses Positif (a) dan Negatif (b).....	60
Gambar V.9 <i>Wordcloud</i> Ulasan Aspek Layanan Positif (a) dan Negatif (b).....	61
Gambar V.10 <i>Wordcloud</i> Ulasan Aspek Keamanan Positif (a) dan Negatif (b) .....	62
Gambar V.11 Jumlah Sentimen per Bulan.....	64
Gambar V.12 Jumlah Aspek per Bulan .....	64
Gambar V.13 Jumlah Sentimen per Aspek pada Setiap Bulan.....	65

## DAFTAR TABEL

<i>Table II.1 Perbandingan Akurasi Algoritma</i> .....	18
<i>Table II.2 Confusion Matrix</i> .....	23
<i>Table II.3 Penelitian Terdahulu</i> .....	29
<i>Table IV.1 Case Folding</i> .....	38
<i>Table IV.2 Removal of Punctuation</i> .....	39
<i>Table IV.3 Removal of Number</i> .....	39
<i>Table IV.4 Spelling Correction</i> .....	40
<i>Table IV.5 Stemming</i> .....	40
<i>Table IV.6 Stopword Removal</i> .....	41
<i>Table IV.7 Hasil Topik</i> .....	42
<i>Table IV.8 Coherence Score</i> .....	44
<i>Table IV.9 Kata Kunci tiap Aspek</i> .....	45
<i>Table IV.10 Identifikasi Aspek</i> .....	46
<i>Table IV.11 Splitting Dataset per Rasio</i> .....	49
<i>Table IV.12 Splitting Dataset setiap Aspek per Rasio</i> .....	49
<i>Table IV.13 Sampel Pembobotan Kata</i> .....	50
<i>Table IV.14 Perhitungan TF dan IDF</i> .....	50
<i>Table IV.15 Perhitungan TF-IDF</i> .....	51
<i>Table V.1 Jumlah Akurasi pada setiap Rasio</i> .....	53
<i>Table V.2 Classification Report pada setiap Aspek</i> .....	55
<i>Table V.3 Jumlah Kemunculan Kata Aspek Fitur</i> .....	59
<i>Table V.4 Jumlah Kemunculan Kata Aspek Akses</i> .....	60
<i>Table V.5 Jumlah Kemunculan Kata Aspek Layanan</i> .....	61
<i>Table V.6 Jumlah Kemunculan Kata Aspek Keamanan</i> .....	63

## DAFTAR ISTILAH

Istilah	Deskripsi	Halaman pertama kali digunakan
<i>Fintech</i>	: <i>Financial Technology</i>	1
APJI	: Asosiasi Pengelola Jurnal Indonesia	1
<i>E-commerce</i>	: ( <i>Electronic Commerce</i> ) yaitu perdagangan secara elektronika atau pembelian dan penjualan produk serta jasa melalui internet	1
AFTECH	: Asosiasi <i>Fintech</i> Indonesia	2
QRIS	: <i>Quick Response Code Indonesia Standard</i>	2
P2P	: <i>Peer to Peer</i>	3
LDA	: ( <i>Latent Dirichlet Allocation</i> ) Metode pemodelan topik untuk mengetahui topik apa yang dibahas pada sekumpulan dokumen	5
ABSA	: ( <i>Aspect Based Sentiment Analysis</i> ) Analisis sentimen yang berfokus pada aspek atau atribut khusus pada teks dalam mengekstraksi sentimen	5
SVM	: <i>Support Vector Machine</i>	6
TF-IDF	: ( <i>Term Frequency-Inverse Document Frequency</i> ) Metode yang digunakan dalam mengukur frekuensi kata dan menggambarkan seberapa penting suatu kata dalam dokumen.	10
NB	: <i>Naïve Bayes</i>	12
DT	: <i>Decision Tree</i>	24
BOW	: ( <i>Bag of Words</i> ) Metode yang digunakan dalam mengubah teks menjadi vektor dengan menghitung frekuensi kemunculan term/kata pada dokumen.	25
NLP	: ( <i>Natural Language Processing</i> ) bagian dari kecerdasan buatan yang berfokus pada interaksi antara manusia serta komputer melalui bahasa manusia yang alami	25
TN	: <i>True Negative</i>	28

TP	:	<i>True Positive</i>	28
FN	:	<i>False Negative</i>	28
FP	:	<i>False Positive</i>	28

# **BAB I**

## **PENDAHULUAN**

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

*Fintech* atau *Financial Technology* telah menjadi pendorong perubahan global yang paling menonjol dalam industri keuangan. Transformasi ini telah terjadi dalam layanan seperti pembayaran digital, asuransi, atau manajemen kekayaan aktif, karena faktor-faktor seperti evolusi teknologi yang cepat, perubahan ekonomi makro, lingkungan regulasi, dan ekspektasi konsumen yang berubah (Moreira-Santos dkk., 2022). Perkembangan teknologi *smartphone* saat ini memang luar biasa. *Smartphone* telah menjadi kebutuhan primer bagi setiap manusia untuk memenuhi kebutuhan komunikasi dan informasi sehari-hari. Pengguna *internet* di Indonesia tahun 2017 berjumlah 143,26 juta orang dari 262 juta penduduk Indonesia dan 50,08% pengguna *smartphone* sebanyak 131 juta orang. Indonesia memiliki modal besar untuk mendukung perkembangan *fintech*, yaitu jumlah masyarakat kelas menengah yang mencapai 45 juta orang, dan total pengguna internet yang mencapai 150 juta. Karena itu, Indonesia menjadi pengguna *smartphone* terbesar keempat di dunia setelah China, India, dan Amerika (Warjiyono dkk., 2019).

*Fintech* telah merambah pasar dunia. Tingkat adopsi di Cina dan India mencapai 87%. Di negara lain, *fintech* juga menunjukkan gangguan pasar yang cepat, seperti di Rusia dan Afrika Selatan yang mencapai 82%. Australia, Kanada, Hong Kong, Singapura, Inggris Raya, dan AS mencapai 60% pada tahun 2019 (Nanggala, 2020). Perkembangan teknologi informasi di Indonesia saat ini terus meningkat, Indonesia sebagai negara berkembang harus selalu mengikuti *trend* penggunaan teknologi yang ada. Salah satu bentuk perkembangan teknologi informasi digital di Indonesia saat ini adalah meningkatnya pengguna internet dan munculnya berbagai jenis perusahaan di bidang bisnis *e-commerce* di Indonesia. Berdasarkan hasil survei APJI dan *Polling* Indonesia jumlah pengguna internet di Indonesia tahun 2018 meningkat 27,91 juta (10,12%) menjadi 171,18 juta orang dari tahun sebelumnya. Artinya, penetrasi internet di Indonesia meningkat menjadi 64,8% dari total populasi 264,16 juta jiwa. Sedangkan pada tahun yang sama, jumlah



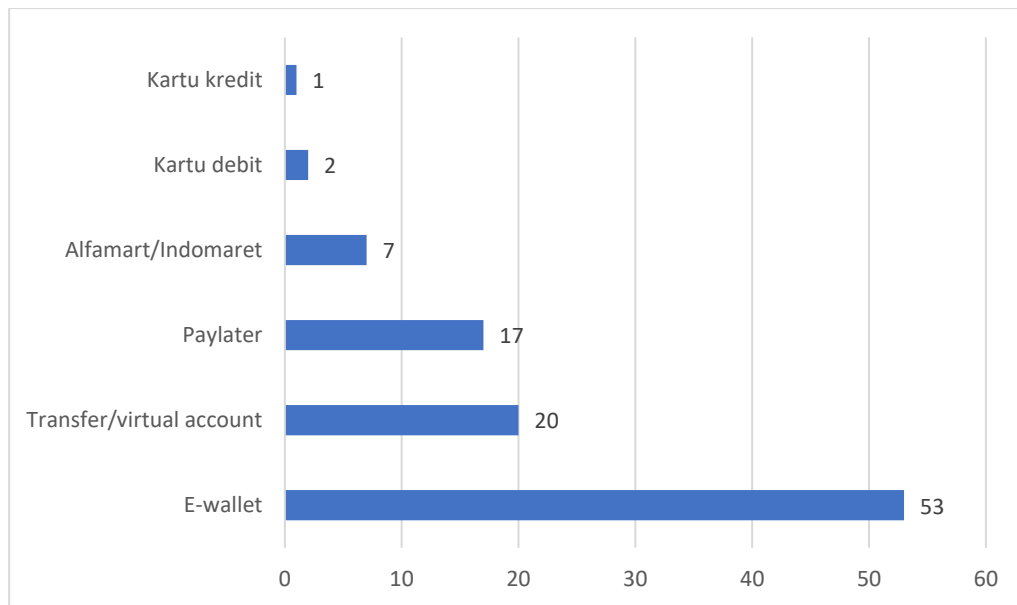
transaksi *e-commerce* di Indonesia telah mencapai sekitar 144 triliun Rupiah di Indonesia (Putra & Sfenrianto, 2020).

Keberadaan *fintech* di Indonesia terus berkembang pesat. Hal ini terbukti dari hasil survei yang dilakukan oleh AFTECH (Asosiasi *Fintech* Indonesia) Jumlah perusahaan *fintech* di Indonesia semakin bertambah dan mereka memberikan kontribusi yang semakin besar terhadap perekonomian negara. Pada tahun 2021, total investasi pada layanan *fintech* mencapai 904 juta USD, sebuah angka yang signifikan. Investasi besar ini bahkan telah menciptakan tiga dari delapan perusahaan *fintech* "unicorn" yang saat ini aktif beroperasi di sektor *fintech*. Adopsi teknologi *fintech* juga mengalami peningkatan yang cukup pesat pada tahun 2021. Contohnya adalah penggunaan uang elektronik, di mana nilai transaksi mencapai 58,5% atau setara dengan 35 triliun rupiah. Tidak hanya itu, keberhasilan dalam mengadopsi sistem *Quick Response Code Indonesia* (QRIS) juga patut diperhatikan, karena telah berhasil melampaui target dengan mencapai 12 juta pedagang yang menggunakan sistem ini. Dalam hal penyaluran pinjaman melalui *fintech*, jumlah rekening yang terlibat mencapai 13,47 juta rekening dengan total dana pinjaman sebesar kurang lebih 13,6 triliun Rupiah pada Desember 2021 (AFTECH, 2021).

Industri *fintech* menunjukkan prospek perkembangan pesat karena aspek inovasi yang dinamis. pada beberapa tahun ini berkembang sangat pesat di beberapa negara terutama di Indonesia, hal tersebut terjadi karena tentunya yang mendasari yaitu penggunaan internet di Indonesia yang meningkat dengan cepat dan pengguna internet yang terus bertambah setiap waktunya, lalu permintaan dari konsumen akan adanya inovasi untuk mendapatkan pelayanan atau pengalaman yang nyaman, cepat, efisien serta murah dalam menggunakan layanan keuangan terutama yang berasal dari generasi yang tumbuh bersama dengan teknologi digital pada generasi *millennials* dan generasi z tidak hanya pada aspek keuangan namun juga pada banyak aspek lainnya, seperti kesehatan, pendidikan, transportasi, ritel, hotel dan lainnya (Nizar, 2020).

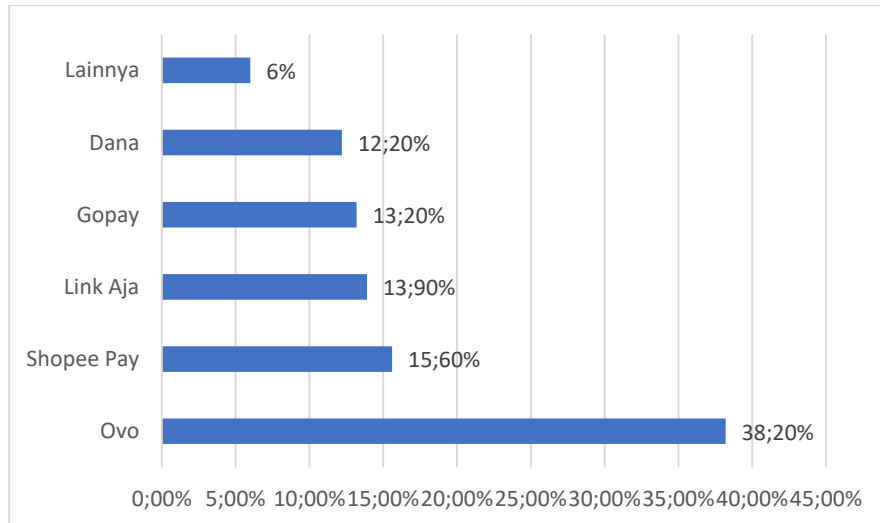
Perbankan juga tidak punya pilihan selain merangkul digitalisasi, sepenuhnya mengubah cara menyediakan layanan dan berinteraksi dengan pelanggan

komersial dan korporat, karena dua faktor utama: digitalisasi yang cepat di semua sektor yang membentuk ekonomi dan munculnya pesaing baru di sektor keuangan yang modelnya didasarkan pada teknologi baru (Barroso & Laborda, 2022). Kemajuan teknologi informasi dan perangkat komunikasi telah mengubah proses bisnis transaksi keuangan. Jika sebelumnya transaksi keuangan banyak dilakukan secara langsung atau tatap muka, kemudian beralih ke transaksi dengan sistem yang melibatkan kabel, dan kini trennya beralih ke transaksi *online* (Abdillah, 2019).



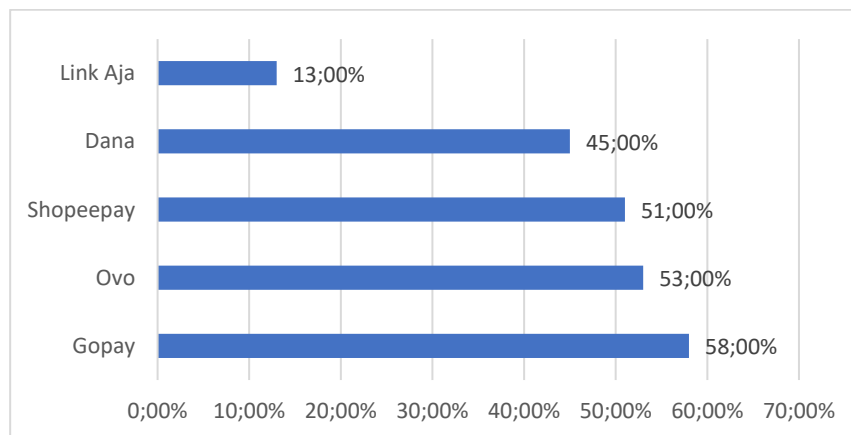
Gambar I.1 Metode Pembayaran yang Paling Sering Digunakan, sumber: katadata.co.id

*Fintech* merupakan ranah yang mencakup berbagai macam jenis layanan, dan di Indonesia, pemanfaatan *e-wallet* telah mengemuka sebagai salah satu produk yang paling mendominasi dalam penerapannya oleh masyarakat, seperti yang terungkap dari data yang dirilis oleh databoks (databoks, 2022), Meski begitu, tak hanya *e-wallet* yang mendapat perhatian, sebab jenis-jenis *fintech* lainnya seperti P2P lending dan layanan investasi juga turut meramaikan lanskap *fintech* di negara ini.



Gambar I.2 *E-Wallet* yang Paling Sering Dipakai di Indonesia, sumber: buku.com

Berdasarkan Laporan *Mobile Wallets Report 2021* menunjukkan bahwa OVO sebagai *e-wallet* yang paling banyak digunakan, disertai dengan beberapa *e-wallet* lainnya di Indonesia seperti ShopeePay, Link Aja, dan Gopay (Annur, 2021).



Gambar I.3 *E-wallet* yang paling banyak digunakan dalam 3 bulan terakhir sumber: Insight Asia

Namun, berdasarkan pada gambar I.3 pada tahun 2022 berdasarkan laporan data dari Insight Asia, OVO berhasil disusul oleh Gopay menjadi dompet digital yang paling sering digunakan dalam 3 bulan terakhir oleh para responden (Insight Asia, 2022). Berdasarkan *rating* aplikasi yang didapatkan dari Google Play Store, aplikasi OVO mengalami penurunan pada *rating* (Rosyid dkk., 2022) yaitu (4.1) dibandingkan dengan Gopay sebesar (4,5). Penurunan pada *rating* tersebut dikarenakan banyak pengguna yang mengeluh terhadap kinerja aplikasi yang lambat, batasan fitur, *limit* transaksi, dan lainnya (Rosyid dkk., 2022). Menurut

pandangan Weil dan Rosen, tanda-tanda keluhan yang dilaporkan oleh pengguna mencerminkan bahwa mereka sedang mengalami *Technostress* (Tacy, 2015). *Technostress* menjadi salah satu faktor penyebab penurunan jumlah pengguna karena ditemukan beberapa konsumen memutuskan berhenti menggunakan layanan digital seperti *fintech* ketika mereka mengalami *technostress* (Lee, 2022). bahkan konsumen muda dan terpelajar pun cenderung merasa kesulitan untuk terus-menerus memperoleh teknologi digital baru karena hal ini berubah dengan cepat dari hari ke hari. Selain tekanan untuk memperoleh teknologi digital baru yang terus diperbarui, ada banyak jenis *technostress* lainnya, seperti masalah invasi privasi, ketidakstabilan keamanan digital, kesulitan dalam menggunakan perangkat digital yang kompleks, dan tekanan untuk mengganti perangkat digital baru karena terus-menerus pembaruan teknologi digital (Lee, 2021).

Sehingga untuk dapat memahami apa saja aspek yang mempengaruhi penurunan pengguna dan penyebab maka diperlukan untuk menganalisis perilaku penerimaan pengguna *fintech*, penulis ingin mengetahui apa yang membuat pengguna tidak memilih menggunakan layanan OVO, lalu apa yang membuat mereka tidak puas dan beralih ke kompetitor dan kepuasan terhadap layanan *fintech* OVO. Dan karena hal tersebut diputuskan bahwa ovo akan digunakan sebagai bahan penelitian pada topik kali ini karena paling banyak digunakan oleh masyarakat indonesia dengan maksud agar data yang didapatkan akan lebih banyak sehingga dapat memudahkan berjalannya penelitian ini dengan menggunakan proses *sentiment analysis* menggunakan *machine learning*.

Menurut Medhat dkk. (2014) *Sentiment Analysis* (SA) adalah studi komputasi tentang pendapat, sikap, dan emosi orang terhadap suatu entitas. Entitas dapat mewakili individu, peristiwa, atau topik. Namun, analisis sentimen tradisional berfokus pada mengklasifikasikan sentimen keseluruhan yang diungkapkan dalam teks tanpa menentukan tentang apa sentimen itu. Ini mungkin tidak cukup jika teks secara bersamaan mengacu pada topik atau entitas yang berbeda (juga dikenal sebagai aspek), mungkin mengekspresikan sentimen yang berbeda terhadap aspek yang berbeda. Mengidentifikasi sentimen yang terkait dengan aspek-aspek tertentu dalam teks adalah tugas yang lebih kompleks yang dikenal sebagai

analisis sentimen berbasis aspek (ABSA) (Hoang dkk., 2019). Untuk mendapatkan aspek-aspek yang berhubungan dengan *technostress* dipilih menggunakan topic modelling LDA karena LDA adalah metode pemodelan topik yang efektif untuk membuat konteks kumpulan dokumen dan memberikan topik yang lebih berkualitas (Albalawi dkk., 2020). Topik ini kemungkinan besar akan dicakup oleh ulasan. Sehingga dapat mengukur sentimen seseorang terhadap teknologi atau produk.

Selama ini orang menilai penerimaan teknologi hanya dengan menggunakan model-model dikembangkan sejak dulu dan instrumen menggunakan kuesioner sementara kuesioner memiliki kekurangan seperti tingkat respon yang rendah, sampel yang terbatas, biaya yang dikeluarkan, kemungkinan resiko kesalahan yang besar (Alderman & Salem, 2010). Sehingga penulis memilih menggunakan *machine learning* untuk menganalisis data yang ada dengan lebih akurat dan otomatis dengan algoritma SVM karena merupakan algoritma terbaik menurut beberapa studi (Pinem dkk., 2018). SVM juga telah ditemukan memberikan akurasi yang lebih baik dalam hal mengklasifikasikan teks. Karena SVM adalah pengklasifikasi biner, mereka lebih cocok dalam mengklasifikasikan polaritas kalimat (Mahtab dkk., 2018). Sehingga dalam penelitian ini akan menggunakan analisis sentimen berbasis aspek pada ulasan pengguna OVO dari Google Play Store dalam menilai tingkat kepuasan pengguna serta mengetahui hal-hal yang tidak memuaskan dari pengguna dari sisi *technostress* yang klasifikasikan menjadi kategori sentimen positif dan negatif. Serta penelitian ini diharapkan dapat menjadi sebagai masukan akan perbaikan terhadap keseluruhan aspek dari OVO sehingga kualitas dari aplikasi OVO dapat lebih ditingkatkan dalam masa yang akan datang.

## **I.2 Perumusan Masalah**

Rumusan masalah yang mendasari penelitian ini adalah:

1. Bagaimana penggunaan *topic modelling* dalam penentuan aspek atau topik yang terdapat pada data?
2. Bagaimana implementasi analisis sentimen berbasis aspek menggunakan algoritma *Support Vector Machine* berdasarkan aspek yang telah ditentukan?

3. Aspek-aspek apa saja yang menyebabkan technostress dari pengguna OVO?

### **I.3 Tujuan Penelitian**

Penelitian ini bertujuan untuk:

1. Mengetahui penggunaan *topic modelling* dalam penentuan aspek atau topik yang terdapat pada data.
2. Mengimplementasi analisis sentimen berbasis aspek menggunakan algoritma *Support Vector Machine* berdasarkan aspek yang telah ditentukan.
3. Mendapatkan informasi serta insight yang didapatkan dari hasil analisis sentimen berbasis aspek.

### **I.4 Batasan Penelitian**

Batasan penelitian ini melingkupi sebagai berikut:

1. Jenis *Fintech* yang dipilih sebagai bahan penelitian yaitu OVO.
2. Penelitian berasal dari data ulasan yang didapatkan Google Play.
3. Data yang diambil untuk penelitian ini hanya untuk periode rentang waktu 1 januari 2023 hingga 13 maret 2023.
4. Algoritma yang dipilih untuk digunakan pada analisis sentimen yaitu *Support Vector Machine* (SVM).
5. *Topic modeling* yang digunakan yaitu *Latent Dirichlet Allocation* (LDA).
6. Perilaku pengguna yang menjadi bahan penelitian ini yaitu *technostress*
7. Aspek yang diidentifikasi sebagai faktor *technostress* adalah aspek yang mempunyai ulasan sentimen negatif yang tinggi.
8. Pelabelan menggunakan *Sentistrength* yang dilakukan setelah *text-preprocessing*.

### **I.5 Manfaat Penelitian**

Manfaat dari penelitian ini:

1. Bagi Industri,  
Informasi sentimen pengguna OVO berguna untuk evaluasi dan peningkatan layanan. Perusahaan dapat menyesuaikan kinerja mereka berdasarkan umpan balik pengguna, seperti meningkatkan respons layanan pelanggan atau

memperluas fitur yang disukai pengguna. memberikan informasi bagi perusahaan OVO akan sentimen yang diberikan oleh pengguna layanan tersebut yang dapat berguna sebagai bahan evaluasi dalam peningkatan layanan dalam aspek-aspek yang dibahas.

2. Bagi Akademisi,

Informasi tentang OVO dan *fintech* menjadi wawasan dan referensi untuk penelitian, terutama terkait implementasi algoritma dan penerimaan teknologi. Penelitian ini membantu mengembangkan pemahaman tentang *fintech* dan memandu penelitian masa depan. dapat dijadikan sebagai wawasan ataupun referensi bagi penelitian yang akan dikembangkan di masa yang akan datang terutama dalam hal implementasi algoritma dan penerimaan teknologi khususnya *technostress*.

3. Bagi penulis,

Informasi tentang perkembangan *fintech* dan penggunaan OVO digunakan untuk memperkaya karya penulis. Hal ini membantu menyajikan informasi yang akurat dan terkini kepada pembaca serta menggambarkan implikasi *fintech* dalam kehidupan sehari-hari. menambah wawasan dan informasi terkait perkembangan serta penggunaan *fintech*.

## **I.6 Sistematika Penulisan**

Sistematika penulisan pada penelitian ini adalah sebagai berikut:

### **Bab I      Pendahuluan**

Pada bab ini mencakup beberapa aspek penting, termasuk latar belakang penelitian, perumusan masalah, tujuan penelitian, batasan penelitian, dan manfaat penelitian. Latar belakang menyajikan alasan di balik pelaksanaan penelitian ini, diikuti oleh perumusan masalah yang menguraikan pertanyaan-pertanyaan penelitian yang akan dijawab oleh tujuan penelitian. Batasan penelitian digunakan untuk membatasi ruang lingkup penelitian, sedangkan manfaat penelitian mencakup kontribusi peneliti dalam memberikan hasil penelitian yang bermanfaat.

## **Bab II Tinjauan Pustaka**

Pada bab ini menguraikan tentang aspek ilmiah dasar atau teori yang relevan dengan permasalahan yang ada dalam penelitian ini. Lebih dari itu, bab ini mencakup penemuan-penemuan sebelumnya yang terkait dengan topik penelitian ini, bertujuan untuk melengkapi pemahaman tentang subjek yang sedang diteliti.

## **Bab III Metodologi Penelitian**

Bab ini menjelaskan pendekatan yang diterapkan dalam penelitian ini. Selain itu, bab ini juga mencakup metode konseptual, sistematika penyelesaian masalah, proses pengumpulan dan pengolahan data, serta metode evaluasi yang digunakan dalam penelitian ini. Dalam bagian langkah-langkah penyelesaian masalah, terdapat tahap identifikasi masalah, analisis sentimen berbasis aspek, dan kesimpulan serta saran.

## **Bab IV Analisis dan Perancangan**

Pada bab ini dijelaskan secara detil terkait proses-proses seperti, tahapan praproses data, penentuan model topik, pembentukan ABSA, pembobotan kata, pemisahan *dataset* berdasarkan rasio dan perancangan model klasifikasi *Support Vector Machine*.

## **Bab V Hasil dan Pembahasan**

Pada bab ini mencakup perincian tentang temuan dari eksperimen model *Support Vector Machine*, evaluasi, serta langkah validasi menggunakan *Confusion Matrix* dan pengujian *overfitting*. Disertai pula dengan proses pembuatan visualisasi berupa *wordcloud*, grafik *line chart*, dan *heatmap*.

## **Bab VI Kesimpulan dan Saran**

Pada bab ini dijelaskan kesimpulan dari penyelesaian masalah yang dilakukan serta jawaban dari rumusan permasalahan yang ada pada bagian pendahuluan. Saran dari solusi dikemukakan pada bab ini untuk tugas akhir serta penelitian selanjutnya.



## BAB II

### TINJAUAN PUSTAKA

#### **II.1 *Financial Technology***

*Fintech* adalah kata yang dibuat dengan menggabungkan "keuangan" dan "teknologi," dan dapat dikatakan bahwa itu adalah teknologi yang menerapkan TI ke dunia keuangan. *Fintech* pada dasarnya sesuatu yang memiliki potensi untuk mengubah cara keuangan beroperasi, teknologi yang dapat berkontribusi pada bentuk keuangan baru dan yang akan melahirkan layanan keuangan baru (Nakashima, 2018). Inovasi ini termasuk, misalnya, solusi pembayaran seluler untuk konsumen dan pedagang, pasar *online*, pinjaman, tabungan algoritmik dan alat investasi, mata uang virtual, pelanggan digital biometrik identifikasi dan otentikasi, dan fungsi perusahaan *mid-office* dan *back-office* otomatis, seperti seperti penggunaan algoritma, big data, kecerdasan buatan, dan *link analytics* (National Economic Council, 2017).

*Fintech* mempunyai beragam layanan atau produk yang dapat dimanfaatkan. Namun, berdasarkan Bank Indonesia, dibagi menjadi 4 jenis yaitu:

1. *P2P Lending* dan *Crowdfunding*

Bisa disebut juga sebagai *marketplace* finansial. *Platform* ini mampu mempertemukan pihak yang membutuhkan dana dengan pihak yang memberikan dana yang digunakan untuk investasi atau modal. Biasanya, proses pada *P2P lending* ini lebih mudah karena bisa dilakukan pada satu *online platform* (H Kara, 2014).

2. Manajemen Risiko Investasi

Digunakan untuk mengawasi kondisi keuangan serta merencanakan keuangan secara lebih praktis dan mudah (Kara, 2014).

3. Sistem Pembayaran

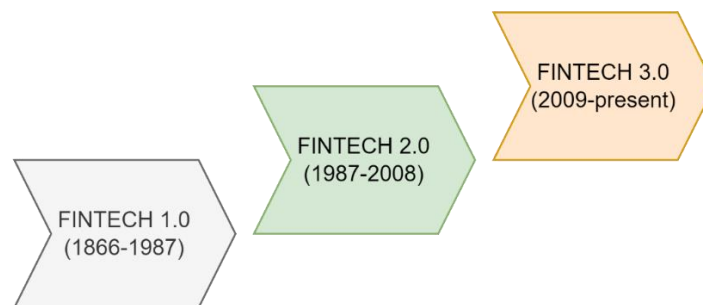
Menurut PBI No. 19/12/PBI No.2017, pengertian *fintech* di sistem pembayaran yaitu sebagai berikut: penggunaan teknologi dalam sistem keuangan yang menghasilkan produk, layanan, teknologi, dan/atau model bisnis baru dan dapat berdampak pada stabilitas sistem keuangan, stabilitas

moneter, dan/atau keandalan keamanan, kelancaran, dan efisiensi sistem pembayaran(AFTECH, 2021).

#### 4. *Market Aggregator*

Jenis *fintech* saat ini mengacu pada portal yang mengumpulkan berbagai informasi keuangan untuk disajikan kepada kelompok sasaran atau pengguna. Biasanya, *fintech* jenis ini mencakup berbagai informasi, kiat keuangan, kartu kredit, dan investasi. Dengan *fintech* jenis ini, harapannya adalah Anda dapat mengambil banyak informasi sebelum mengambil keputusan keuangan (H Kara, 2014).

Evolusi *fintech* telah terjadi sejak awal perkembangan keuangan dan teknologi, berikut perkembangan evolusi *fintech* diantaranya:



Gambar II.1 Evolusi *Fintech*

1. *Fintech* 1.0 (1866-1987), evolusi *fintech* dimulai dengan munculnya sistem administrasi keuangan di Mesopotamia dan perkembangan mata uang fiksi di perekonomian modern. Proses ini juga terlihat dalam perkembangan teknologi perhitungan dan perdagangan (misalnya dengan membiayai dan menjamin kapal dan infrastruktur seperti jembatan, jalur kereta api, dan kanal) maupun dalam mendukung produksi barang untuk perdagangan tersebut, serta akuntansi *double entry* yang fundamental bagi perekonomian modern. Revolusi keuangan di Eropa pada akhir 1600-an memainkan peran penting dalam Revolusi Industri dengan mendukung perkembangan teknologi yang menjadi landasan dari perkembangan industri (Arner, 2015) .
2. *Fintech* 2.0 (1987-2008): Pengembangan Layanan Keuangan Digital Tradisional *Fintech* 2.0 merupakan tahap selanjutnya dari perkembangan *fintech* yang terjadi antara tahun 1987 hingga 2008. Pada tahap ini, terjadi

integrasi lebih lanjut antara keuangan dan teknologi, dengan munculnya layanan keuangan digital yang ditawarkan oleh perusahaan *fintech*. Layanan keuangan digital ini meliputi layanan investasi, layanan perbankan digital, dan lainnya. Selain itu, terjadi juga integrasi lebih lanjut antara *fintech* dan industri keuangan tradisional, seperti bank dan lembaga keuangan lainnya. Pada tahap ini, juga terjadi perkembangan teknologi yang lebih canggih, seperti penggunaan kecerdasan buatan (AI) dan analitik big data dalam *fintech*. Selain itu, terjadi juga munculnya internet sebagai faktor yang menjadi landasan perkembangan *fintech* selanjutnya (Arner, 2015).

3. *Fintech* 3.0 (2009 – present), setelah terjadi krisis keuangan global pada tahun 2008, terjadi perubahan pandangan dari pelanggan retail tentang siapa yang memiliki sumber daya dan legitimasi untuk menyediakan layanan keuangan. Pada tahap ini, *fintech* telah mengalami pertumbuhan yang signifikan karena didukung oleh faktor-faktor seperti persepsi publik, pengawasan regulator, kebutuhan politik, dan kondisi ekonomi. Pada tahap ini, terjadi perkembangan yang signifikan dalam layanan keuangan digital yang ditawarkan oleh *fintech*, termasuk layanan perbankan, pembayaran, investasi, dan lainnya. Selain itu, terjadi juga peningkatan akses terhadap layanan keuangan ini, dengan banyaknya pengguna internet di seluruh dunia yang memungkinkan akses yang lebih luas ke layanan keuangan *digital* ini. Setelah tahun 2008, terjadi kecocokan kondisi pasar yang mendukung munculnya pelaku pasar yang inovatif di industri jasa keuangan, termasuk *fintech* yang menerapkan teknologi dalam layanan keuangan (Arner, 2015).

## II.2 OVO



Gambar II.2 Logo OVO, sumber: berita99.co

PT. Visionet Internasional merupakan salah satu perusahaan pembayaran terbesar di Indonesia, yang salah satu produknya bernama OVO yang disetujui oleh Bank

Indonesia pada 7 Agustus 2017. OVO tersedia gratis dan dapat diunduh di App Store dan Google Play (Silaen & Prabawani, 2020).

### **II.3 Technostress**

Dr. Craig Brod menjelaskan *Technostress* adalah penyakit adaptasi modern yang disebabkan oleh ketidakmampuan untuk mengatasi teknologi komputer baru dengan cara yang sehat. Ini memanifestasikan dirinya dalam dua cara yang berbeda tetapi terkait: dalam perjuangan untuk menerima teknologi komputer, dan dalam bentuk identifikasi berlebihan yang lebih khusus dengan teknologi komputer (Norulkamar & Ahmad, 2009). Selain itu, mereka telah mengidentifikasi lima komponen *technostress*, juga dikenal sebagai pencipta *technostress*, yaitu:

1. *Techno-overload*: Situasi di mana pengguna TIK dipaksa untuk bekerja lebih cepat dan lebih lama.
2. *Techno-invasion*: Situasi di mana pengguna TIK merasa bahwa mereka dapat dihubungi kapan saja atau terus-menerus “terhubung” yang menyebabkan kaburnya konteks terkait pekerjaan dan pribadi.
3. *Techno-complexity*: Situasi di mana pengguna TIK merasa keterampilan mereka tidak memadai karena kompleksitas terkait TIK. Akibatnya, mereka terpaksa menghabiskan waktu dan tenaga untuk mempelajari dan memahami berbagai aspek TIK.
4. *Techno-insecurity*: Situasi di mana pengguna TIK merasa terancam akan kehilangan pekerjaan, baik digantikan oleh TIK baru atau oleh orang lain yang lebih baik dalam TIK dibandingkan dengan mereka.
5. *Techno-uncertainty*: Situasi di mana pengguna TIK merasa tidak pasti dan tidak tenang karena TIK terus berubah dan perlu ditingkatkan.

*Technostress* didefinisikan sebagai stres atau penyakit psikosomatik yang disebabkan oleh bekerja dengan teknologi komputer setiap hari. Psikolog klinis Craig Brod menciptakan istilah *technostress* pada awal 1980-an, sehingga mendefinisikannya sebagai gangguan psikologis yang dialami oleh individu ketika mereka berinteraksi dengan teknologi. *Technostress* didefinisikan oleh Weil & Rosen sebagai "setiap efek negatif pada sikap, pikiran, perilaku, dan

psikologi manusia yang secara langsung atau secara tidak langsung dihasilkan dari teknologi" (Tacy, 2015). *Technostress* mempengaruhi kinerja pengguna dan niat berkelanjutan untuk menggunakan berbagai teknologi (Chou & Chou, 2021).

#### **II.4 Google Play Store**

Google Play Store, yang juga disebut sebagai "Google Play," adalah platform digital yang dimiliki oleh perusahaan Google. Secara bawaan, Google Play telah terpasang di perangkat Android seperti ponsel, tablet, TV Android, dan perangkat Google TV. Selain itu, Google Play Store juga dapat diakses secara *online* melalui *browser web*. Tujuan utama Google Play Store adalah menyediakan tempat bagi pengguna Android untuk mengunduh aplikasi dan game. Selain aplikasi, game, dan konten lainnya, Google Play Store juga menawarkan kategori seperti "Film & TV" dan "Buku". Namun, perubahan terjadi dalam hal konten musik, yang kemudian dipisahkan dan diperkenalkan secara terpisah melalui layanan YouTube *Music*. Secara sederhana, Google Play Store adalah *platform* tempat pengguna dapat mengunduh dan membeli berbagai produk digital yang ditawarkan oleh Google (Kompasiana, 2021).

#### **II.5 Analisis Sentimen**

Analisis sentimen adalah metode untuk menganalisis dan mengekstrak pengetahuan menggunakan pernyataan subjektif sebagai sumber data, yang dipublikasikan melalui Internet. Data yang digunakan dalam analisis idealnya perlu mengungkapkan secara eksplisit pendapat penulis tentang aspek subjek, termasuk merek, produk, dan layanan. Tugas klasifikasi analisis berarti memberi label data tekstual sebagai "Positif", "Negatif", atau "Netral". Namun, dalam penelitian ini, tujuannya adalah untuk memisahkan data menjadi "Positif" dan "Negatif" saja. Berbagai metode dan teknik telah diterapkan pada penelitian-penelitian sebelumnya. Metode-metode tersebut dapat dikategorikan menjadi tiga kategori: Teknik sederhana, seperti *K-Nearest Neighbor* (KNN), *Decision Tree*, *Logistic Regression*, dan *Naïve Bayes*. Teknik Lanjutan, biasanya dibuat untuk bentuk data tertentu, seperti beberapa Algoritma Pembelajaran Mendalam. *Ensemble* atau kombinasi teknik, seperti *Random Forest*, *Light Gradient Boosting*

*Algorithm* (LGBM), *eXtreme Gradient Boosting* XGBoost, *Adaptive Boosting* (AdaBoost), dan *Voting Classifier* (Vot) (Utami dkk., 2022).

### **II.5.1 Analisis Sentimen Berbasis Aspek**

*Aspect Based Sentiment Analysis* (ABSA) adalah tugas yang lebih kompleks daripada analisis sentimen tingkat teks tradisional. Ini berfokus pada mengidentifikasi atribut atau aspek dari suatu entitas yang disebutkan dalam sebuah teks, bersama dengan sentimen yang diungkapkan terhadap setiap aspek (Hoang dkk., 2019).

*Aspect Based Sentiment Analysis* (ABSA) adalah solusi untuk masalah dalam analisis sentimen yang tidak dapat mengklasifikasikan ulasan ke dalam kelas positif atau negatif berdasarkan kategori aspek. Pada level ini, sebuah model akan dapat mengklasifikasikan sebuah dokumen teks ke dalam kategori aspek dan polaritas sentimen. Misalnya pada kalimat review “Makanannya enak tapi harganya lumayan mahal”, model ABSA mampu mengklasifikasikan kalimat “makanannya enak” ke dalam aspek makanan dan kelas positif, kemudian kalimat “harganya lumayan mahal ” ke dalam aspek harga dan kelas negatif (Amalia & Winarko, 2021).

Tugas berikutnya dalam *Aspect Based Sentiment Analysis* (ABSA) adalah untuk mengenali sentimen yang terkait dengan menentukan nilai polaritas sentimen untuk setiap aspek yang telah diekstraksi. Nilai tersebut menunjukkan tingkat positif atau negatif dari pendapat tentang aspek tersebut. Ada banyak metode yang telah diusulkan untuk identifikasi sentimen, salah satunya adalah menggunakan pendekatan leksikon yang menggunakan kamus kata-kata dengan label sentimen untuk menilai skor sentimen (D’Aniello dkk., 2022).

### **II.5.2 Machine Learning**

*Machine Learning* bergantung pada algoritma yang berbeda untuk memecahkan masalah data. *Machine learning* digunakan untuk mengajari mesin cara menangani data dengan lebih efisien. Terkadang setelah melihat data, kita tidak dapat menginterpretasikan ekstrak informasi dari data tersebut. Sehingga

diterapkan penggunaan *machine learning* (Malviya dkk., 2020). Algoritma yang biasa digunakan pada machine learning diantaranya:

1. *Supervised*, Dalam *dataset* pelatihan algoritma ini dengan kelas pra-label diberikan dan berdasarkan *dataset* terlatih ini, input diberi label dengan kelas/hasil keluaran, Algoritma ini mengklasifikasikan set data input dengan bantuan *classifier* terlatih (Malviya dkk., 2020). Contohnya seperti *decision tree*, naïve bayes, dan *support vector machine* (Malviya dkk., 2020).
2. *Unsupervised*, Jenis algoritma pembelajaran mesin ini mengambil data input yang tidak berlabel dan kemudian dengan bantuan algoritma yang berbeda, struktur/pola tersembunyi ditemukan. Berbeda dengan *supervised*, teknik ini tidak menggunakan data pralabel untuk melatih pengklasifikasi (Malviya dkk., 2020).
3. *Semi-supervised*, adalah kombinasi dari *supervised* dan *unsupervised*. Jenis algoritma ini berurusan dengan set data berlabel dan tidak berlabel (Malviya dkk., 2020). Contohnya seperti *generative models*, *self training*, dan *transductive support vector machine*.

### **II.5.3 Text Preprocessing**

*Text preprocessing* adalah sebuah tahapan seleksi data yang dilakukan pada setiap dokumen dan hasil dari *preprocessing* dapat berdampak pada akurasi dari dokumen klasifikasi. Terdapat beberapa teknik dalam proses *preprocessing* (Pradha dkk., 2019), diantaranya:

#### 1. *Lower Text*

Pada proses ini dilakukan perubahan pada setiap huruf kapital yang ada menjadi huruf kecil.

#### 2. *Removal of Punctuation*

Pada proses ini dilakukan penghapusan pada tanda baca atau kata non-alfanumerik pada teks asli.

#### 3. *Spelling Correction*

Pada proses ini dilakukan perbaikan pada ejaan kata yang kurang tepat atau tidak baku menjadi bentuk asalnya.

#### 4. *Stemming*

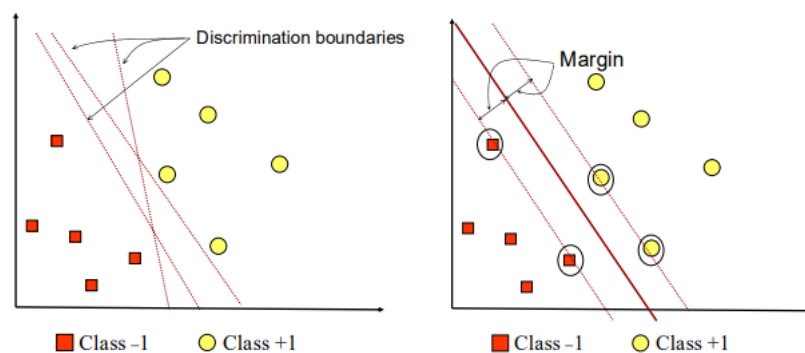
Pada proses ini dilakukan perubahan kata menjadi makna dasarnya. Itu membuat mengurangi jumlah kata sehingga dapat memfasilitasi dalam kecepatan komputasi.

### 5. *Stopword Removal*

Pada proses ini dilakukan penghapusan pada kata yang tidak memiliki makna seperti kata sambung dan kata penghubung.

## II.5.4 *Support Vector Machine*

*Support Vector Machine* adalah salah satu algoritma *machine learning*. Algoritma ini ditemukan oleh Vapnik dkk, pada sekitar 1960-an. Saat ini *support vector machine* sering digunakan dan merupakan salah satu algoritma dasar yang dipelajari dalam proses pembelajaran *machine learning*. Algoritma ini pada dasarnya ditujukan untuk menemukan *hyperplane* terbaik untuk memisahkan dua kelas yang berbeda. *Hyperplane* memaksimalkan margin antara data kelas positif dan negatif terdekat dari *hyperplane*. Data terdekat dari *hyperplane* disebut vektor pendukung. Algoritma *Support Vector Machine*, dalam konteks pemrosesan bahasa alami, akan mengklasifikasikan kata, frasa, atau kalimat ke dalam kategori berdasarkan set fitur. Data yang dimasukkan akan ditransformasikan secara matematis dengan menggunakan fungsi *kernel*, di mana fungsi *kernel* ini akan menciptakan pemisahan data secara linier dari berbagai kategori (Pinem dkk., 2018).



Gambar II.3 *Hyperplane Support Vector Machine*

$$w \cdot x + b = 0$$

Keterangan:



w : parameter *hyperplane* yang dicari (garis tegak lurus garis *hyperplane* dan titik *support vector*)

x : data *input* SVM (x1 = indeks kata, x2 = bobot kata)

b : parameter *hyperplane* yang dicari (nilai bias)

## II.6 Perbandingan Algoritma *Support Vector Machine*, *Naïve Bayes* dan *Decision Tree*

Table II.1 Perbandingan Akurasi Algoritma

No	Penelitian	Akurasi		
		SVM	Naïve Bayes	Decision Tree
1.	<i>Aspect Based Sentiment Analysis using POS Tagging and TF IDF</i>	94.52	75.72	87.30
2.	<i>Improving the performance of aspect based sentiment analysis using fine-tuned Bert Base Uncased model</i>	81.33	80.12	-
3.	<i>Sentiment Analysis on Bangladesh Cricket with Support Vector Machine</i>	73.49	70.47	64.76
4.	<i>Aspect-Based Sentiment Analysis on Candidate Character Traits in Indonesian Presidential Election</i>	87.56%	81.29%	-

Berdasarkan perbandingan akurasi dari beberapa penelitian dalam Tabel II.2 yang membandingkan Algoritma SVM, NB, dan DT, disimpulkan bahwa dalam penelitian ini, penulis memilih Algoritma *Support Vector Machine* sebagai metode klasifikasi sentimen.

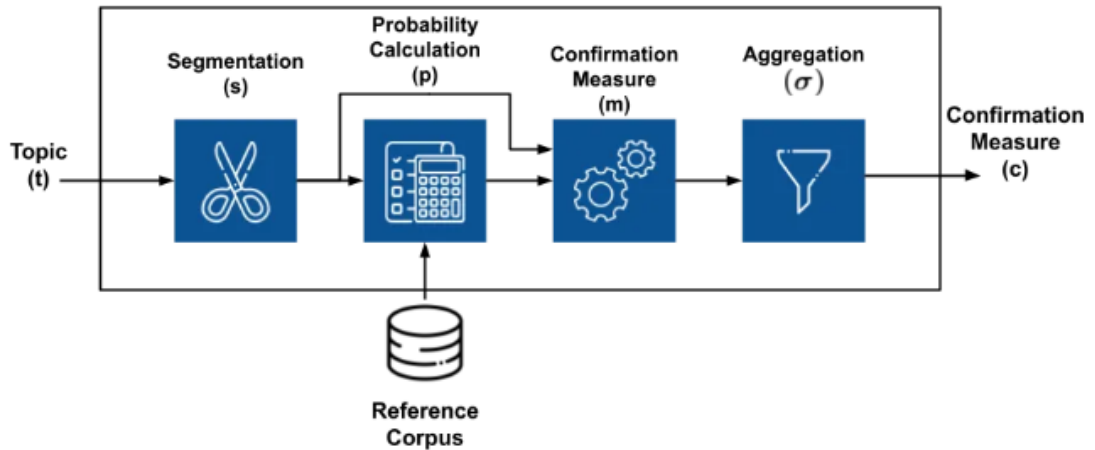
## **II.7 LDA**

*Latent Dirichlet allocation* (LDA) merupakan algoritma pemodelan topik yang paling umum digunakan, adalah model probabilistik generatif untuk koleksi data diskrit seperti kumpulan teks. LDA adalah *three-level hierarchical bayesian model*, di mana masing-masing item koleksi dimodelkan sebagai campuran terbatas atas serangkaian topik yang mendasarinya (Blei dkk., 2002). Model LDA adalah model yang paling populer dan sangat dipelajari di banyak domain, Metode LDA dapat menghasilkan serangkaian topik yang menggambarkan seluruh korpus, yang dapat dimengerti secara individual dan juga menangani korpus dokumen-kata skala besar tanpa perlu memberi label teks apa pun (Albalawi dkk., 2020). Dokumen teks berisi pola semantik tersembunyi yang disebut "topik", dan masing-masing topik ini ditentukan oleh distribusi probabilitas pada sekumpulan kata tetap. Karena LDA adalah metode tanpa pengawasan untuk pemodelan topik, ini tidak memerlukan set pelatihan, tag, atau metadata apa pun untuk pembelajaran, sehingga sejumlah besar dokumen tekstual dapat dianalisis dalam waktu singkat. Model LDA sering digunakan dalam analisis isi berdasarkan pemodelan topik (Gurcan, Ozyurt, 2022).

## **II.8 Topic Coherence**

*Topic Coherence* merupakan alat evaluasi untuk mengukur kualitas model topik dengan meneliti tingkat koherensi dari topik yang dihasilkan dalam suatu teknik pemodelan topik (Amoualian dkk., 2017). Dalam hubungan dengan *coherence*, seringkali kita membahas tentang sifat kerjasama. Sebagai contoh, suatu rangkaian argumen dianggap *coherence* apabila saling mendukung satu sama lain. Metrik *topic coherence* mengevaluasi sejauh mana sebuah topik 'didukung' dari sekumpulan teks yang disebut korpus referensi. Metrik ini mengandalkan statistik dan probabilitas yang diperoleh dari korpus referensi, terutama menitikberatkan

pada konteks kata, guna memberikan *coherence score* pada suatu topik (Pedro, 2022).



Gambar II.4 Perhitungan Coherence Score, sumber: João Pedro (2022)

Pada gambar II.4 ditampilkan struktur alur dalam perhitungan *topic coherence*, yang mencakup:

1. *Segmentation*

*Segmentation* Merupakan suatu proses untuk menentukan jenis pembagian segmen yang digunakan guna memecah rangkaian kata menjadi bagian-bagian yang lebih kecil.

2. *Probability Calculation*

*Probability Calculation* menggunakan korpus referensi, dilakukan perhitungan probabilitas kata-kata. Probabilitas-probabilitas ini menjadi landasan utama koherensi. Selanjutnya, probabilitas tersebut digunakan dalam tahap berikutnya untuk menyatukan skor dari topik yang bersangkutan.

3. *Confirmation Measure*

*Confirmation Measure* merupakan inti dalam *topic coherence*. *Confirmation measure* dihitung dengan mempertimbangkan kombinasi *subset* segmen melalui perhitungan probabilitas dalam langkah Perhitungan Probabilitas.

4. *Aggregation*

Dalam *proses aggregation*, dilakukan penggabungan seluruh nilai-nilai yang dihasilkan dari *confirmation measure* menjadi satu nilai akhir yang disebut sebagai *coherence score*.

## II.9 Bag Of Words

*Bag-of-Words* adalah metode untuk merepresentasikan teks menjadi fitur numerik (Hasanli & Rustamov, 2019). Model BoW semakin populer karena kinerjanya yang tinggi dan sederhana. Karena kesuksesan besar model BoW dalam klasifikasi, model ini telah digunakan dalam bidang pemrosesan gambar, untuk mengklasifikasikan gambar dan keperluan video. Model BoW adalah representasi yang disederhanakan yang digunakan dalam *Natural Language Processing* (NLP) dan *Information Retrieval* (IR). Model BoW sering digunakan dalam metode klasifikasi dokumen, di mana fitur untuk melatih pengklasifikasi dihasilkan dari kemunculan atau frekuensi setiap kata. Dalam model ini, teks seperti seperti kalimat, atau dokumen direpresentasikan sebagai kantong (*bag*) kata, hanya mempertimbangkan duplikat kata, dan mengabaikan tata bahasa dan urutan kata (Qader dkk., 2019).

## II.10 TF IDF

TF-IDF merupakan gabungan dari dua kata yang berbeda yaitu *Term Frequency* dan *Inverse Document Frequency*. TF digunakan untuk mengukur berapa kali sebuah kata terdapat dalam sebuah dokumen. IDF adalah metode yang digunakan untuk menghitung tingkat kepentingan kata-kata dalam sebuah dokumen. Beberapa algoritma seringkali menganggap semua kata kunci memiliki bobot yang sama, tanpa memperhatikan apakah kata tersebut adalah kata penghubung seperti "dari" yang sebenarnya tidak relevan. Namun, setiap kata kunci sebenarnya memiliki tingkat kepentingan yang berbeda (Qaiser & Ali, 2018).

### 1. *Term Frequency* (TF)

Metode untuk mengukur seberapa sering sebuah kata atau istilah tertentu muncul dalam suatu dokumen atau teks. Tujuan utama dari metode ini adalah untuk memberikan representasi numerik tentang seberapa penting kata tersebut dalam konteks dokumen tersebut. Semakin sering kata tersebut muncul dalam dokumen, semakin tinggi pula nilai *Term Frequency*-nya.

$$tf_{i,j} = \frac{n_{i,j}}{\text{Total Number of terms in document}}$$

$$tf(t_i, d_j) = f(d_j, t_i)$$

## 2. Inverse Document Frequency (IDF)

Digunakan untuk menilai seberapa penting sebuah kata dalam suatu koleksi dokumen atau korpus teks. Kata-kata yang muncul lebih jarang dalam korpus dianggap lebih penting daripada kata-kata yang muncul secara umum.

$$\begin{aligned}idf_{(w)} &= \ln \left( \frac{N}{df_t} \right) \\ &= \ln \left( \frac{N + 1}{df_t + 1} \right)\end{aligned}$$

## 3. TF-IDF

Mengevaluasi pentingnya sebuah kata dalam suatu dokumen dalam konteks kumpulan dokumen yang lebih besar. Tujuan utamanya adalah untuk mengidentifikasi kata-kata yang paling relevan atau karakteristik dari sebuah dokumen dalam kumpulan dokumen tersebut.

$$w_{i,j} = tf_{i,j} \times idf(t)$$

Keterangan:

$t_{i,j}$  : frekuensi kemunculan kata- $i$  pada dokumen ke- $j$

$n_{i,j}$  : jumlah kemunculan  $i$  pada dokumen  $j$

$N$  : total dokumen

$df_t$  : jumlah dokumen yang mengandung kata ke- $t$

$\ln$  : suatu fungsi integral yang didefinisikan sebagai  $\ln x = \int_1^x \frac{1}{x} dx$  dan semua  $x > 0$ .

$W$  : bobot dokumen ke  $i$  terhadap dokumen  $j$

$tf$  : banyaknya term  $i$  pada dokumen  $j$

$idf$  : *Inversed Document Frequency*

## II.11 Confusion Matrix

*Confusion matrix* memberikan detail tentang hasil klasifikasi suatu sistem, yang menunjukkan klasifikasi aktual dan prediksi. Ini adalah cara umum untuk menilai kinerja sistem klasifikasi menggunakan informasi yang ada dalam matriks.

Berikut ini merupakan tabel yang menunjukkan *confusion matrix* untuk pengklasifikasian dengan dua kelas (Santra & Christy, 2012).

Table II.2 *Confusion Matrix*

<i>Actual</i>	<i>Classified as</i>	
	<i>Positive</i>	<i>Negative</i>
<i>Positive</i>	<i>True Positive (TP)</i>	<i>False Negative (FN)</i>
<i>Negative</i>	<i>False Positive (FP)</i>	<i>True Negative (TN)</i>

Keterangan:

- *True Positive (TP)*: Jumlah prediksi positif yang diklasifikasikan dengan benar oleh model.
- *True Negative (TN)*: Jumlah prediksi negatif yang diklasifikasikan dengan benar oleh model.
- *False Positive (FP)*: Jumlah prediksi positif yang diklasifikasikan dengan salah oleh model.
- *False Negative (FN)*: Jumlah prediksi negatif yang salah diklasifikasikan salah oleh model.

Ada beberapa cara untuk meringkas informasi dalam *confusion matrix* (Pajankar & Joshi, 2022), diantaranya:

1. *Accuracy*

*Accuracy* adalah jumlah prediksi yang benar (TP dan TN) dibagi dengan jumlah semua sampel (semua entri dari *confusion matrix* dijumlahkan).

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

2. *Precision*

*Precision* mengukur berapa banyak sampel yang diprediksi sebagai positif sebenarnya positif. Presisi digunakan sebagai metrik kinerja ketika tujuannya adalah untuk membatasi jumlah *false positive*. Sebagai contoh, bayangkan sebuah model untuk memprediksi apakah suatu obat baru akan efektif dalam mengobati suatu penyakit dalam uji klinis. Uji klinis terkenal mahal, dan

perusahaan farmasi hanya mau melakukan percobaan jika sangat yakin bahwa obat tersebut benar-benar akan bekerja. Oleh karena itu, penting agar model tersebut tidak menghasilkan banyak kesalahan positif—dengan kata lain, model tersebut memiliki presisi yang tinggi.

$$\text{Presisi} = \frac{TP}{TP+FP}$$

### 3. *Recall*

*Recall* mengukur berapa banyak sampel positif yang ditangkap oleh prediksi positif. *Recall* digunakan sebagai metrik kinerja saat kita perlu mengidentifikasi semua sampel positif; yaitu, ketika penting untuk menghindari *false negative*.

$$\text{Recall} = \frac{TP}{TP+FN}$$

### 4. *F1-Score*

Jadi, meskipun *precision* dan *recall* adalah pengukuran yang sangat penting, hanya melihat salah satunya tidak akan memberi gambaran lengkap. Salah satu cara untuk meringkasnya adalah *f-score* atau *f-measure*, yaitu dengan rata-rata harmonik *precision* dan *recall*.

$$\text{F1-Score} = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}}$$

## II.12 *SentiStrength*

*SentiStrength* merupakan metode yang digunakan untuk menghasilkan skor yang menunjukkan kekuatan sentimen dalam sebuah teks, serta mengklasifikasikan teks tersebut ke dalam tiga kategori yaitu positif, negatif, dan netral. *SentiStrength* menerapkan metode berbasis leksikon dengan memanfaatkan aturan dan informasi linguistik tambahan (non-leksikal) untuk mengidentifikasi tingkat sentimen dalam suatu teks. *SentiStrength* akan mengeluarkan skor positif dan negatif dalam kisaran 1 hingga 5. Skor 1 menunjukkan kalimat tanpa sentimen positif atau negatif, sementara skor 5 menunjukkan kalimat dengan sentimen sangat positif atau sangat negatif (Thelwall, Buckley, 2013).

Skor akhir suatu kalimat ditentukan berdasarkan skor positif dan skor negatif tertinggi yang diperoleh dari kata-kata yang membentuk kalimat tersebut. Sebagai contoh, jika kita mengambil kalimat "Saya kesal saat login sering gagal", hasil analisis menunjukkan bahwa frasa "Saya kesal" mendapatkan skor sentimen negatif -4, sementara kata "gagal" mendapatkan skor sentimen negatif -5. Jadi, secara keseluruhan, kalimat tersebut memiliki skor akhir <skor: 1, -5>. Angka dalam tanda "[...]" menunjukkan skor sentimen dari setiap kata dalam kalimat, dan angka dalam tanda "<skor: ..., ...>" menunjukkan skor akhir kalimat. Untuk menentukan sentimen akhir dari kalimat tersebut, maka diterapkan aturan berikut (Wahid, 2017):

- Jika nilai positif > nilai negatif, maka sentimen adalah positif.
- Jika nilai positif < nilai negatif, maka sentimen adalah negatif.
- Jika nilai positif = nilai negatif, maka sentimen adalah netral.

## **II.13 Penelitian Terdahulu**

### **II.13.1 *Effective Comparison of LDA with LSA for Topic Modelling***

Proses mengubah data yang tidak terstruktur menjadi format terstruktur yang dapat dibaca menjadi sulit dari hari ke hari. Hingga hari ini, setiap organisasi memiliki lebih dari 80% data operasionalnya dalam format yang tidak dapat dibaca. Metode yang diusulkan membantu dalam mengubah data yang tidak dapat dibaca menjadi format terstruktur yang dapat dibaca dengan bantuan *machine learning* adalah klasifikasi, dan pengelompokan memainkan peran penting dalam mengubah data operasional menjadi model data dan memvisualisasikan informasi yang diproses kepada pengguna akhir. Karena organisasi memiliki persyaratan khusus, dengan mempertimbangkannya, kami akan menerapkan *Latent Dirichlet Allocation* (LDA) dan *Latent Semantic Analysis* (LSA), yang mampu menangani data diskrit. Selain itu, dilakukan perbandingan untuk menguji divergensi, throughput, kualitas, dan waktu respon, karena keduanya dapat mengklasifikasikan data berdasarkan konten dan dengan memberi label pada masing-masing kategori. Algoritma dengan divergensi yang lebih baik diimplementasikan yang dapat menangani kebutuhan organisasi dengan menghadirkan area teratas yang perlu ditingkatkan/dikonsentrasikan tergantung



pada analitik yang dibuat oleh algoritma pada data diskrit yang tersedia, dan dengan menerapkan teknik visualisasi, hasilnya akan ditampilkan secara merata dalam format grafis (Kalepalli, 2020).

Penelitian ini ditulis oleh Yaswanth Kalepalli, Pasupuleti Durga Phani Teja, Shaik Tasneem dan Dr. Suneetha Manne.

### **II.13.2 *Aspect-Based Sentiment Analysis in Beauty Product Reviews Using TF-IDF and SVM Algorithm***

Ulasan produk sangat penting dalam *e-commerce* karena dapat membantu calon pembeli membuat keputusan sebelum melakukan pembelian dan membantu penjual mengukur produk mereka. Sebuah produk bisa memiliki ribuan ulasan, sehingga menyulitkan calon pembeli dan penjual untuk menarik kesimpulan dari ulasan yang melimpah tersebut. Penelitian ini membangun sistem yang menerapkan *Aspect-based Sentiment Analysis* (ABSA) dengan *dataset* dari *review* produk di *website Female Daily*. Sistem ini dibangun menggunakan TF-IDF sebagai kombinasi metode ekstraksi fitur dengan kata bigram dan kata bigram. Algoritma *Support Vector Machine* (SVM) digunakan untuk mengklasifikasikan sentimen. Hasil percobaan ini menunjukkan bahwa tahap *preprocessing*, khususnya proses *stemming* dan *stopwords removal* sangat mempengaruhi hasil akurasi. Pemilihan kata N-gram juga sangat menentukan, di mana penelitian ini menunjukkan bahwa kata unigram memberikan akurasi yang lebih tinggi daripada kata bigram. Hasil akhir dari penelitian ini menunjukkan bahwa kombinasi TF-IDF dengan *word unigram* dan SVM dengan *kernel linier* menghasilkan akurasi terbaik, yaitu 88,35% (Arthamevia, 2014).

Penelitian ini ditulis oleh Nadira Putri Arthamevia, Adiwijaya, dan Mahendra Dwifebri Purbolaksono.

### **II.13.3 *Impact of Digital Technostress and Digital Technology Self-Efficacy on Fintech Usage Intention of Chinese Gen Z Consumers***

Studi ini menyelidiki hubungan antara empat subdimensi *technostress* (kompleksitas, kelebihan beban, invasi, dan ketidakpastian), efikasi diri teknologi digital, dan niat penggunaan tekfin. Data dari total 266 konsumen Gen Z Tiongkok

digunakan dalam analisis regresi berganda. Hasil penelitian secara umum mendukung bahwa semua sub-dimensi *technostress* berhubungan negatif dengan niat penggunaan tekfin. Terkait dengan efek moderasi *self-efficacy* teknologi digital pada hubungan antara empat sub-dimensi *technostress* dan niat penggunaan *fintech*, ditemukan efek interaksi yang signifikan dengan kompleksitas dan kelebihan beban. Akhirnya, penelitian ini membahas implikasi teoritis dan manajerial dari temuan penelitian (Lee, 2021).

Penelitian ini ditulis oleh You-Kyung Lee.

#### **II.13.4 A Comparative Analysis of Social Communication Applications Using Aspect Based Sentiment Analysis 2022**

Studi ini bertujuan untuk mengembangkan model berbasis pembelajaran mendalam untuk menilai ulasan secara otomatis yang dapat diadaptasi untuk analisis data di masa mendatang. Kami mengusulkan metodologi pembelajaran mendalam dengan menggunakan *Aspect Based Sentiment Analysis* (ABSA) memanfaatkan ulasan aplikasi komunikasi yang diambil dari Google Play Store menggunakan aplikasi *scraper* Google Play. Penelitian ini menggunakan teknik text mining untuk ABSA pada *review* pengguna. Untuk ekstraksi Topik, kami menggunakan *Latent Dirichlet Allocation* (LDA) dan metode pembelajaran mendalam *Long Short-Term Memory* (LSTM) untuk klasifikasi topik. Hasilnya menunjukkan bahwa model yang kami usulkan memberikan hasil yang menjanjikan dengan akurasi 90% dengan menggunakan model LSTM (Irfan dkk., 2022).

Penelitian ini ditulis oleh Laiba Irfan, Shabir Hussain, Muhammad Ayoub, Yang Yu, dan Akmal Khan.

#### **II.13.5 Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM**

Penelitian ini melakukan analisis sentimen berbasis aspek terhadap ulasan para pemakai aplikasi KAI Access menggunakan metode *Naïve Bayes* dan *Support Vector Machine* (SVM), menggunakan 3 skenario yang berbeda, Skenario 1 dengan menggunakan *Multinomial Naive Bayes*, skenario ke 2 dengan

menggunakan SVM dengan parameter yaitu default *library Sklearn*, dan skenario ke 3, dengan SVM dan *hyperparameter tuning*, dan data yang diambil bersumber dari Google Play Store. Hasil dari penelitian ini menunjukkan sebagian besar sentimen pengguna adalah negatif pada setiap aspek, dan yang terbanyak dibahas yaitu aspek *errors* menunjukkan bahwa terdapat kesalahan atau gangguan sistem yang cukup tinggi pada *KAI Access*. Hasil pengujian menghasilkan model terbaik dari skenario ke 3 dengan hasil rata-rata skor akurasi 91,63%, *f1-score* 75,55%, presisi 77,60%, dan recall 74,47% (Mustakim & Priyanta, 2022).

Penelitian ini dilakukan oleh Huda Mustakim dan Sigit Priyanta.

#### **II.13.6 Analisis Sentimen Berbasis Aspek Pada Aplikasi Tokopedia Menggunakan LDA Dan *Naïve Bayes***

Penelitian yang dilakukan oleh (Astuti, 2020) fokus pada klasifikasi sentimen dari ulasan-ulasan Tokopedia menggunakan metode *Naïve Bayes*. Jumlah aspek ditentukan melalui proses *clustering* topik LDA, yang menghasilkan empat topik utama: kebermanfaatan, pelayanan, pengalaman belanja, dan tampilan. Proses *clustering* ini menjadi acuan manual dalam anotasi aspek. Evaluasi penelitian ini menggunakan kurva ROC dan AUC untuk mengukur kinerja model. Setelah dilakukan *oversampling*, akurasi tertinggi yang berhasil dicapai mencapai 92,5%, yang menunjukkan peningkatan yang signifikan. Selain itu, nilai AUC yang diperoleh juga sangat baik, yaitu sebesar 0,95.

Penelitian ini dilakukan oleh Shinta Prima Astuti

Table II.3 Penelitian Terdahulu

No	Nama Penelitian	Jenis Penelitian	Pembahasan	Output
1	<i>Effective Comparison of LDA with LSA for Topic Modelling</i>	Jurnal Internasional	Penelitian ini membahas mengenai perbandingan LDA dan LSA pada klasifikasi yang dilakukan berdasarkan konten lalu pemberian label pada setiap aspek atau kategori	LSA ditemukan memiliki divergensi yang lebih sedikit setelah 250 iterasi dengan pembesaran awal adalah 75.300209 dan divergensi di sini setelah iterasi 2000 adalah 1.235335 karena membutuhkan kumpulan dokumen dan kosa kata yang sangat besar untuk mendapatkan hasil yang akurat. Sedangkan LDA ditemukan memiliki divergensi yang lebih baik daripada LSA setelah 250 iterasi dengan pembesaran awal di 82.572319 dan setelah iterasi 2000 adalah 1.224222 yang disarankan untuk kumpulan dokumen yang lebih kecil.
2	<i>Aspect-Based Sentiment Analysis in Beauty Product Reviews Using TF-IDF and SVM Algorithm</i>	Jurnal Internasional	Penelitian ini membahas mengenai <i>Aspect-based Sentiment Analysis (ABSA)</i> dengan <i>dataset</i> dari review produk di <i>website Female Daily</i> . Sistem ini dibangun	Hasilnya TF-IDF dengan <i>word unigram</i> dan SVM dengan <i>linear kernel</i> pada dataset dengan word stemming dan tanpa stopwords removal menghasilkan

No	Nama Penelitian	Jenis Penelitian	Pembahasan	Output
			menggunakan TF-IDF sebagai kombinasi metode ekstraksi fitur dengan kata bigram dan kata bigram. Algoritma <i>Support Vector Machine</i> (SVM) digunakan untuk mengklasifikasikan sentimen.	akurasi terbaik yaitu 88,35%.
3	<i>Impact of Digital Technostress and Digital Technology Self-Efficacy on Fintech Usage Intention of Chinese Gen Z Consumers</i>	Jurnal International	Penelitian ini membahas hubungan antara empat subdimensi <i>technostress</i> (kompleksitas, kelebihan beban, invasi, dan ketidakpastian), efikasi diri teknologi digital, dan niat penggunaan <i>fintech</i> .	Hasilnya pemasar <i>fintech</i> harus mengembangkan materi berbasis media, seperti gambar, animasi, atau video, agar konsumen dapat lebih mudah dan cepat memahami fitur teknologi digital baru dan cara menggunakannya, dengan mempertimbangkan ciri perilaku konsumen Gen Z. Kedua, pemasar <i>fintech</i> harus menghadirkan norma dan peraturan yang lebih tinggi untuk masalah privasi dan keamanan pribadi. Ketiga, penting untuk berhati-hati untuk tidak secara langsung memaparkan konsumen pada pembaruan atau

No	Nama Penelitian	Jenis Penelitian	Pembahasan	Output
				perubahan teknologi digital yang terlalu sering dan untuk menetapkan strategi pemasaran yang lebih cermat untuk mengurangi peningkatan beban kognitif dan emosional konsumen akibat teknologi digital.
4	<i>A Comparative Analysis of Social Communication Applications Using Aspect Based Sentiment Analysis 2022</i>	Jurnal Nasional	Penelitian ini mengembangkan model metodologi deep learning dengan menggunakan <i>Aspect-based Sentiment Analysis (ABSA)</i> dengan memanfaatkan ulasan aplikasi komunikasi yang diambil dari Google Play Store menggunakan aplikasi <i>Google Play Scraper</i> . Penelitian ini menggunakan teknik text mining untuk ABSA pada ulasan pengguna. Untuk ekstraksi Topik, kami menggunakan <i>Latent Dirichlet Allocation (LDA)</i> dan metode pembelajaran mendalam <i>Long Short-</i>	Hasilnya menunjukkan bahwa model memberikan hasil yang menjanjikan dengan akurasi 90% dengan menggunakan model LSTM. Otoritas WhatsApp dapat menggunakan hasilnya untuk mengoptimalkan aplikasi komunikasi dengan menambahkan fitur yang lebih efisien dan memperbaruinya

No	Nama Penelitian	Jenis Penelitian	Pembahasan	Output
			<i>Term Memory</i> (LSTM) untuk klasifikasi topik.	
5	<i>Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM</i>	Jurnal Nasional	Penelitian ini melakukan analisis sentimen berbasis aspek terhadap ulasan para pemakai aplikasi KAI Access menggunakan metode <i>Naïve Bayes</i> dan <i>Support Vector Machine</i> (SVM), menggunakan 3 skenario yang berbeda, Skenario 1 dengan menggunakan <i>Multinomial Naïve Bayes</i> , skenario ke 2 dengan menggunakan SVM dengan parameter yaitu default <i>library Sklearn</i> , dan skenario ke 3, dengan SVM dan <i>hyperparameter tuning</i> , dan data yang diambil bersumber dari Google Play Store.	Sebagian besar sentimen pengguna adalah negatif pada setiap aspek, dan yang terbanyak dibahas yaitu aspek errors menunjukkan bahwa terdapat kesalahan atau gangguan sistem yang cukup tinggi pada KAI Access. Hasil pengujian menghasilkan model terbaik dari skenario ke 3 dengan hasil rata-rata skor akurasi 91,63%, <i>f1-score</i> 75,55%, presisi 77,60%, dan <i>recall</i> 74,47%
6	Analisis Sentimen Berbasis Aspek Pada Aplikasi Tokopedia Menggunakan LDA Dan <i>Naïve Bayes</i>	Jurnal Nasional	Studi ini melakukan analisis sentimen berbasis aspek pada ulasan Tokopedia dengan memanfaatkan metode <i>Naïve Bayes</i> . Jumlah aspek ditentukan dengan menggunakan teknik	Setelah melakukan oversampling, dilakukan evaluasi dengan menggunakan kurva ROC dan AUC. Didapatkan akurasi tertinggi sebesar 92,5%, dengan nilai AUC sebesar 0,95.

No	Nama Penelitian	Jenis Penelitian	Pembahasan	Output
			<p>clustering topik LDA yang menghasilkan 4 topik. Untuk mengatasi ketidakseimbangan data, dilakukan resampling menggunakan tiga teknik berbeda, yaitu <i>RandomUnderSampler</i>, <i>RandomOverSampler</i>, dan SMOTEENN.</p>	

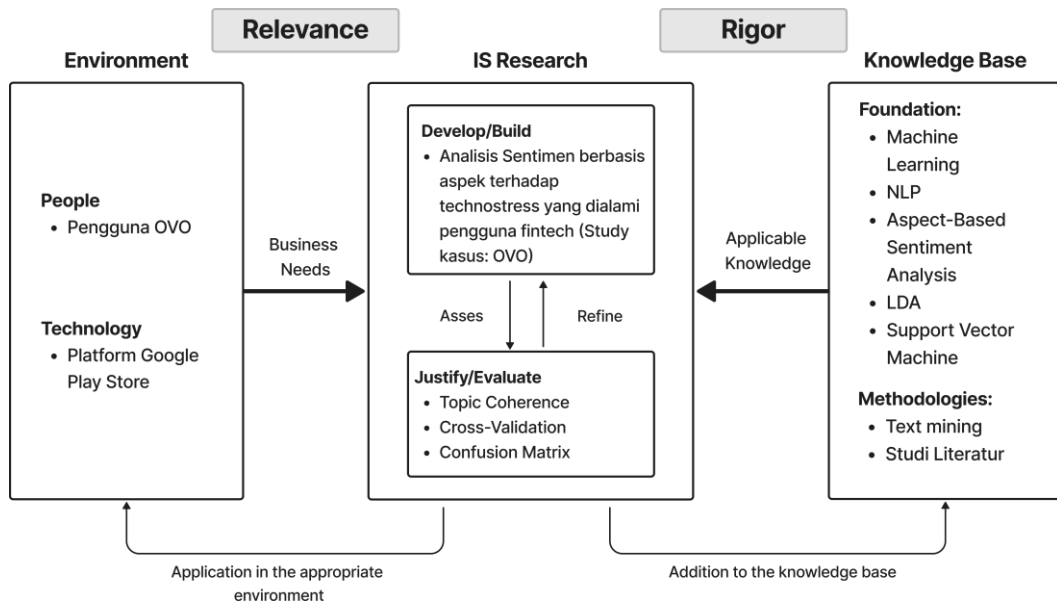


## BAB III

### METODOLOGI PENELITIAN

#### III.1 Kerangka Berpikir

Penelitian ini menggunakan model konseptual yang diadopsi dari *Design Science in Information System Research* (Hevner, 2004).

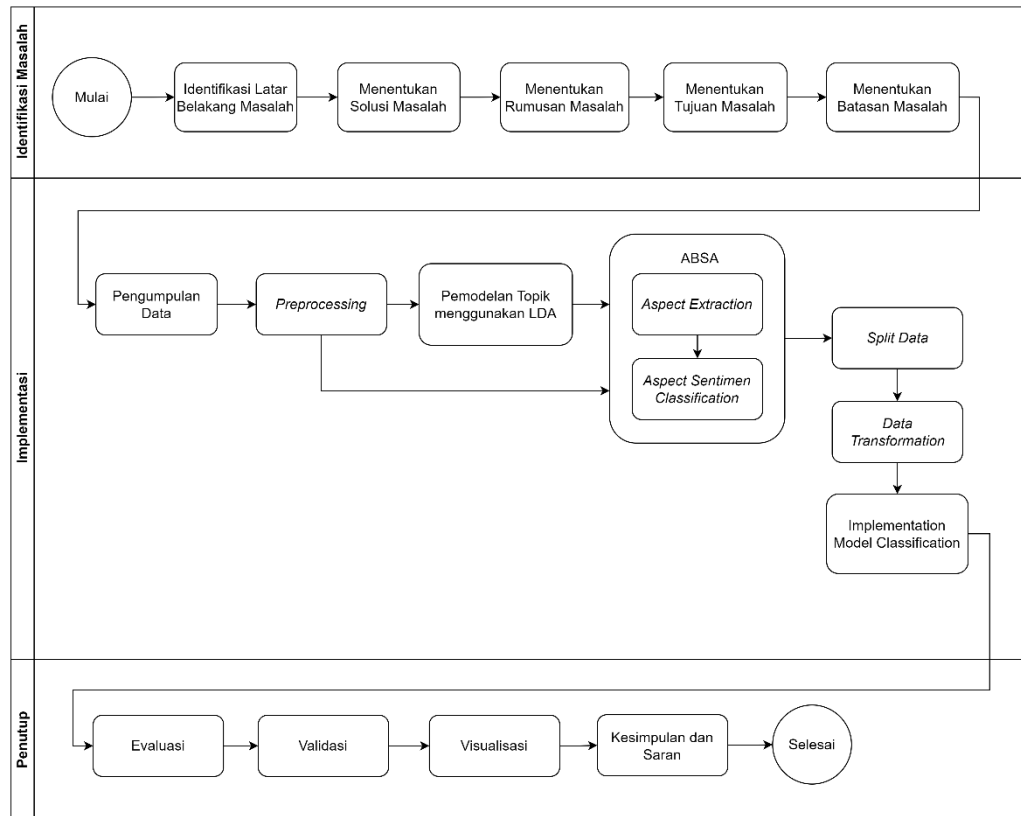


Gambar III.1 Model Konseptual

Gambar III.1 merupakan model konseptual yaitu representasi visual dari konsep yang saling terkait yang digunakan untuk memetakan persoalan pada penelitian ke dalam *framework* pemikiran dari kerangka teori pada bab 2, juga sebagai dasar pengembangan teori, alat komunikasi, membantu dalam memecahkan masalah, dan membantu memahami konsep-konsep yang kompleks. Penelitian ini berfokus pada pengguna OVO pada bagian people dan Google Play Store sebagai *technology*. Adapun penelitian ini mengkaji sentimen pengguna untuk mendapatkan aspek apa saja yang menyebabkan *technostress* pada pengguna OVO. Metodologi penelitian melibatkan berbagai teknik termasuk *Machine Learning*, *Natural Language Processing*, *Aspect-Based Sentiment Analysis*, LDA, dan *Support Vector Machine*.

### III.2 Sistematika Penyelesaian Masalah

Terdapat tiga tahap pada sistematika penyelesaian masalah pada penelitian ini yaitu tahap identifikasi masalah, tahap implementasi, dan tahap penutup. Gambar III.2 menggambarkan sistematika penyelesaian masalah pada penelitian ini.



Gambar III.2 Alur Sistematika Penyelesaian Masalah

#### III.2.1 Tahap Identifikasi Masalah

Tahap yang dilakukan pertama yaitu tahap identifikasi masalah, pada tahap ini dilakukan identifikasi masalah yang terjadi pada produk *fintech* yaitu dompet digital. Lalu pengumpulan studi literatur sebagai dasar penentuan metode yang tepat. Kemudian menentukan tujuan serta batasan dalam penelitian ini.

#### III.2.2 Tahap Implementasi

Tahap yang kedua yaitu tahap implementasi yang dilakukan dengan mengambil data dari review OVO dari aplikasi Google Play Store, *text preprocessing* untuk mempersiapkan data sehingga dapat diolah, *topik modeling* menggunakan LDA untuk menentukan topik-topik yang terdapat pada data. Selanjutnya dilakukan

*aspect extraction* dengan tujuan menentukan aspek terhadap tiap ulasan, lalu dilakukan pelabelan data dengan memberikan label sentimen. Kemudian menggunakan metode TF-IDF untuk mengukur bobot atau pentingnya kata-kata pada teks. Selanjutnya melakukan pembagian data dengan beberapa rasio untuk pelatihan serta pengujian model. Yang terakhir dilakukan klasifikasi menggunakan algoritma *Support Vector Machine*.

### III.2.3 Tahap Penutup

Tahap yang terakhir yaitu tahap Kesimpulan dan Saran, pada tahap ini yang dilakukan yaitu proses analisis terhadap hasil dan setelah dianalisis maka akan dilakukan penarikan kesimpulan dari hasil yang didapatkan dari penelitian ini.

### III.3 Pengumpulan Data

Penelitian ini menggunakan teknik *scraping* untuk melakukan pengumpulan data yang memanfaatkan *library google play scraper*. Prosesnya menggunakan Python. Data diambil pada rentang waktu 2018 hingga 2023 dan didapatkan 100.000 ulasan, sebanyak 33.120 ulasan dipilih untuk mengambil ulsan pada tanggal 1 Januari 2023 hingga 14 Maret 2023.

	A	B	C	D	E	F	G	H	I	J	K	L
		userName	score	at	content							
1												
2	2024	ahmad siswanto	1	2023-03-14 04:13:39	Saya menyesal pernah download ovo aplikasi sangat merugikan saya							
3	2093	safarimitrasukses	5	2023-03-13 13:09:55	mudah sekali dan simple							
4	81914	danu wihara purba	1	2023-03-13 13:01:28	tolong dong servernya di perbaiki lagi, ini ovo masih banyak kendalanya, banyak pengguna ovo yang mengeluh, apa server pusat ovo keceal aj							
5	3161	Dzaky lucu	5	2023-03-13 12:57:26	Mudah dan praktis....							
6	64943	Faliz Almunawar25	5	2023-03-13 12:56:52	Darderdor mantap banget							
7	529	Ahmad Samsi	1	2023-03-13 12:52:20	Aplikasi ovo ini pembohong setelah AQ mendapatkan 1juta 500. Eh malah apk ini tidak dapat di buka . Setelah qu instal ulang apk ovo.lalu qu i							
8	38182	Riyan Ramdani	1	2023-03-13 12:51:06	ERROR TERUSS!!! GABISA TRANSFER KE SESAMA OVO ATAU PUN QRIS MOHON UNTUK PERBAIKI PIHAK MONTOON							
9	22769	Ridwan Cahyadi	1	2023-03-13 12:49:52	Lama banget kalo isi pulsa pake ovo hadeh							
10	10461	Rekustra Chanel	2	2023-03-13 12:48:38	Semua nya udh bagus, hanya aja, masa kirim ke sesama ovo ada admin nya juga, bingung							
11	13462	Sabilla Amalia	2	2023-03-13 12:45:21	Sumpah w kira ivo membantu kaya merek sebelah tapi ternyata tidak w udah ngisi berulang kali di ovo tapi giliran w mau keke ke grab gk bisa2							
12	29476	Dinni Ariska	5	2023-03-13 12:41:03	bayar pake ovo jadi lebih mudah							
13	60257	Achmad Rizqi fathor	1	2023-03-13 12:34:02	Aplikasi lemot proses nya kelamaan							
14	2657	Bung Kus	5	2023-03-13 12:20:43	Lumayan gampang dan membantu							
15	64760	Somad Somadun (Pa	5	2023-03-13 12:16:58	OVO Oke dan simpel							
16	520	Adi Sorjaya	5	2023-03-13 11:57:51	Sangat membantu sekali...semoga kedepan nya lebih maju dan bermanfaat dan selu memberikan kemudahan transaksi							
17	611	Selly Valentina	3	2023-03-13 11:42:21	Akhir" ini pembelian voucher grab sangat lama,gak kayak dulu lagi 10 detik selesai.kasih 3 star dulu							
18	2331	Fadli Muzaki	1	2023-03-13 11:38:10	Saldo sering hangus pas ngetransfer ke nomer laen padahal saldo berkurang							
19	13463	Atifa Raihana	5	2023-03-13 11:35:35	Pake ovo itu gampang dan menyenangkan							
20	33948	Bayu Pitono	5	2023-03-13 11:31:04	Transaksi jdi lepas cepat dan mudah							
21	4181	Kucing TerbangKelai	5	2023-03-13 11:29:26	Pemakaian masi enak. Dan dah di dukung juga pilihan isi saldo melalui alfamart maupun Indomaret. Pernah sebelumnya bisa isi saldo di alfam							

Gambar III.3 Hasil *Scraping*

Berdasarkan tabel III.1 Beberapa atribut terpenting diambil seperti atribut *username* untuk mengetahui pemberi ulasan terhadap OVO, atribut tanggal untuk mengetahui kapan ulasan dibuat dan atribut *content* yang berisi ulasan pengguna.

### III.4 Pengolahan Data

Di tahap pengolahan data ini dilakukan proses *text pre-processing* untuk mengelola isi teks menjadi lebih siap dan bersih sehingga siap untuk diolah. Kemudian peneliti akan melakukan proses pemodelan topik untuk menentukan aspek-aspek yang terdapat pada teks untuk dilakukan analisa. Setelah itu, peneliti melakukan topic modelling menggunakan metode LDA (*Latent Dirichlet Allocation*) guna mengidentifikasi aspek-aspek yang relevan. Dengan memanfaatkan metode evaluasi *Topic Coherence*, peneliti dapat memperoleh nilai *coherence score* tertinggi untuk menentukan jumlah topik yang optimal. Selanjutnya dilakukan *aspect extraction* untuk tujuan menentukan aspek terhadap tiap ulasan, lalu dilakukan pelabelan data dengan memberikan label sentimen. Kemudian dilakukan proses TF-IDF untuk mengukur bobot atau pentingnya kata-kata pada teks. Kemudian dilakukan pembagian data menjadi data uji dan data latih dengan beberapa skenario skala yaitu 20:80, 25:75, dan 30:70 yang diambil dari hasil penelitian oleh (Zahoor dkk., 2020) untuk mendapatkan hasil akurasi model terbaik. Selanjutnya dilakukan pemodelan dengan algoritma yang dipilih yaitu *Support Vector Machine*.

### III.5 Metode Evaluasi

Tahap yang dilakukan selanjutnya yaitu melakukan evaluasi. Metode evaluasi yang dipakai pada penelitian ini yaitu *Confusion Matrix* untuk menganalisis hasil prediksi suatu model. Ia memberikan gambaran tentang seberapa baik model tersebut dalam mengklasifikasikan data. Memberikan hasil performa yang berisi *True Positive (TP)*, *True Negative (TN)*, *False Positive (FP)*, dan *False Negative (FN)* yang dapat menghitung metrik evaluasi lainnya seperti akurasi, presisi, *recall*, dan *F1-score*, yang memberikan informasi lebih rinci tentang kinerja model dalam mengklasifikasikan data. Selain itu, dalam penelitian ini, dilakukan validasi pengujian *overfitting* dengan melihat bias dan varians antara data *test* dan data *train* pada model.

## BAB IV

### ANALISIS DAN PERANCANGAN

#### **IV.1 Text Preprocessing**

*Text preprocessing* adalah langkah-langkah yang dilakukan pada teks sebelum diproses lebih lanjut, dengan tujuan membersihkan, mengubah format, dan mempersiapkannya untuk analisis atau pemrosesan lebih lanjut. Dengan menerapkan teknik-teknik ini dalam *text preprocessing*, teks dapat dibersihkan dan dipersiapkan dengan lebih baik sebelum digunakan dalam analisis atau pemrosesan lebih lanjut.

##### **IV.1.1 Pemisahan Kalimat**

Pada proses ini dilakukan pemisahan kalimat pada ulasan yang dipisahkan oleh tanda baca titik serta konjungsi (“namun”, “tapi”, dan “tetapi”). Dengan tujuan agar setiap ulasan dapat diketahui jika terdapat dua aspek yang berbeda. Setelah dipisah maka kalimat yang dipisah akan menjadi ulasan yang baru atau tidak bersatu dengan ulasan sebelumnya. Setelah proses pemisahan kalimat maka jumlah ulasan bertambah menjadi 47467 ulasan.

##### **IV.1.2 Case Folding**

Dalam proses ini, dilakukan pengubahan semua huruf dalam teks menjadi huruf kecil untuk menghindari perbedaan dalam pengenalan kata yang memiliki kapitalisasi yang berbeda.

*Table IV.1 Case Folding*

<b>Sebelum</b>	<b>Sesudah</b>
Ovo Kenapa Sih Kalau Beli Saldo Pasti Gak 100% saldo masuk pasti Sll Kena Potongan 20, Kenapa Gak Adminnya, Pas Pembelian Isi Saldo Aja, Biar Masuknya 100%, isi 100.0 , Masuk 98.0 buat dikirim Transaksi lain, Jadi Gak Bisa 100.0 , Begitu Juga Isi 50.0 yg masuk 48.0 , Kaya gopay linkaja 100% Saldo masuk	ovo kenapa sih kalau beli saldo pasti gak 100% saldo masuk pasti sll kena potongan 20, kenapa gak adminnya, pas pembelian isi saldo aja, biar masuknya 100%, isi 100.0 , masuk 98.0 buat dikirim transaksi lain, jadi gak bisa 100.0 , begitu juga isi 50.0 yg masuk 48.0 , kaya gopay linkaja 100% saldo masuk

#### IV.1.3 Removal of Punctuation

Proses ini melibatkan penghapusan tanda baca seperti titik, koma, tanda tanya, dan tanda baca lainnya dari teks. Tanda baca biasanya tidak memberikan informasi tambahan yang relevan dalam pemrosesan teks, sehingga dihilangkan untuk menyederhanakan teks.

Table IV.2 Removal of Punctuation

Sebelum	Sesudah
ovo kenapa sih kalau beli saldo pasti gak 100% saldo masuk pasti sll kena potongan 20, kenapa gak adminnya, pas pembelian isi saldo aja, biar masuknya 100%, isi 100.0 , masuk 98.0 buat dikirim transaksi lain, jadi gak bisa 100.0 , begitu juga isi 50.0 yg masuk 48.0 , kaya gopay linkaja 100% saldo masuk	ovo kenapa sih kalau beli saldo pasti gak 100 saldo masuk pasti sll kena potongan 20 kenapa gak adminnya pas pembelian isi saldo aja biar masuknya 100 isi 100 0 masuk 98 0 buat dikirim transaksi lain jadi gak bisa 100 0 begitu juga isi 50 0 yg masuk 48 0 kaya gopay linkaja 100 saldo masuk

#### IV.1.4 Removal of Number

Proses ini melibatkan penghapusan angka dari teks. Karena angka seringkali tidak relevan dalam analisis teks tertentu.

Table IV.3 Removal of Number

Sebelum	Sesudah
ovo kenapa sih kalau beli saldo pasti gak 100 saldo masuk pasti sll kena potongan 20 kenapa gak adminnya pas pembelian isi saldo aja biar masuknya 100 isi 100 0 masuk 98 0 buat dikirim transaksi lain jadi gak bisa 100 0 begitu juga isi 50 0 yg masuk 48 0 kaya gopay linkaja 100 saldo masuk	ovo kenapa sih kalau beli saldo pasti gak saldo masuk pasti sll kena potongan kenapa gak adminnya pas pembelian isi saldo aja biar masuknya isi masuk buat dikirim transaksi lain jadi gak bisa begitu juga isi yg masuk kaya gopay linkaja saldo masuk

#### IV.1.5 Spelling Correction

Proses ini mengoreksi kata-kata yang salah ketik atau salah eja dalam teks sehingga kualitas teks dapat ditingkatkan.

Table IV.4 Spelling Correction

Sebelum	Sesudah
ovo kenapa sih kalau beli saldo pasti gak saldo masuk pasti sll kena potongan kenapa gak adminnya pas pembelian isi saldo aja biar masuknya isi masuk buat dikirim transaksi lain jadi gak bisa begitu juga isi yg masuk kaya gopay linkaja saldo masuk	ovo kenapa sih kalau beli saldo pasti tidak saldo masuk pasti selalu kena potongan kenapa tidak adminnya pas pembelian isi saldo saja biar masuknya isi masuk buat dikirim transaksi lain jadi tidak bisa begitu juga isi yang masuk kayak gopay linkaja saldo masuk

#### IV.1.6 Stemming

Proses ini mereduksi kata-kata ke bentuk dasarnya (kata dasar atau kata akar) dengan menghapus imbuhan atau akhiran kata. Tujuan dari stemming adalah untuk mengatasi variasi kata yang memiliki akar kata yang sama, sehingga kata-kata tersebut dapat dianggap sebagai satu entitas dalam analisis.

Table IV.5 Stemming

Sebelum	Sesudah
ovo kenapa sih kalau beli saldo pasti tidak saldo masuk pasti selalu kena potongan kenapa tidak adminnya pas pembelian isi saldo saja biar masuknya isi masuk buat dikirim transaksi lain jadi tidak bisa begitu juga isi yang masuk kayak gopay linkaja saldo masuk	ovo kenapa sih kalau beli saldo pasti tidak saldo masuk pasti selalu kena potong kenapa tidak adminnya pas beli isi saldo saja biar masuk isi masuk buat kirim transaksi lain jadi tidak bisa begitu juga isi yang masuk kayak gopay linkaja saldo masuk

#### IV.1.7 Stopword Removal

Proses ini melibatkan penghapusan kata-kata yang tidak umum dan tidak memiliki makna khusus yang tidak memberikan informasi penting dalam analisis dan dapat dihilangkan untuk mengurangi dimensi dan meningkatkan efisiensi pemrosesan.

Table IV.6 Stopword Removal

Sebelum	Sesudah
ovo kenapa sih kalau beli saldo pasti tidak saldo masuk pasti selalu kena potong kenapa tidak adminnya pas beli isi saldo saja biar masuk isi masuk buat kirim transaksi lain jadi tidak bisa begitu juga isi yang masuk kayak gopay linkaja saldo masuk	ovo kalau beli saldo pasti saldo masuk selalu kena potong adminnya beli isi saldo biar masuk isi masuk buat kirim transaksi bisa isi masuk gopay linkaja saldo masuk

#### IV.1.8 Null Value Removal

Proses ini menghapus data yang memiliki nilai “*null*”, Nilai *null* biasanya muncul ketika tidak ada data yang tersedia untuk atribut tertentu. Proses penghapusan nilai *null* penting dalam analisis data karena nilai *null* dapat mengganggu analisis statistik dan pemodelan yang akurat.

### IV.2 Pemodelan Topik

Proses pemodelan topik dilakukan setelah proses *preprocessing*. Dengan menggunakan algoritma LDA dengan tujuan untuk mendapatkan topik-topik yang terkandung pada data ulasan yang ada sebagai representasi.

#### IV.2.1 Pembobotan Kata

Pada pemodelan topik, perlu dilakukan tahap awal yaitu pembobotan kata yang melibatkan pembuatan *dictionary* yang berisi kumpulan kata-kata unik yang terdapat pada dalam korpus. *Dictionary* atau kamus ini berperan sebagai basis atau referensi untuk merepresentasikan dokumen-dokumen dalam bentuk vektor kata-kata. Lalu setiap dokumen dijadikan representasi vektor, dan juga untuk mengubah dari teks ke matriks numerik. Maka digunakan Metode *Bag of Words* (BoW) mengubah setiap dokumen dalam korpus menjadi vektor yang



mencerminkan frekuensi kata-kata dalam dokumen tersebut. Setelah siap, data tersebut dapat digunakan pada penerapan algoritma LDA (*Latent Dirichlet Allocation*).

#### IV.2.2 Pembuatan Model LDA

Selanjutnya dilakukan tahap pembuatan topic model untuk mengetahui topik yang relevan dengan dokumen. Proses dilakukan dengan menggunakan library **gensim.models.LdaMulticore**, model ini menggunakan pendekatan paralel serta digunakan core CPU untuk meningkatkan kecepatan proses pemodelan. Model LDA dilatih menggunakan data teks yang telah direpresentasikan oleh BoW dengan menggunakan korpus dan *dictionary* pada proses sebelumnya. Terdapat parameter lain yang diberikan pada proses penginisialisasi model seperti jumlah topik, iterasi dan pekerja (*workers*). Model ini dapat menghasilkan distribusi topik yang optimal pada tiap dokumen dalam korpus. Pada model ini, iterasi dilakukan sebanyak 100 iterasi yang sesuai dengan standar pada iterasi model LDA. Berikut adalah contoh hasil topik dengan jumlah topik 5 serta jumlah kata pada topik 10 dengan 100 iterasi.

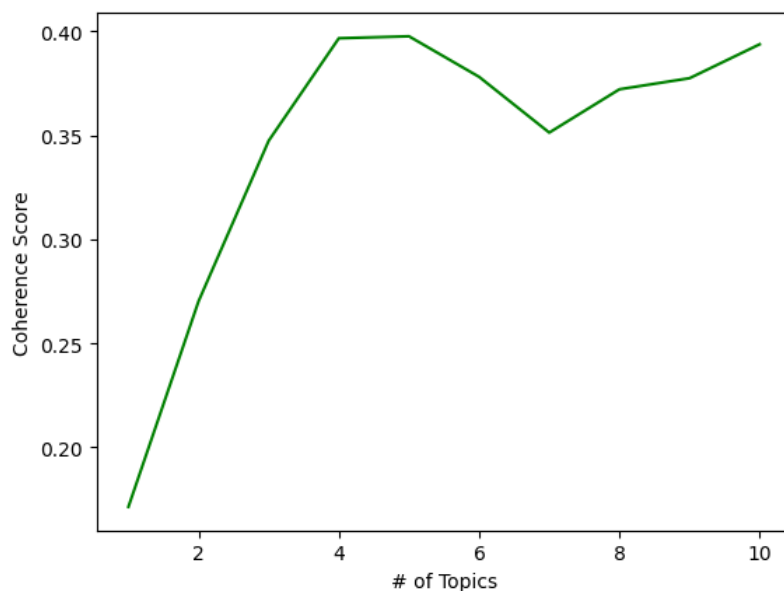
Table IV.7 Hasil Topik

Topic	Probabilitas Kata
0	0.108*"saldo" + 0.050*"transfer" + 0.049*"beli" + 0.025*"kecewa" + 0.025*"bayar" + 0.018*"promo" + 0.017*"token" + 0.017*"topup" + 0.015*"cashback" + 0.014*"tarik" + 0.012*"listrik" + 0.012*"layan" + 0.010*"transaksi" + 0.010*"grab" + 0.010*"kuota"
1	0.146*"transfer" + 0.039*"bank" + 0.036*"lapor" + 0.030*"error" + 0.028*"chat" + 0.027*"kirim" + 0.022*"rekening" + 0.020*"hasil" + 0.020*"biaya" + 0.017*"parah" + 0.016*"admin" + 0.015*"layan" + 0.012*"respon" + 0.010*"topup" + 0.010*"cs"
2	0.064*"bantu" + 0.060*"login" + 0.048*"akun" + 0.046*"nomor" + 0.029*"email" + 0.027*"buka" + 0.025*"handphone" + 0.016*"blokir" + 0.014*"susah" + 0.014*"salah" + 0.012*"daftar" + 0.012*"download" + 0.011*"log" + 0.010*"update" + 0.010*"hilang"

Topic	Probabilitas Kata
3	0.059*"upgrade" + 0.047*"daftar" + 0.042*"gagal" + 0.041*"premier" + 0.031*"update" + 0.023*"log" + 0.021*"akun" + 0.020*"ktp" + 0.020*"bikin" + 0.020*"foto" + 0.016*"server" + 0.015*"sulit" + 0.013*"login" + 0.013*"ribet" + 0.012*"mulu"
4	0.141*"transaksi" + 0.130*"mudah" + 0.085*"cepat" + 0.038*"security" + 0.023*"gampang" + 0.021*"aman" + 0.021*"lancar" + 0.017*"data" + 0.014*"nyaman" + 0.013*"code" + 0.012*"simpler" + 0.011*"praktis" + 0.010*"sukses" + 0.009*"manfaat" + 0.009*"pribadi"

### IV.2.3 Evaluasi

Pada tahap evaluasi, digunakan *coherence score* yaitu metode *CV coherence* dengan metode evaluasi visual dan evaluasi numerik. Alat visualisasi yang populer untuk menampilkan interpretasi topik pada LDA digunakan pada tahap ini yaitu PyLDAvis. Berikut merupakan tampilan grafik yang menunjukkan nilai *coherence* dari 10 topik.



Gambar IV.1 Grafik nilai *coherence*

Pada gambar IV.1 ditampilkan grafik dari hasil nilai *coherence* pemodelan topik yang dibuat. Grafik tersebut menunjukkan bahwa terdapat kenaikan nilai *coherence* seiring jumlah topik yang ikut bertambah. Namun, saat sampai pada topik ke 6 hingga topik ke 7 terdapat penurunan dan kembali meningkat pada topik ke 8 hingga topik ke 9. Nilai *coherence* yang semakin tinggi mengindikasikan semakin baik interpretasi topik yang dihasilkan. Dari 10 percobaan topik yang dilakukan, didapatkan 5 topik dengan nilai *coherence* tertinggi yaitu 0,396449. Nilai tersebut dapat dilihat pada tabel IV.8.

Table IV.8 Coherence Score

<b>Jumlah Topik</b>	<b>Coherence Score</b>
1	0.171413
2	0.270251
3	0.347409
4	0.396672
5	0.397612
6	0.393488
7	0.357394
8	0.376167
9	0.381555
10	0.171413

Setelah mengetahui topik dengan nilai *coherence* tertinggi, dilakukan visualisasi LDA dengan pyLDAvis, dan dari hasil visualisasi 5 topik tersebut ditemukan terdapat 2 *cluster* topik yang beririsan yaitu topik 3 dan 4. Irisan yang terdapat pada *cluster* tersebut dapat diidentifikasi bahwa kedua *cluster* tersebut dapat digabungkan menjadi 1 *cluster* topik. Berikut hasil visualisasi LDA pada gambar IV.2



Gambar IV.2 Visualisasi LDA

Pada gambar IV.2 yang menampilkan visualisasi LDA, pada setiap cluster menunjukkan terdapat *top words* atau disebut juga kata kunci yang terdapat pada setiap topik. Berikut merupakan tabel yang berisi *cluster* topik serta kata-kata kunci yang dihasilkan dari setiap topik.

Table IV.9 Kata Kunci tiap Aspek

No Topik	Kategori Aspek	Kata Kunci
0	Fitur	fitur, transaksi, transfer, promo, premier, poin, qris, bank, tunai, paylater, topup, invest, pedulilindungi, pulsa, pln, pdam, bpjs, telkom, lazada, tokopedia, grab, prakerja.
1	Pelayanan	layan, cs, call, center, callcenter, customer, service, hubungi, respon, email, chat, telepon, balas, pusat bantuan, staf, lapor, admin, komplain, sms, hubungi
2-3	Akses	akses, login, log, sign, akun, server, daftar, registrasi, buka
4	Keamanan	aman, keamanan, secure, security, privasi, pribadi

Pada tabel IV.9 ditampilkan aspek dan kata kunci yang terkandung didalamnya. Pada aspek fitur terdapat beberapa kata kunci yang berisi beberapa fitur yang terdapat pada aplikasi OVO. Lalu pada aspek akses terdapat beberapa kata kunci yang menggambarkan proses mengakses aplikasi yang dilakukan pengguna. Selanjutnya pada aspek pelayanan terdapat beberapa kata kunci yang menggambarkan pelayanan *customer service* pada OVO. Dan pada aspek keamanan menggambarkan kata kunci tentang keamanan data pengguna.

### IV.3 Analisis Sentimen Berbasis Aspek

*Aspect-Based Sentiment Analysis* atau ABSA adalah metode analisis sentimen yang digunakan untuk mengidentifikasi dan menganalisis sentimen atau pendapat yang terkait dengan aspek-aspek tertentu dalam suatu teks. Pada ABSA terdapat beberapa hal yang perlu untuk diperhatikan dalam prosesnya, yaitu *aspect extraction* dan *aspect sentiment classification*.

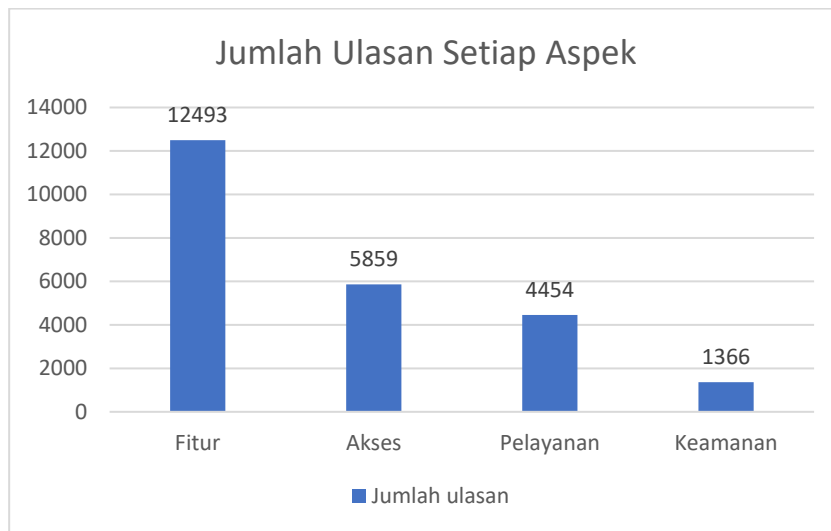
#### IV.3.1 Aspect Extraction

Setelah topik telah ditentukan pada proses LDA. Selanjutnya maka akan dilakukan proses pelabelan aspek atau *aspect extraction* yang diambil dari hasil LDA yaitu aspek fitur, akses, pelayanan, dan keamanan. Aspek-aspek tersebut memiliki beberapa kata-kata tertentu yang dapat menjadi petunjuk terhadap aspek yang dibahas dari pengguna. Berikut merupakan identifikasi aspek pada tabel IV.10.

Table IV.10 Identifikasi Aspek

Topik	Kalimat Ulasan	Interpretasi Topik
1	<b>fitur</b> pengapusan history sejarah transaksi keren	Fitur
2	cepat <b>layananya</b> mudah gampang	Pelayanan
3	<b>akses</b> lambat pening nengoknya	Akses
4	Tinggal sekali klik..sgt mudah smga <b>keamanan</b> terjamin	Keamanan

Tabel IV.10 Menunjukkan contoh kalimat yang membahas suatu topik tertentu. Kata-kata yang dicetak tebal merupakan kata-kata yang berperan penting dalam menginterpretasikan topik-topik yang ada. Keempat topik ini menjadi aspek-aspek yang akan dilakukan analisis serta diteliti pada tahap yang akan dilakukan selanjutnya.



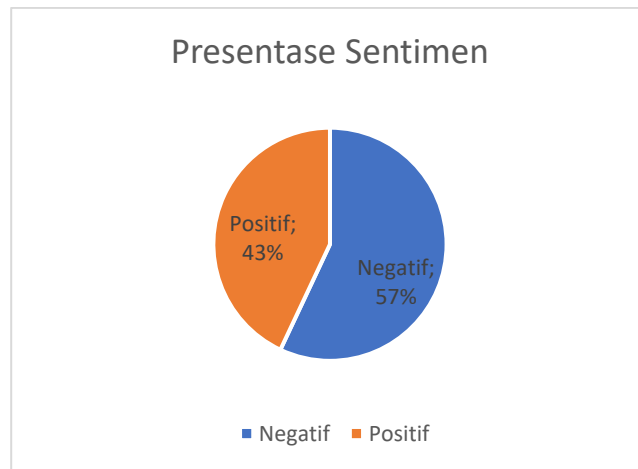
Gambar IV.3 Jumlah Ulasan Setiap Aspek

Pada gambar IV.3 menampilkan jumlah ulasan pada setiap aspek berdasarkan kata kunci yang ditentukan. Aspek fitur menjadi aspek dengan ulasan terbanyak pada aplikasi OVO sebanyak 12.493 ulasan. Dari total 43.452 ulasan, terdapat 24.172 ulasan yang mengandung empat aspek tersebut, dan sisanya sebanyak 19.280 ulasan sama sekali tidak membahas keempat aspek tersebut.

#### IV.3.2 Aspect Sentiment Classification

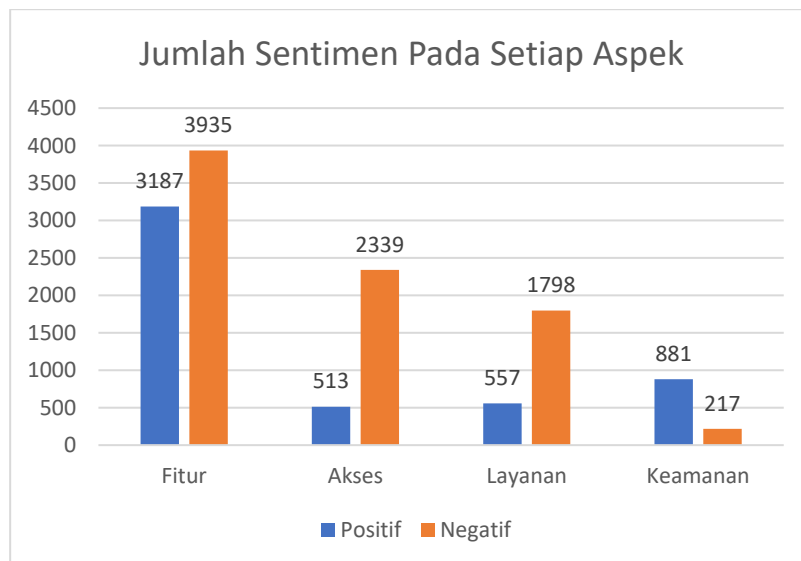
Pada proses ini, ulasan-ulasan yang telah dilakukan pengkategorian ke beberapa aspek akan dilabeli dengan sentimen yaitu menjadi positif, negatif, dan netral. Proses melabeli ini dilakukan dengan *senti strength*. Berdasarkan hasil klasifikasi yang didapatkan, menunjukkan bahwa ulasan sentimen netral memiliki jumlah yang lebih tinggi dibanding dengan sentimen positif dan negatif. Hal ini dapat mengakibatkan ketidakseimbangan pada data. Tidak hanya itu, sentimen netral juga tidak memberikan wawasan yang dapat berguna pada penelitian ini. Maka berdasarkan hal tersebut, ulasan dengan sentimen netral dihapus. Setelah dihapus, jumlah ulasan pada data berkurang sebanyak 21.398 sehingga tersisa 22.054

ulasan. Berikut merupakan hasil klasifikasi pada sentimen positif serta negatif dan jumlah sentimen keseluruhan dan pada setiap aspek.



Gambar IV.4 Presentase sentimen

Pada gambar IV.4 Menunjukkan jumlah sentimen secara keseluruhan aspek. Selain itu juga ditemukan bahwa jumlah dari sentimen negatif lebih besar dibandingkan dengan jumlah dari sentimen positif.



Gambar IV.5 Jumlah sentimen pada setiap aspek

Lalu pada gambar IV.5 Menampilkan jumlah sentimen pada setiap aspek. Jumlah sentimen pada spek fitur tetap menjadi yang terbanyak dengan total sebanyak 7.122 ulasan. Dari jumlah 22.054 ulasan, ada 13.418 ulasan yang termasuk dalam empat aspek tersebut dan sisanya sebanyak 8.636 ulasan tidak membahas keempat

aspek tersebut. Dan jumlah sentimen negatif pada gambar IV.5 masih yang terbanyak pada aspek selain aspek keamanan.

#### IV.4 Split Dataset

Pada tahap *split dataset*, proses yang dilakukan yaitu dengan membagi *dataset* menjadi data *train* dan data *test*. Data *train* berguna untuk melatih model, sedangkan *data test* berguna untuk menguji performa model. Skenario rasio pembagian yang digunakan yaitu pembagian 70:30, 75:25, dan 80:20.

Table IV.11 Splitting Dataset per Rasio

<b>Rasio</b>	<b>Data Train</b>	<b>Data Test</b>	<b>Jumlah</b>
70:30	15.437	6.617	22.054
75:25	16.540	5.514	
80:20	17.643	4.411	

Pada tabel IV.14 menunjukkan pembagian rasio data *train* dan data *test* terhadap 22.054 data. Pada rasio 70:30 terdapat 15.437 data *train* dan 6.617 data *test*, lalu pada rasio 75:25 terdapat 16.540 data *train* dan 5.514 data *test*, serta pada rasio 80:20 terdapat 17.643 data *train* dan 4.411 data *test*.

Table IV.12 Splitting Dataset setiap Aspek per Rasio

<b>Aspek</b>	<b>Rasio</b>						<b>Jumlah</b>
	<b>70:30</b>		<b>75:25</b>		<b>80:20</b>		
	<i>Data Train</i>	<i>Data Test</i>	<i>Data Train</i>	<i>Data Test</i>	<i>Data Train</i>	<i>Data Test</i>	
Fitur	4958	2137	5341	5697	5697	1425	7122
Akses	1966	856	2139	713	2281	571	2852
Pelayanan	1648	707	1766	589	1884	471	2355
Keamanan	768	330	823	275	878	220	1098

Setelah membagi data *train* dan data *test* pada keseluruhan data, selanjutnya dilakukan juga pembagian data *train* dan data *test* pada setiap aspek. Pada tabel



IV.15 menunjukkan jumlah pembagian data *train* dan data *test* pada rasio 70:30, 75:25, dan 80:20 pada setiap aspek. Sebagai contoh yaitu pada aspek fitur pada rasio 70:30 terdapat 4.958 data latih dan 2.137 data uji.

#### IV.5 TF-IDF

Pada proses ini teks diubah menjadi representasi numerik menggunakan metode TF-IDF, yang menggabungkan nilai TF (*Term Frequency*) dan IDF (*Inverse Document Frequency*) untuk menghitung bobot dari setiap kata dalam dokumen. Berikut merupakan contoh tiga dokumen dan perhitungan pada pembobotan TF-IDF.

Table IV.13 Sampel Pembobotan Kata

Ulasan data	Dokumen 1	lambat transaksi
	Dokumen 2	cepat gampang akurat
	Dokumen 3	mudah cepat

Table IV.14 Perhitungan TF dan IDF

Term	TF			df	D/df	IDF ( $\ln(N/df)$ )
	d1	d2	d3			
lambat	0.5	0	0	1	3	0.693
transaksi	0.5	0	0	1	3	0.693
cepat	0	0.33	0	2	1.5	0.287
gampang	0	0.33	0	1	3	0.693
Akurat	0	0.33	0	1	3	0.693
mudah	0	0	0.5	1	3	0.693
cepat	0	0.33	0.5	2	1.5	0.287
<b>Panjang Dokumen</b>				<b>Total bobot tiap dokumen</b>		

Table IV.15 Perhitungan TF-IDF

Term	TF-IDF		
	D1	D2	D3
lambat	0.3465	0	0
transaksi	0.3465	0	0
cepat	0	0.0958	0.3465
gampang	0	0.2310	0
Akurat	0	0.3662	0
mudah	0	0	0.5493

Tabel IV.13 menunjukkan hasil perhitungan TF-IDF untuk beberapa *term* yang terdapat dalam tiga dokumen. Kolom "TF" mencerminkan seberapa sering *term* tersebut muncul dalam masing-masing dokumen, sementara kolom "DF" menunjukkan berapa banyak dokumen yang mengandung *term* tersebut. "D/DF" adalah rasio antara jumlah total dokumen dengan *Document Frequency*, dan "IDF" merupakan logaritma natural dari nilai D/DF. Selanjutnya, kolom "TF-IDF" menunjukkan hasil perkalian antara nilai TF dan IDF. Nilai TF-IDF ini memberikan indikasi tentang pentingnya *term* tersebut dalam konteks keseluruhan koleksi dokumen.

#### IV.6 Implementasi *Support Vector Machine*

Pada tahap ini, dilakukan pengembangan model untuk melakukan prediksi sentimen. Sebelumnya, data dibagi menjadi dua bagian: data latih (*train*) digunakan untuk melatih model, dan data uji (*test*) digunakan untuk menguji model. Data latih ini berperan penting dalam memprediksi klasifikasi sentimen positif dan negatif pada ulasan. Algoritma yang digunakan dalam penelitian ini adalah *Support Vector Machine* dengan *kernel linier*. Berikut merupakan *pseudocode* untuk membuat algoritma SVM.

1. *Split dataset* menjadi *training set* ( $X_{train}, y_{train}$ ) dan *test set* ( $X_{test}, y_{test}$ ).

2. Pilih *kernel* yang cocok, dan penelitian ini dipilih kernel linier yang sesuai untuk analisis sentimen pembagian positif dan negatif.
3. Latih model SVM menggunakan set pelatihan ( $X_{train}, y_{train}$ ) dan temukan *hyperplane* optimal yang memaksimalkan *margin*.
4. Uji model SVM
5. Evaluasi performa dari model SVM menggunakan *confusion matrix*

Membuat fungsi penentu keputusan klasifikasi algoritma SVM dengan *kernel* linier karena berdasarkan pengujian menggunakan *gridsearch* dari kernel linier, rbf, dan poly didapatkan kernel linier mendapatkan akurasi tertinggi sebesar 0.980559 . Dengan algoritma ini, diperoleh *hyperplane* terbaik dengan persamaan berikut.

$$f(\mathbf{x}) = \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ 4673 \\ 4674 \\ 4675 \end{matrix} \begin{pmatrix} 0.31873577 \\ 0 \\ 0 \\ \vdots \\ -0.28043967 \\ 0 \\ -0.42555839 \end{pmatrix}^T \cdot \mathbf{x} + (-0.2534382)$$

Berdasarkan persamaan diatas didapatkan nilai bobot berupa *matrix* berukuran 1x4675 dan nilai *intercept* (b) sebesar -0.2534382. Sebagai contoh perhitungan proses klasifikasi, pada *baseline* model SVM dengan rasio 75:25 dengan menggunakan teks "aplikasi ini bikin duit cepet aabis aaja yy" sebagai contoh. Maka didapatkan persamaan sebagai berikut.

$$f(\mathbf{x}) = \begin{matrix} 1 \\ 2 \\ 3 \\ \vdots \\ 4673 \\ 4674 \\ 4675 \end{matrix} \begin{pmatrix} 0.31873577 \\ 0 \\ 0 \\ \vdots \\ -0.28043967 \\ 0 \\ -0.42555839 \end{pmatrix}^T \cdot [0.55412298, 0.55412298, 0 \dots 0,0,0] + (-0.2534382)$$

Setelah teks tersebut dimasukan kedalam persamaan, maka didapatkan hasil  $f(\mathbf{x}) = -1$  .Hasil ini menunjukkan bahwa teks "aplikasi ini bikin duit cepet aabis aaja yy" termasuk kedalam teks yang memiliki label negatif.

## BAB V

### HASIL DAN PEMBAHASAN

#### V.1 Evaluasi dan Validasi

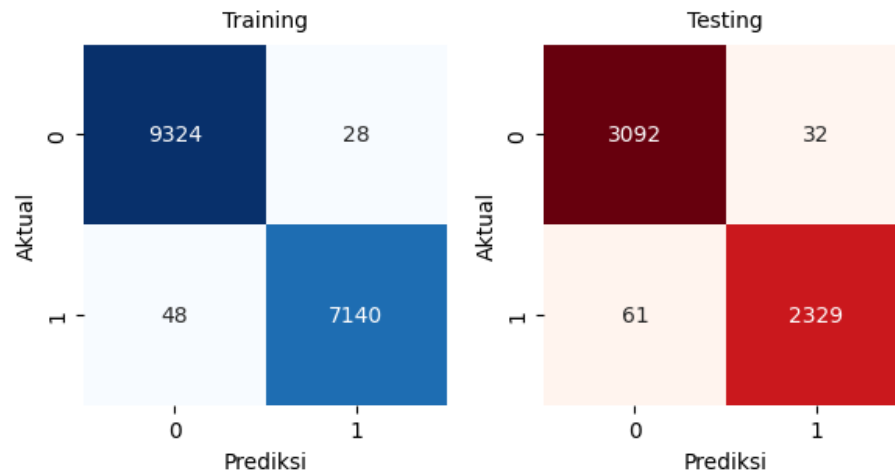
Pada proses ini dilakukan evaluasi serta validasi untuk mengukur tingkat performa model yang telah dibuat menggunakan algoritma *Support Vector Machine*. Proses ini membandingkan dari ketiga rasio pembagian data yaitu 70:30, 75:25, dan 80:20, sehingga dapat diketahui performa dari masing-masing rasio. Dengan tujuan mengetahui efektivitas performa model dalam memprediksi sentimen melalui beberapa pembagian rasio data *train* serta data *test* yang berbeda. Metode evaluasi yang digunakan pada penelitian ini yaitu *confusion matrix*.

**Table V.1 Jumlah Akurasi pada setiap Rasio**

<b>Ratio</b>	<b><i>Splitting Data</i></b>	<b><i>True Positive</i></b>	<b><i>True Negatif</i></b>	<b><i>Total Prediction Data</i></b>	<b><i>Total Actual Data</i></b>	<b><i>Accuracy</i></b>
70:30	<i>Data Train</i>	6662	8704	15366	15437	99.5%
	<i>Data Test</i>	2798	3704	6502	6617	98.2%
75:25	<i>Data Train</i>	7140	9324	16464	16540	99.5%
	<i>Data Test</i>	2369	3092	5461	5514	98.32%
80:20	<i>Data Train</i>	7588	9965	17553	17463	99.5%
	<i>Data Test</i>	1886	2453	4339	4411	98.3%

Pada tabel V.1 menunjukkan hasil rasio pemisahan data (*splitting data*) dalam tiga skenario yang berbeda: 70:30, 75:25, dan 80:20. Rasio ini mengacu pada pembagian data menjadi data pelatihan (*data train*) dan data pengujian (*data test*). Hitung akurasi dengan membagi jumlah prediksi yang benar dengan jumlah total data prediksi. Hasil yang didapatkan yaitu nilai akurasi tertinggi terdapat pada rasio 75:25 dengan hasil akurasi pada *data train* sebesar 99.5% dan hasil akurasi

pada data *test* sebesar 98,32%. Berikut merupakan hasil *confusion matrix* pada rasio 75:25.



Gambar V.1 *Confusion Matrix* Rasio 75:25

Pada gambar V.1 menunjukkan hasil *confusion matrix* dari data *training* terdapat 9324 data yang terklasifikasikan dengan benar sebagai TN (*True Negatif*) dan 7140 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 28 data yang salah sebagai FN (*False Negative*) dan 48 data yang salah sebagai FP (*False Positive*). Lalu pada hasil *confusion matrix* dari *data testing* terdapat 3092 data yang terklasifikasikan dengan benar sebagai TN (*True Negatif*) dan 2329 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 32 data yang salah sebagai FN (*False Negative*) dan 61 data yang salah sebagai FP (*False Positive*). Hasil tersebut menunjukkan model berhasil melakukan prediksi dengan akurasi yang baik untuk positif dan negatif.

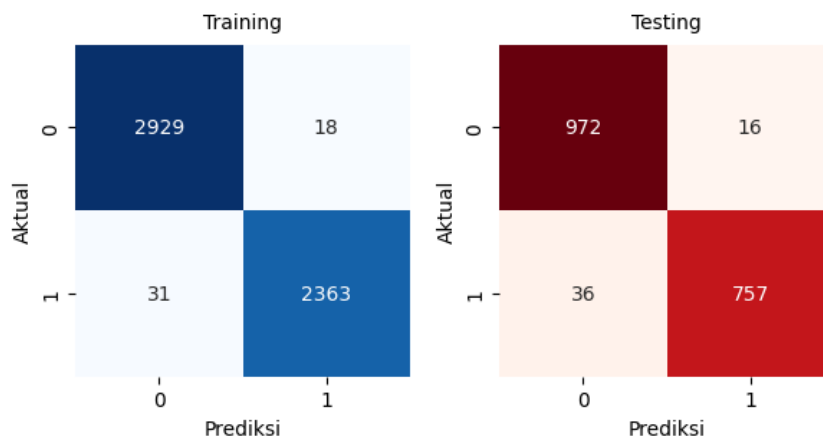
Dalam rangka memverifikasi kehandalan model dan mengurangi risiko *overfitting*, penulis memeriksa varians serta bias antara data *train* dan data *test* dalam hasil model, ketika model memiliki bias dan varians yang tinggi maka disebut *overfitting* yang didapatkan dari penjelasan oleh (Christanto & Setiabudi, 2020). Bias dapat diketahui dari nilai akurasi, akurasi yang semakin tinggi maka bias akan semakin rendah. Lalu varians dapat diketahui dari penjumlahan nilai antara selisih rata-rata akurasi dengan data *train* dan *test* dibagi dua. Semakin tinggi nilai varians maka semakin tinggi kemungkinan terjadi *overfitting* pada model

Nilai ini menandakan bahwa model menunjukkan konsistensi yang tinggi dalam kinerjanya, karena mampu memberikan tingkat akurasi yang tinggi secara konsisten pada berbagai *subset* data yang digunakan dalam proses validasi. Selain itu, hasil uji *overfitting* juga menunjukkan bahwa tidak ada indikasi *overfitting* pada model, karena hasil pelatihan dan pengujian menghasilkan hasil yang konsisten. Sehingga rasio pembagian 75:25 akan digunakan juga untuk pemodelan pada aspek-aspek lainnya.

Table V.2 Classification Report pada setiap Aspek

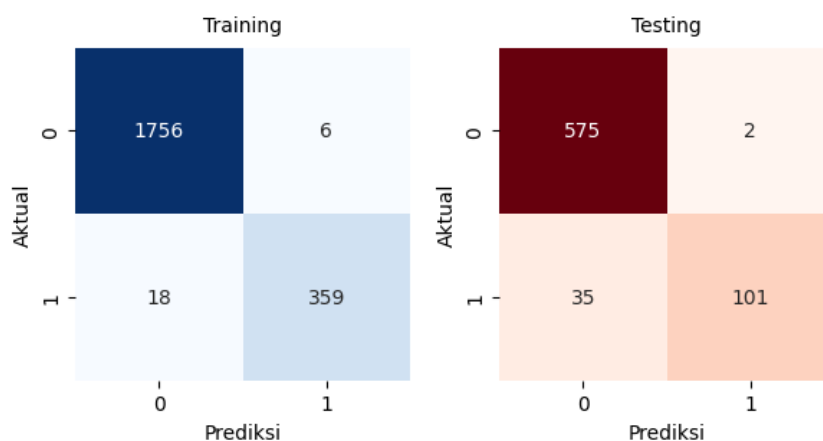
Aspek	Klasifikasi	Presisi	Recall	F1-Score	Akurasi
Fitur	Positif	0.98	0.95	0.97	97.08%
	Negatif	0.96	0.98	0.97	
Akses	Positif	0.98	0.74	0.85	94.81%
	Negatif	0.94	1.00	0.97	
Pelayanan	Positif	0.93	0.69	0.79	91.34%
	Negatif	0.91	0.98	0.95	
Keamanan	Positif	0.97	0.99	0.98	96.36%
	Negatif	0.95	0.88	0.91	

Hasil klasifikasi pada tabel V.3 diatas menunjukkan beberapa aspek pada OVO, yang pertama pada Fitur positif memiliki presisi 0,98, *recall* 0,95 dan *f1 score* sebesar 0,97 hasil akurasi yang tinggi yaitu sebesar 97,08% menandakan model dapat mengklasifikasi dengan baik. Yang kedua pada akses positif memiliki presisi 0,98, *recall* 0,74 dan *f1 score* 0,85 sebesar hasil akurasi yang tinggi yaitu sebesar 94,81% menandakan model tersebut sedikit kesulitan dalam mengenali dan mengklasifikasian dengan tepat akses yang positif. Yang ketiga pada aspek pelayanan positif memiliki presisi 0,93, *recall* 0,69 dan *f1 score* 0,79 sebesar hasil akurasi yang tinggi yaitu sebesar 91,34% menandakan model tersebut sedikit kesulitan dalam mengenali dan mengklasifikasian dengan tepat pelayanan yang positif. yang keempat pada aspek keamanan positif memiliki presisi 0,97, *recall* 0,99 dan *f1 score* sebesar 0,98 hasil akurasi yang tinggi yaitu sebesar 96,36% menandakan model dapat mengklasifikasi dengan baik.



**Gambar V.2 Confusion Matrix Aspek Fitur**

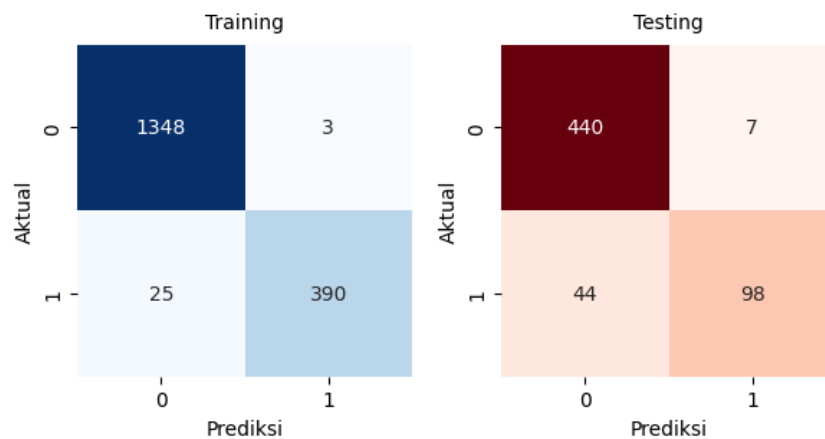
Pada gambar V.2 menampilkan hasil *confusion matrix* pada aspek fitur dari data *training* terdapat 2929 data yang terklasifikasikan dengan benar sebagai TN (*True Negative*) dan 2363 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 18 data yang salah sebagai FN (*False Negative*) dan 31 data yang salah sebagai FP (*False Positive*). Dan akurasi pada data *training* aspek fitur sebesar 99,08% dan pada data *testing* sebesar 97,08%. Lalu pada hasil *confusion matrix* dari data *testing* terdapat 972 data yang terklasifikasikan dengan benar sebagai TN (*True Negative*) dan 757 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 16 data yang salah sebagai FN (*False Negative*) dan 36 data yang salah sebagai FP (*False Positive*).



**Gambar V.3 Confusion Matrix Aspek Akses**

Pada gambar V.3 menampilkan hasil *confusion matrix* pada aspek akses dari data *training* terdapat 1756 data yang terklasifikasikan dengan benar sebagai TN (*True*

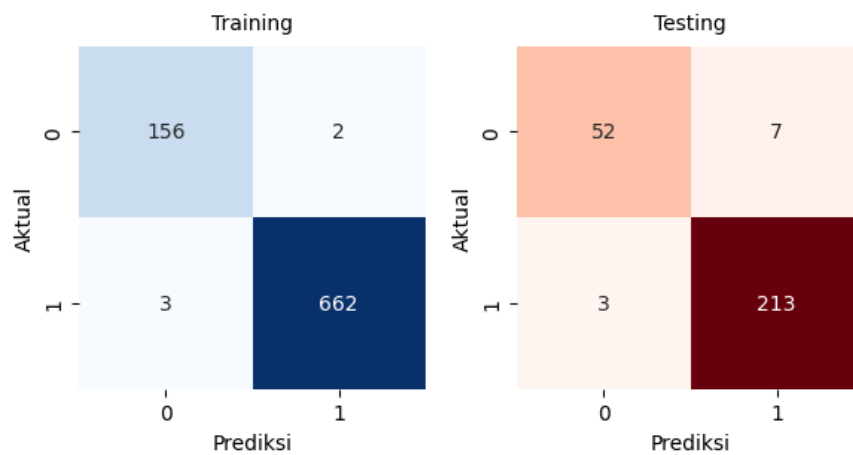
*Negative*) dan 359 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 6 data yang salah sebagai FN (*False Negative*) dan 18 data yang salah sebagai FP (*False Positive*). Dan akurasi pada data *training* aspek akses sebesar 98,87% dan pada data *testing* sebesar 94,53%. Lalu pada hasil *confusion matrix* dari data *testing* terdapat 575 data yang terklasifikasikan dengan benar sebagai TN (*True Negative*) dan 101 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 2 data yang salah sebagai FN (*False Negative*) dan 35 data yang salah sebagai FP (*False Positive*).



**Gambar V.4 Confusion Matrix Aspek Pelayanan**

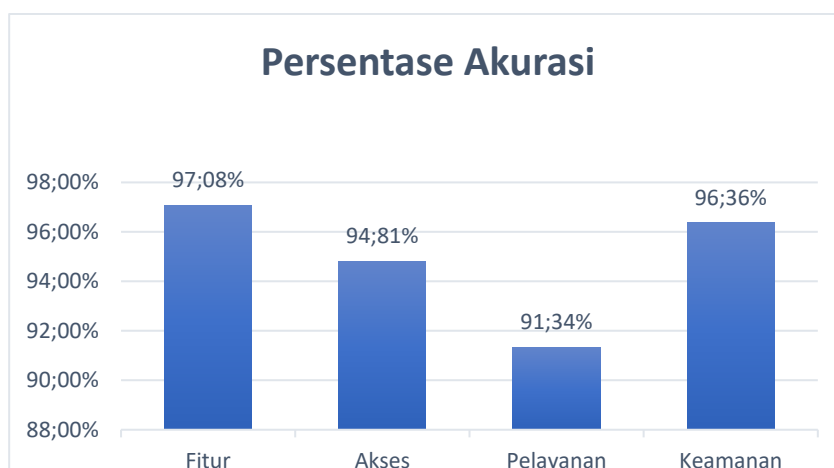
Pada gambar V.4 menampilkan hasil *confusion matrix* pada aspek pelayanan dari data *training* terdapat 1348 data yang terklasifikasikan dengan benar sebagai TN (*True Negatif*) dan 390 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 3 data yang salah sebagai FN (*False Negative*) dan 25 data yang salah sebagai FP (*False Positive*). Dan akurasi pada data *training* aspek pelayanan sebesar 98,4% dan pada data *testing* sebesar 91,34%. Lalu pada hasil *confusion matrix* dari data *testing* terdapat 440 data yang terklasifikasikan dengan benar sebagai TN (*True Negatif*) dan 98 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 7 data yang salah sebagai FN (*False Negatif*) dan 44 data yang salah sebagai FP (*False Positive*).





**Gambar V.5 Confusion Matrix Aspek Keamanan**

Pada gambar V.5 menampilkan hasil *confusion matrix* pada aspek keamanan dari *data training* terdapat 156 data yang terklasifikasikan dengan benar sebagai TN (*True Negatif*) dan 663 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 2 data yang salah sebagai FN (*False Negatif*) dan 2 data yang salah sebagai FP (*False Positive*). Dan akurasi pada *data training* aspek keamanan sebesar 99.5% dan pada data *testing* sebesar 96,72%. Lalu pada hasil *confusion matrix* dari data *testing* terdapat 52 data yang terklasifikasikan dengan benar sebagai TN (*True Negative*) dan 214 data yang terklasifikasikan TP (*True Positive*). Namun terdapat 7 data yang salah sebagai FN (*False Negative*) dan 2 data yang salah sebagai FP (*False Positive*).

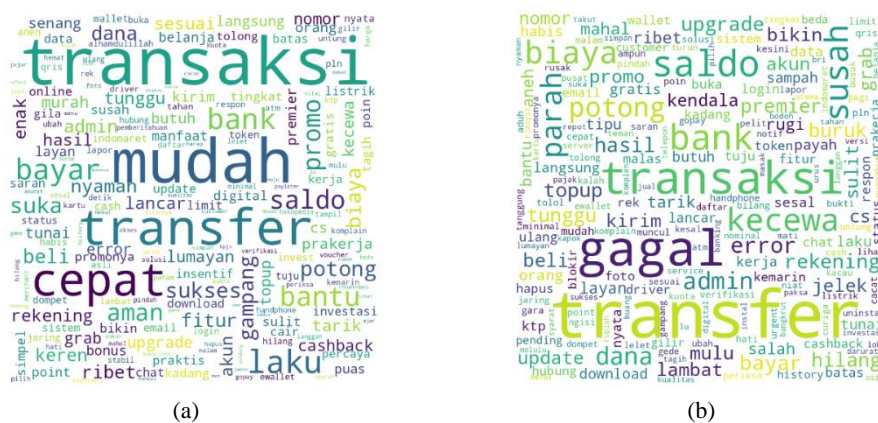


**Gambar V.6 Persentase Akurasi tiap Aspek**

Berdasarkan Gambar V.6 yang membandingkan akurasi pada setiap aspek, dapat disimpulkan bahwa aspek Fitur memiliki tingkat akurasi tertinggi, yaitu 97,58%, dan memiliki standar deviasi yang terendah, hanya 0,02. Penyebabnya adalah karena aspek Fitur memiliki jumlah data yang lebih besar dan distribusi kelas yang lebih seimbang. Keadaan ini memberikan kesempatan bagi model untuk belajar dari lebih banyak contoh, sehingga menghasilkan prediksi yang lebih akurat.

## V.2 Visualisasi Word Cloud

Wordcloud digunakan dalam memberi bantuan dalam menggambarkan frekuensi kemunculan kata-kata dalam bentuk visual. Berikut visualisasi *word cloud* pada keempat aspek.



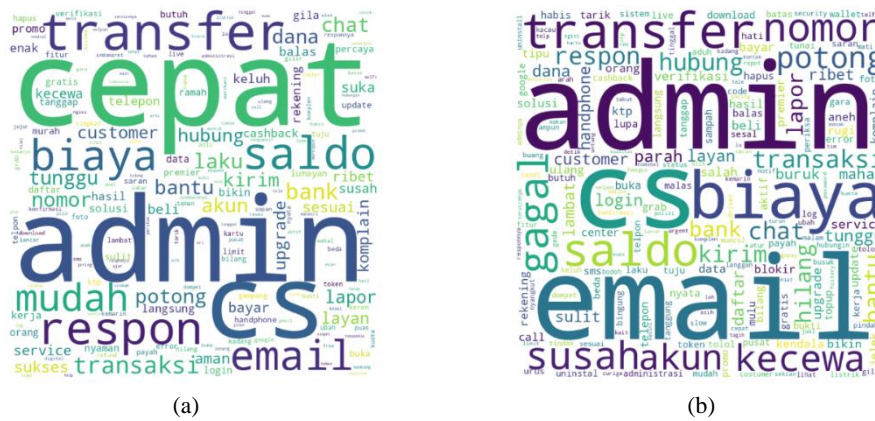
Gambar V.7 Wordcloud Ulasan Aspek Fitur Positif (a) dan Negatif (b)

Table V.3 Jumlah Kemunculan Kata Aspek Fitur

Fitur			
Positif		Negatif	
Kata	Jumlah	Kata	Jumlah
transaksi	1861	transfer	2621
mudah	1604	gagal	1332
transfer	1007	transaksi	1115
cepat	831	saldo	652
laku	315	bank	616
bank	225	susah	522
bayar	253	kecewa	495
bantu	224	biaya	476
saldo	198	parah	408
promo	175	potong	361



Terlihat pada gambar V.8 yang merupakan *wordcloud* aspek akses. Lalu pada tabel V.6 terlihat pada sentimen positif terlihat banyak kata *login*, buka, akun, mudah. Bisa disimpulkan banyak pengguna yang merasa akses masuk ke aplikasi OVO mudah. Pada sentimen negatif terlihat banyak kata akun, *login*, susah. Bisa disimpulkan bahwa juga terdapat pengguna merasa susah dalam *login* akun ovo mereka. OVO perlu melakukan peningkatan pada beberapa bagian terkait akses pengguna, terutama saat pengguna mencoba *login* setelah *logout*. Proses verifikasi saat mengakses kembali akun sebelumnya saat *login* yang rumit sehingga menyulitkan proses *login*. Selain itu, pengguna menghadapi kendala dalam mendaftar kembali karena batasan penggunaan akun sebelumnya.



Gambar V.9 Wordcloud Ulasan Aspek Layanan Positif (a) dan Negatif (b)

Table V.5 Jumlah Kemunculan Kata Aspek Layanan

Layanan			
Positif		Negatif	
Kata	Jumlah	Kata	Jumlah
cepat	169	admin	528
admin	161	email	494
cs	142	cs	419
transfer	121	biaya	415
respon	93	tranfer	387
saldo	90	saldo	332
biaya	81	gagal	267
email	81	nomor	249
mudah	75	susah	234
transaksi	72	kecewa	211



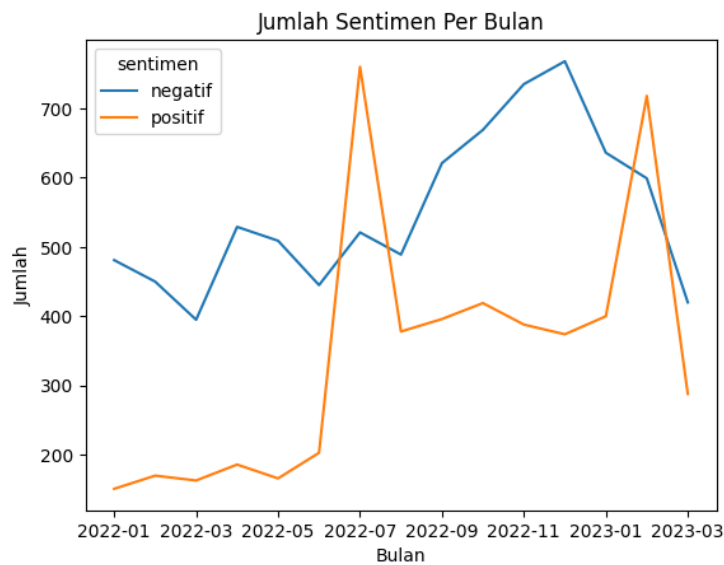
Table V.6 Jumlah Kemunculan Kata Aspek Keamanan

<b>Keamanan</b>			
<b>Positif</b>		<b>Negatif</b>	
<b>Kata</b>	<b>Jumlah</b>	<b>Kata</b>	<b>Jumlah</b>
aman	571	<i>security</i>	116
nyaman	332	gagal	83
mudah	239	<i>code</i>	77
cepat	175	salah	44
transaksi	146	transfer	41
transfer	82	<i>login</i>	36
akun	46	saldo	31
bayar	46	nyaman	31
percaya	42	susah	30
saldo	38	transaksi	28

Pada tabel V.8 terlihat sentimen positif terlihat banyak kata aman, mudah, transaksi. Bisa disimpulkan banyak pengguna yang merasa keamanan pada OVO aman dan mudah. Pada sentimen negatif terlihat banyak kata *security*, gagal, *code*. Bisa disimpulkan bahwa juga terdapat pengguna merasa kendala dalam mendapatkan kode *security*, serta mengalami kegagalan. Meski begitu, sentimen negatif terhadap aspek keamanan OVO jauh lebih jarang dibandingkan dengan sentimen positif.

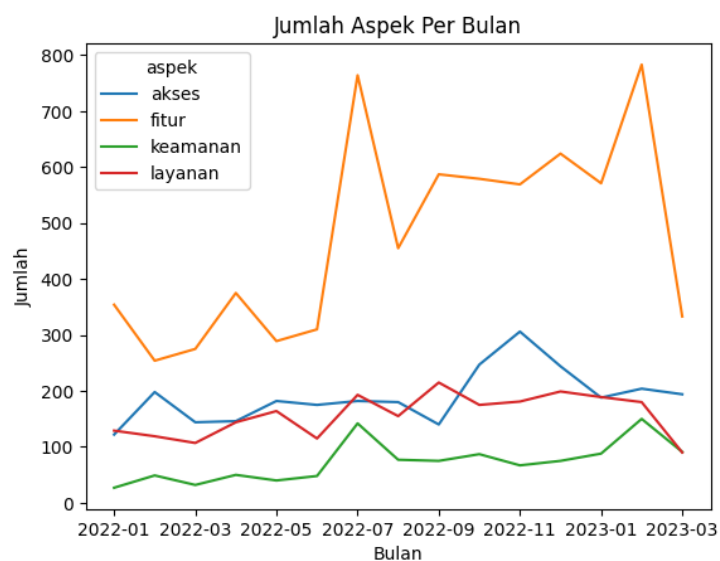
### V.3 Grafik

Pada tahap ini, dilakukan proses visualisasi grafik yang menunjukkan jumlah data ulasan dalam rentang waktu bulan ke bulan. Gambar V.17 menampilkan jumlah ulasan berdasarkan sentimen. Sementara itu, gambar V.18 menampilkan jumlah ulasan untuk setiap aspek dari bulan ke bulan. Grafik tersebut memberikan informasi sentimen apa yang paling tinggi dan yang paling rendah pada ulasan di bulan-bulan tertentu pada ulasan aplikasi OVO di Google Play Store. Lalu dari periode tersebut dicari tau apakah terdapat kejadian peristiwa yang berkaitan dengan aplikasi seperti promo atau gangguan sehingga memberi pengaruh terhadap jumlah sentimen yang muncul pada tiap bulan.



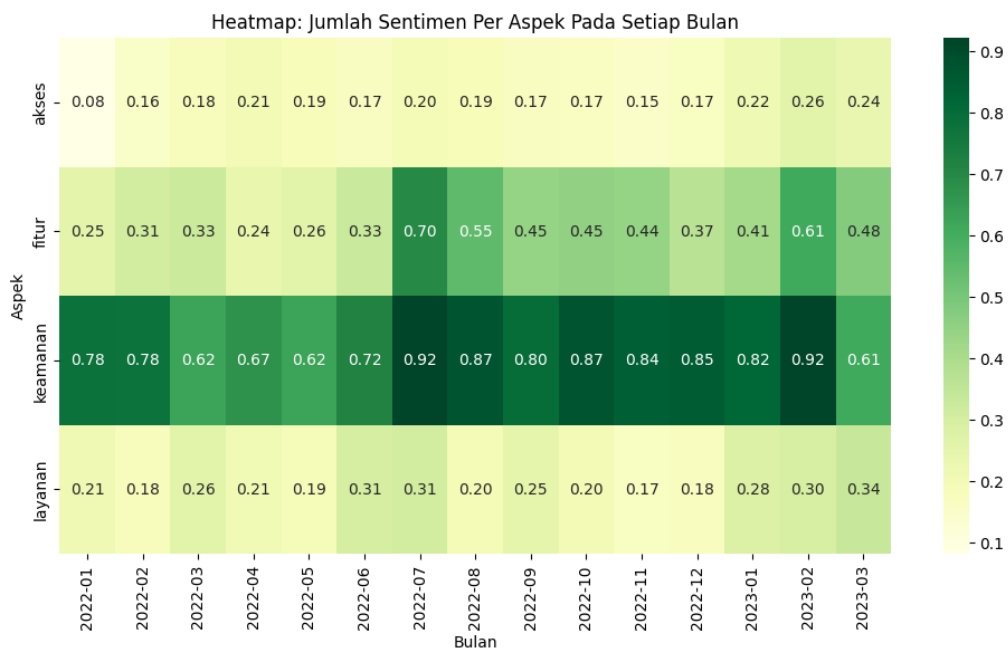
Gambar V.11 Jumlah Sentimen per Bulan

Pada gambar V.17, terlihat bahwa jumlah ulasan pengguna OVO dengan sentimen negatif tertinggi terjadi pada bulan Desember 2022 (2022-12) dengan 768 ulasan, sementara jumlah ulasan sentimen negatif terendah terjadi pada bulan Maret 2022 (2022-03) hanya sebanyak 395 ulasan. Sentimen positif mencapai puncaknya pada bulan Juli 2022 dengan 760 ulasan, namun mengalami penurunan drastis pada bulan Agustus. Namun, sentimen positif kembali meningkat menjadi 718 ulasan pada bulan Februari 2023 (2023-02). Jumlah ulasan sentimen positif terendah terjadi pada bulan Januari 2022 (2022-01) dengan hanya 151 ulasan.



Gambar V.12 Jumlah Aspek per Bulan

Sementara itu, pada gambar V.18 yang menampilkan jumlah ulasan untuk setiap aspek, aspek fitur memiliki jumlah ulasan terendah pada bulan Februari 2022 (2022-02) dengan 254 ulasan, namun satu tahun kemudian, jumlah ulasan mengenai fitur mencapai puncaknya dengan 783 ulasan. Pada aspek akses, jumlah ulasan tertinggi terjadi pada bulan November 2022 (2022-11) dengan 306 ulasan, sementara jumlah ulasan terendah terjadi pada bulan Januari 2022 (2022-01) dengan hanya 122 ulasan. Aspek pelayanan keamanan memiliki jumlah ulasan tertinggi pada bulan September 2022 (2022-09) dengan 215 ulasan, sementara jumlah ulasan terendahnya terjadi pada bulan Maret 2023 (2023-03) hanya sebanyak 90 ulasan. Pada aspek keamanan, jumlah ulasan tertinggi tercatat pada bulan Februari 2023 (2023-02) dengan 150 ulasan, sementara yang terendah terjadi pada bulan Januari 2022 (2022-01) hanya dengan 27 ulasan.



Gambar V.13 Jumlah Sentimen per Aspek pada Setiap Bulan

Gambar V.13 merupakan sebuah *heatmap* yang menjelaskan jumlah sentimen pada setiap bulan berdasarkan aspek yang terkait. Warna pada *heatmap* mencerminkan tingkat sentimen dari masing-masing aspek dalam setiap bulannya. Semakin gelap warna pada *heatmap*, semakin tinggi jumlah sentimen positif yang diperoleh oleh aspek pada bulan tersebut. Sebagai contoh, pada bulan Juli 2022, aspek keamanan ditunjukkan dengan warna hijau gelap dan angka 0,92,



mengindikasikan bahwa pada bulan tersebut, aspek keamanan memiliki jumlah sentimen positif yang lebih tinggi daripada sentimen negatif. Di sisi lain, pada bulan Januari 2022, aspek akses ditunjukkan dengan warna hijau cerah dan angka 0,08, menandakan bahwa pada bulan tersebut, aspek akses memiliki jumlah sentimen negatif yang lebih dominan daripada sentimen positif.

Aspek keamanan lebih didominasi oleh sentimen positif, sedangkan akses dan pelayanan lebih didominasi oleh sentimen negatif. Aspek Fitur menerima banyak ulasan yang tidak menyenangkan pada bulan April 2022 (2022-04). Namun, pada bulan Juli 2022 (2022-07), ulasan positif mengalami peningkatan yang tajam, yang mungkin terkait dengan peluncuran promo Shoptakuler oleh OVO pada rentang tanggal 1 hingga 31 Juli 2022, dimulai dari pukul 15:00 hingga 21:00 WIB setiap harinya (Paramahita, 2022). Setelahnya, sentimen terkait aspek Fitur mengalami fluktuasi mulai dari Agustus 2022 (2022-08) hingga Januari 2023 (2023-01). Menariknya, ulasan positif kembali meningkat pada Februari 2023 (2023-02), yang mungkin juga berhubungan dengan adanya Promo Bareksa OVO Shoptakuler *Online Fest* periode Februari 2023 (Dewi, 2023).

## **BAB VI**

### **KESIMPULAN DAN SARAN**

#### **VI.1 Kesimpulan**

Temuan dari penelitian yang dilaksanakan telah menghasilkan beberapa kesimpulan penting yang dapat dirangkum sebagai berikut:

1. Melalui penerapan metode *Topic Modelling*, terutama dengan menggunakan teknik *Latent Dirichlet Allocation* (LDA), berhasil mengidentifikasi berbagai aspek yang dibicarakan oleh pengguna *fintech* OVO di Google Play Store. Setelah melakukan 100 iterasi dengan 10 topik, ditemukan 5 topik dengan skor kohesi tertinggi sebesar 0.397612. Namun, saat dilakukan visualisasi LDA dengan *pyLDAvis*, terlihat bahwa terdapat 2 *cluster* topik (topik 3 dan 4) yang tumpang tindih, sehingga kedua *cluster* tersebut dapat digabungkan menjadi 1 *cluster* topik. Dengan demikian, ada 4 topik yang dapat dianalisis, yaitu Fitur, Akses, Pelayanan, dan Keamanan.
2. Penggunaan metode klasifikasi *Support Vector Machine* dalam implementasi analisis sentimen berbasis aspek berhasil memberikan hasil yang positif dalam mengklasifikasikan sentimen pengguna terhadap berbagai aspek pelayanan OVO. Dengan menggunakan pembagian *dataset* optimal 75:25, tingkat akurasi yang tinggi berhasil dicapai untuk setiap aspek. Akurasi klasifikasi mencapai 97,58% untuk aspek Fitur, 95,23% untuk aspek Akses, 92,52% untuk aspek Pelayanan, dan 96,72% untuk aspek Keamanan. Rata-rata akurasi keseluruhan juga mencapai tingkat yang memuaskan. Selain itu, uji *overfitting* menunjukkan bahwa model yang digunakan untuk setiap aspek tidak menunjukkan tanda-tanda *overfitting*, hasil tersebut menegaskan bahwa kinerja model dalam menilai sentimen pada setiap aspek sangat baik.
3. Melalui analisis sentimen dan visualisasi *Wordcloud* pada setiap aspek pelayanan OVO, telah berhasil mengidentifikasi faktor-faktor yang menyebabkan *technostress* pada pengguna. Metode visualisasi *Wordcloud* membantu memahami seberapa sering kata-kata tertentu muncul dalam teks dan mengenali sentimen negatif pada setiap aspek yang berpotensi menyebabkan *technostress*. Hasil penelitian menunjukkan bahwa beberapa

faktor penyebab *technostress* pada pengguna OVO termasuk potongan saldo, biaya, fitur *login*, kesulitan pendaftaran akun, pelayanan admin atau *customer service*, serta transfer. Analisis juga menemukan bahwa aspek fitur merupakan yang paling mendapatkan sentimen negatif yang tinggi, diikuti oleh aspek akses dan pelayanan. Meskipun aspek keamanan juga memiliki sentimen negatif, jumlahnya cenderung lebih rendah dibandingkan dengan sentimen positif, menandakan kepuasan pengguna OVO yang lebih tinggi terhadap aspek keamanan dibandingkan dengan aspek lainnya.

## VI.2 Saran

Dalam penelitian ini, penulis menyadari adanya beberapa kekurangan yang perlu diperhatikan dan ditingkatkan di masa mendatang. Berdasarkan kesimpulan penelitian, penulis ingin memberikan beberapa saran serta rekomendasi yang dapat menjadi panduan untuk pengembangan dan perbaikan di penelitian berikutnya dan dalam peningkatan pada pelayanan OVO.

1. Pemodelan topik menggunakan algoritma LDA telah berhasil menghasilkan kelompok aspek yang diteliti, meskipun nilai coherence tertinggi hanya mencapai 0.396449. Terdapat beberapa faktor yang dapat memengaruhi angka tersebut, seperti kompleksitas data, ukuran *dataset*, pengaturan hyperparameter, proses konvergensi model, dan tahapan proses *preprocessing*. Untuk meningkatkan nilai coherence, penting untuk mempertimbangkan faktor-faktor tersebut, dan alternatifnya, juga dapat mencoba menggunakan algoritma pemodelan topik lain, misalnya LSA.
2. Metode klasifikasi *Support Vector Machine* telah digunakan untuk membangun model klasifikasi analisis sentimen berbasis aspek yang menunjukkan akurasi yang sangat baik. Meskipun demikian, hasil tersebut dapat berubah jika terdapat penambahan pada data, maka penting untuk terus memantau performa model dan memeriksa terhadap *overfitting* ketika melakukan perubahan atau menambahkan data baru. Selain itu, untuk penelitian berikutnya, dapat dimungkinkan untuk menggunakan sumber data lainnya yang terdapat pada *platform* selain Google Play Store. Selain *Support*

*Vector Machine*, model klasifikasi juga bisa diujicobakan menggunakan pendekatan *deep learning* atau metode *machine learning* lainnya.

3. Hasil dari penelitian ini menunjukkan beberapa aspek yang perlu diperhatikan dan ditingkatkan oleh OVO dalam hal kualitas fitur, akses, dan pelayanan guna meningkatkan tingkat kepuasan pelanggan, meningkatkan penilaian, dan mengurangi ulasan negatif.

Beberapa langkah yang disarankan oleh peneliti untuk dilakukan meliputi:

- 1) Meningkatkan Kualitas Fitur Transaksi dan Transfer: Area Fitur merupakan bagian yang sering mendapat kritik negatif, terutama fitur transaksi dan transfer. Untuk mengatasi hal ini, perbaikan pada fitur transaksi dan transfer menjadi prioritas utama karena aktivitas ini sangat penting dalam aplikasi. Hal ini diharapkan dapat meningkatkan pengalaman pengguna dan mengurangi kemungkinan adanya ulasan negatif.
- 2) Optimalisasi Akses Pengguna: Aspek Akses juga memiliki tingkat kritik negatif yang signifikan sehingga perlu ditingkatkan. OVO harus memperbaiki proses *login* dan verifikasi yang seringkali mengalami masalah, sehingga pengguna dapat dengan mudah mengakses akun mereka secara lancar dan aman.
- 3) Meningkatkan pelayanan Pelanggan: Peningkatan dalam Aspek Pelayanan menjadi fokus penting. Pengguna mengharapkan respon cepat dan bantuan yang memadai ketika menghadapi masalah. Selain itu, mengurangi biaya admin pada aktivitas transfer juga dapat meningkatkan kepuasan pengguna dan mengurangi kritik negatif terkait aspek pelayanan.
- 4) Memastikan Keamanan Data Pengguna: Aspek Keamanan mendapatkan respon positif secara keseluruhan, namun sistem keamanan pada OVO tetap harus selalu ditingkatkan karena hal tersebut sangat penting bagi pengguna dengan cara meningkatkan dan mengikuti perkembangan teknologi keamanan guna menjaga keamanan data para pengguna.

Dengan mengimplementasikan langkah-langkah ini, maka OVO dapat meningkatkan kualitas layanan dan mengurangi tingkat *technostress* yang dirasakan oleh pengguna, sehingga dapat meningkatkan kepuasan mereka.

## DAFTAR PUSTAKA

- Abdillah, L. A. (2019). Proceeding: International Conference on Communication, Information Technology and Youth Study (I-CITYS 2019) (eISBN:978-967-17343-5-3) Bayview Hotel Melaka, Malaysia AN OVERVIEW OF INDONESIAN FINTECH APPLICATION. *Proceeding: International Conference on Communication, Information Technology and Youth Study, Figure 1*, 8–16.
- AFTECH, M. (2021). *Annual Members Survey 2*.
- Albalawi, R., Yeap, T. H., & Benyoucef, M. (2020). *Using Topic Modeling Methods for Short-Text Data: A Comparative Analysis*. 3(July), 1–14. <https://doi.org/10.3389/frai.2020.00042>
- Alderman, A. K., & Salem, B. (2010). Survey research. *Plastic and Reconstructive Surgery*, 126(4), 1381–1389. <https://doi.org/10.1097/PRS.0b013e3181ea44f9>
- Amalia, P. R., & Winarko, E. (2021). *Aspect-Based Sentiment Analysis on Indonesian Restaurant Review Using a Combination of Convolutional Neural Network and Contextualized Word Embedding*. 15(3).
- Amoualian, H., Lu, W., Gaussier, E., Balikas, G., Amini, M. R., & Clausel, M. (2017). Topical coherence in LDA-based models through induced segmentation. *ACL 2017 - 55th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference (Long Papers), 1*, 1799–1809. <https://doi.org/10.18653/v1/P17-1165>
- Annur, C. M. (2021). Survei: OVO Rajai Pangsa Pasar E-Wallet Indonesia pada 2020. *Katadata.co.id*, 2025.
- Arner, D. W. (2015). *The Evolution of FinTech: A New Post-Crisis Paradigm?* 1–45.
- Arthamevia, N. P. (2014). *Aspect-Based Sentiment Analysis in Beauty Product Reviews Using TF-IDF and SVM Algorithm*.

- Astuti, S. P. (2020). *Analisis Sentimen Berbasis Aspek Pada Aplikasi Tokopedia Menggunakan LDA dan Naïve Bayes*.
- Barroso, M., & Laborda, J. (2022). Digital transformation and the emergence of the Fintech sector : Systematic literature review. *Digital Business*, 2(2), 100028. <https://doi.org/10.1016/j.digbus.2022.100028>
- Blei, D. M., Ng, A. Y., & Jordan, M. T. (2002). Latent dirichlet allocation. *Advances in Neural Information Processing Systems*, 3, 993–1022.
- Chou, H., & Chou, C. (2021). Computers & Education A multigroup analysis of factors underlying teachers ' technostress and their continuance intention toward online teaching. *Computers & Education*, 175(March), 104335. <https://doi.org/10.1016/j.compedu.2021.104335>
- Christanto, B., & Setiabudi, D. H. (2020). Penerapan Random Forest dalam Email Filtering untuk Mendeteksi Spam. *Jurnal Infra*, 8(2).
- D'Aniello, G., Gaeta, M., & La Rocca, I. (2022). KnowMIS-ABSA: an overview and a reference model for applications of sentiment analysis and aspect-based sentiment analysis. Dalam *Artificial Intelligence Review* (Vol. 55, Nomor 7). Springer Netherlands. <https://doi.org/10.1007/s10462-021-10134-9>
- databoks. (2022). *KIC : E-Wallet Paling Sering Digunakan untuk Transaksi E-Commerce*. 2022.
- Fatih Gurcan, Ozcan Ozyurt, N. E. C. (2022). *View of Investigation of Emerging Trends in the E-Learning Field Using Latent Dirichlet Allocation | The International Review of Research in Open and Distributed Learning*.
- H Kara, O. A. M. A. (2014). Buku FIntech Wisnu Wingah. Dalam *Paper Knowledge . Toward a Media History of Documents* (Vol. 7, Nomor 2).
- Hasanli, H., & Rustamov, S. (2019). Sentiment Analysis of Azerbaijani tweets Using Logistic Regression, Naive Bayes and SVM. *13th IEEE International Conference on Application of Information and Communication Technologies, AICT 2019 - Proceedings*. <https://doi.org/10.1109/AICT47866.2019.8981793>

- Hevner, A. R. (2004). Design science 97. *AI and Society*, 10(2), 199–217.  
<https://doi.org/10.1007/BF01205282>
- Hoang, M., Alija Bihorac, O., & Rouces, J. (2019). Aspect-Based Sentiment Analysis Using BERT. *Proceedings of the 22nd Nordic Conference on Computational Linguistics*, 187–196.
- Irfan, L., Hussain, S., Ayoub, M., Yu, Y., & Khan, A. (2022). *A Comparative Analysis of Social Communication Applications using Aspect Based Sentiment Analysis*. 44–50.
- Kalepalli, Y. (2020). *Effective Comparison of LDA with LSA for Topic Modelling*. *Iciccs*, 1245–1250.
- Kompasiana. (2021). *Apa Bedanya Google Play Store dengan Google Store?*
- Kusuma Dewi, H. (2023). *Selamat! Ini Pemenang Promo OVO Februari 2023 Berhadiah Reksadana hingga Rp300 Ribu*.  
<https://www.bareksa.com/berita/promo/2023-03-21/selamat-ini-pemenang-promo-ovo-februari-2023-berhadiah-reksadana-hingga-rp300-ribu>
- Lee, Y. (2021). *Impacts of Digital Technostress and Digital Technology Self-Efficacy on Fintech Usage Intention of Chinese Gen Z Consumers*.
- Lee, Y. (2022). *Higher innovativeness , lower technostress ? : comparative study of determinants on FinTech usage behavior between Korean and Chinese Gen Z consumers*. <https://doi.org/10.1108/APJML-05-2022-0402>
- Malviya, S., Tiwari, A. K., Srivastava, R., & Tiwari, V. (2020). Machine Learning Techniques for Sentiment Analysis: A Review. *SAMRIDDHI: A Journal of Physical Sciences, Engineering and Technology*, 12(02), 72–78.
- Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications : A survey. *Ain Shams Engineering Journal*, 5(4), 1093–1113. <https://doi.org/10.1016/j.asej.2014.04.011>

- Mike Thelwall, Kevan Buckley, G. P. (2013). Sentiment Strength Detection in Short Informal Text. *Journal of the American Society for Information Science and Technology*, 64(July), 1852–1863. <https://doi.org/10.1002/asi>
- Moreira-Santos, D., Au-Yong-Oliveira, M., & Palma-Moreira, A. (2022). Fintech Services and the Drivers of Their Implementation in Small and Medium Enterprises. *Information (Switzerland)*, 13(9). <https://doi.org/10.3390/info13090409>
- Nizar. (2020). Financial Technology ( Fintech ): It ' s Concept and Implementation in Indonesia. *Munich Personal RePEc Archive*, 5(98486), 4–10.
- Mustakim, H., & Priyanta, S. (2022). *Aspect-Based Sentiment Analysis of KAI Access Reviews Using NBC and SVM*. 16(2), 113–124. <https://doi.org/10.22146/ijccs.68903>
- Nakashima, T. (2018). Creating credit by making use of mobility with FinTech and IoT. *IATSS Research*, 42(2), 61–66. <https://doi.org/10.1016/j.iatssr.2018.06.001>
- Nanggala, A. Y. A. (2020). Use of fintech for payment : Approach to technology acceptance model modified. *Journal of Contemporary Information Technology, Management, and Accounting*, 1(1), 1–8.
- National Economic Council. (2017). *A Framework for FinTech*. January, 13.
- Norulkamar, U., & Ahmad, U. (2009). *The Impact of Technostress on Organisational Commitment among Malaysian Academic Librarians*. 38, 103–123.
- Pajankar, A., & Joshi, A. (2022). Introduction to Machine Learning with Scikit-learn. Dalam *Hands-on Machine Learning with Python*. [https://doi.org/10.1007/978-1-4842-7921-2\\_5](https://doi.org/10.1007/978-1-4842-7921-2_5)
- Paramahita, K. (2022). *Nikmati Promo OVO Shoptakuler Juli 2022, Bayar Tagihan Dapat Cashback Rp 200.000*.



<https://www.momsmoney.id/news/nikmati-promo-ovo-shoptakuler-juli-2022-bayar-tagihan-dapat-cashback-rp-200000>

- Pedro, J. (2022). *Understanding Topic Coherence Measures*. <https://towardsdatascience.com/understanding-topic-coherence-measures-4aa41339634c>
- Pinem, F. J., Andreswari, R., & Hasibuan, M. A. (2018). Sentiment analysis to measure celebrity endorsement's effect using support vector machine algorithm. *International Conference on Electrical Engineering, Computer Science and Informatics (EECSI), 2018-October*, 690–695. <https://doi.org/10.1109/EECSI.2018.8752687>
- Pradha, S., Halgamuge, M. N., & Tran Quoc Vinh, N. (2019). Effective text data preprocessing technique for sentiment analysis in social media data. *Proceedings of 2019 11th International Conference on Knowledge and Systems Engineering, KSE 2019*, 1–8. <https://doi.org/10.1109/KSE.2019.8919368>
- Putra, H. R., & Sfenrianto. (2020). Analysis of customer satisfaction factors on e-commerce payment system methods in Indonesia. *International Journal of Advanced Computer Science and Applications*, 11(4), 471–480. <https://doi.org/10.14569/IJACSA.2020.0110463>
- Qader, W. A., Ameen, M. M., & Ahmed, B. I. (2019). An Overview of Bag of Words; Importance, Implementation, Applications, and Challenges. *Proceedings of the 5th International Engineering Conference, IEC 2019*, 200–204. <https://doi.org/10.1109/IEC47844.2019.8950616>
- Qaiser, S., & Ali, R. (2018). Text Mining: Use of TF-IDF to Examine the Relevance of Words to Documents. *International Journal of Computer Applications*, 181(1), 25–29. <https://doi.org/10.5120/ijca2018917395>
- Santra, a. K., & Christy, C. J. (2012). Genetic Algorithm and Confusion Matrix for Document Clustering. *International Journal of Computer Science*, 9(1), 322–328.

- Silaen, E., & Prabawani, B. (2020). *PERSEPSI MANFAAT SERTA PROMOSI TERHADAP MINAT BELI ULANG SALDO E-WALLET OVO*. 1–9.
- Tacy, J. W. (2015). *Technostress Effects on Technology Acceptance by Nurse Faculty*.
- Utami, S. H., Purnama, A. A., & Hidayanto, A. N. (2022). Fintech Lending in Indonesia: A Sentiment Analysis, Topic Modelling, and Social Network Analysis using Twitter Data. *International Journal of Applied Engineering and Technology (London)*, 4(1), 50–56.
- Wahid, D. H. (2017). *Peringkasan Sentimen Esktraktif di Twitter Menggunakan Hybrid TF-IDF dan Cosine Similarity*. 10(2), 207–218.
- Warjiyono, Aji, S., Fandhilah, Hidayatun, N., Faqih, H., & Liesnaningsih. (2019). The Sentiment Analysis of Fintech Users Using Support Vector Machine and Particle Swarm Optimization Method. *2019 7th International Conference on Cyber and IT Service Management, CITSM 2019*. <https://doi.org/10.1109/CITSM47753.2019.8965348>
- Zahoor, K., Bawany, N. Z., & Hamid, S. (2020). Sentiment analysis and classification of restaurant reviews using machine learning. *Proceedings - 2020 21st International Arab Conference on Information Technology, ACIT 2020*. <https://doi.org/10.1109/ACIT50332.2020.9300098>