

**KNOWLEDGE DISCOVERY MELALUI PEMODELAN TOPIK PADA
ULASAN PENGGUNA APLIKASI GOPARTNER MENGGUNAKAN
BERTOPIC, LDA, DAN NMF**

SKRIPSI

Sebagai salah satu syarat untuk menyelesaikan studi
di Program Studi Sistem Informasi S1



Oleh :

Metti Detricia Pratiwi

09031282126102

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS ILMU KOMPUTER

UNIVERSITAS SRIWIJAYA

2024

LEMBAR PENGESAHAN

**KNOWLEDGE DISCOVERY MELALUI PEMODELAN TOPIK PADA
ULASAN PENGGUNA APLIKASI GOPARTNER MENGGUNAKAN
BERTOPIC, LDA, DAN NMF**

SKRIPSI

Sebagai salah satu syarat untuk menyelesaikan studi
di Program Studi Sistem Informasi S1

Oleh

Metti Detricia Pratiwi

09031282126102

Palembang, 23 Desember 2024

Mengetahui,
Ketua Jurusan Sistem Informasi,



Ahmad Rifai, S.T., M.T.
NIP. 197910202010121003

Pembimbing,

A black ink signature of Ken Ditha Tania, consisting of several overlapping loops.

Ken Ditha Tania, M.Kom., PhD.
NIP. 198507182012122003

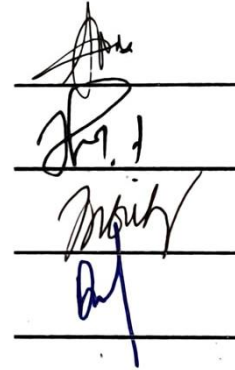
HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari : Jumat
Tanggal : 27 Desember 2024
Nama : Metti Detricia Pratiwi
NIM : 09031282126102
Judul Tugas Akhir : *Knowledge Discovery* Melalui Pemodelan Topik pada
Ulasan Pengguna Aplikasi GoPartner Menggunakan
BERTopic, LDA, dan NMF

Komisi Penguji :

1. Pembimbing : Ken Ditha Tania, M.Kom., PhD.
2. Ketua : Ahmad Rifai, M.T.
3. Sekretaris : Nabila Rizky Oktadini, M.T.
4. Penguji : Dedy Kurniawan, M.Sc.



Mengetahui,
Ketua Jurusan Sistem Informasi



Ahmad Rifai, M.T.
NIP. 97010202010121003

HALAMAN PERNYATAAN BEBAS PLAGIAT

Saya yang bertanda tangan di bawah ini :

Nama : Metti Detricia Pratiwi
NIM : 09031282126102
Program Studi : Sistem Informasi Reguler
Judul Skripsi : *Knowledge Discovery* Melalui Pemodelan Topik Pada
Ulasan Pengguna Aplikasi GoPartner Menggunakan
BERTopic, LDA, dan NMF

Hasil Pengecekan *Software iThenticate/Turnitin*: 3%

Menyatakan bahwa laporan skripsi saya merupakan hasil karya saya sendiri bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan skripsi ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya dengan ketentuan yang berlaku.

Demikian pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan dari pihak manapun.



Indralaya, 23 Desember 2024



Metti Detricia Pratiwi
NIM. 09031282126102

HALAMAN PERSEMBAHAN DAN MOTTO

“Never compare your life with others, like the sun and the moon, they shine when their time comes”

“...Remember me, and I will remember you...” (Qur'an 2:152)

“You are too young to let the world break you” V of BTS

Skripsi ini saya persembahkan untuk:

- ❖ **Diriku sendiri, Metti Detricia Pratiwi**
- ❖ **Ayah, Bunda, Adik, dan Seluruh Keluarga Besar**
- ❖ **Dosen Pembimbing Skripsi dan Dosen Pembimbing Akademik**
- ❖ **Sahabat dan rekan-rekan seperjuangan penulis**
- ❖ **Almamater yang aku banggakan, Universitas Sriwijaya**

KATA PENGANTAR

Puji syukur kehadirat Allah SWT yang telah melimpahkan rahmat, karunia, serta hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **“*Knowledge Discovery Melalui Pemodelan Topik Pada Ulasan Pengguna Aplikasi GoPartner Menggunakan BERTopic, LDA, dan NMF*”** sebagai salah satu syarat untuk menyelesaikan pendidikan S1 Program Studi Sistem Informasi, Fakultas Ilmu Komputer, Universitas Sriwijaya.

Penulisan Tugas Akhir ini tidak lepas dari doa, bimbingan, bantuan, dan dukungan baik secara langsung ataupun secara tidak langsung dari berbagai pihak. Oleh karena itu pada kesempatan ini penulis ingin menyampaikan ungkapan terima kasih yang sebesar-besarnya kepada:

1. Ayah, Bunda, Adik, dan keluarga besar penulis yang selalu memberikan doa, semangat, dan dukungan baik secara materi ataupun moril sehingga penulis dapat berada di titik sekarang menyelesaikan perkuliahan dan Tugas Akhir ini.
2. Bapak Prof. Dr. Erwin, S.Si., M.Si. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya.
3. Bapak Ahmad Rifai, M.T. selaku Ketua Jurusan Sistem Informasi Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Ibu Ken Ditha Tania, M.Kom., PhD. selaku Dosen Pembimbing Tugas Akhir yang telah banyak memberikan bantuan dan bimbingan kepada penulis selama proses pengerjaan tugas akhir ini sehingga penulis dapat menyelesaikan tugas akhir ini dengan baik dan lancar.

5. Ibu Putri Eka Sevdiyuni, S.Si., M.T. selaku Dosen Pembimbing Akademik yang telah membimbing penulis mulai dari awal hingga akhir semester.
6. Bapak Dedy Kurniawan, M.Sc. selaku Dosen Penguji, Ibu Nabila Rizky Oktadini, M.T. selaku Sekretaris Sidang, dan Bapak Ahmad Rifai, M.T. selaku Ketua Sidang.
7. Seluruh Dosen Fakultas Ilmu Komputer Universitas Sriwijaya yang telah berbagi ilmu pengetahuan serta pengalaman berharga selama masa perkuliahan.
8. Seluruh Staf dan Karyawan Fakultas Ilmu Komputer Universitas Sriwijaya yang telah membantu dan mempermudah kegiatan administratif selama masa perkuliahan.
9. Sahabat penulis sejak di bangku SMP, Jesicha Andini dan Nur Fira Syafiyah, yang selalu memberikan energi positif.
10. Sahabat yang sudah seperti keluarga bagi penulis, Yulia Astriani, Safira Nida Munzilina, dan Dila Okta Dwi Putri. Terimakasih sudah menjadi orang berhati baik dan selalu ada di sisi penulis.
11. Rekan-rekan seperjuangan yang selalu berbagi informasi selama perkuliahan hingga penyusunan tugas akhir, yaitu Alfirah, Dea, Cindy, Yohana, dan teman-teman SIREG C 2021 lainnya. Terimakasih atas kebersamaan dan segala kebaikannya.
12. Semua orang baik yang tidak dapat penulis sebutkan namanya satu per satu, terimakasih atas semua bantuan dan segala ucapan penyemangat yang telah diberikan kepada penulis.

13. Terakhir, untuk diriku yang sangat berharga, terimakasih sudah bertahan dan selalu mau mencoba, mengusahakan yang terbaik, dan percaya bahwa dari semua hal sulit yang dilalui pasti ada hal indah dan kebahagiaan yang menanti di akhir. *Thank you for always trying to be the best version of yourself*♥.

Dalam penulisan Tugas Akhir ini, penulis menyadari bahwa laporan ini masih jauh dari kata sempurna, maka dari itu segala kritik dan saran yang membangun sangat diperlukan untuk menyempurnakan penulisan Tugas Akhir ini. Penulis juga berharap Tugas Akhir ini dapat memberikan manfaat untuk menambah ilmu pengetahuan serta berguna bagi penelitian selanjutnya.

Indralaya, 23 Desember 2024



Metti Detricia Pratiwi
NIM. 09031282126102

**KNOWLEDGE DISCOVERY MELALUI PEMODELAN TOPIK PADA
ULASAN PENGGUNA APLIKASI GOPARTNER MENGGUNAKAN
BERTOPIC, LDA, DAN NMF**

Oleh :

Metti Detricia Pratiwi

09031282126102

ABSTRAK

Layanan transportasi dan pesan-antar makanan merupakan salah satu sektor penggerak ekonomi digital di Indonesia. Laporan e-Conomy SEA 2023 menunjukkan bahwa sektor transportasi dan layanan pesan-antar makanan mengalami penurunan GMV pada tahun 2023 sebesar 8% dari tahun sebelumnya. Penurunan GMV ini mengindikasikan adanya penurunan nilai transaksi di sektor jasa transportasi dan pesan-antar. GoPartner merupakan aplikasi yang dikembangkan oleh GoTo untuk membantu mitra driver dalam menjalankan berbagai layanan yang ada di dalam aplikasi Gojek yang merupakan salah satu aplikasi yang bergerak di sektor transportasi dan layanan pesan-antar makanan. Driver sebagai orang yang memberikan pelayanan secara langsung kepada konsumen tentunya menjadi salah satu faktor yang mempengaruhi perilaku konsumen dalam menggunakan jasa. Untuk mengetahui permasalahan yang dihadapi driver, penelitian ini melakukan *knowledge discovery* melalui pemodelan topik pada ulasan aplikasi GoPartner dengan menggunakan BERTopic, LDA, dan NMF yang masing-masing metode tersebut memiliki pendekatan yang berbeda. Berdasarkan hasil penelitian dan kualitas topik yang dihasilkan, BERTopic dan LDA memiliki kualitas yang lebih baik dalam menganalisis ulasan pengguna GoPartner.

Kata kunci: *Topic Modeling*, BERTopic, LDA, NMF, *Knowledge Discovery*

KNOWLEDGE DISCOVERY THROUGH TOPIC MODELING ON GOPARTNER USER REVIEWS USING BERTOPIC, LDA, AND NMF

By :

Metti Detricia Pratiwi

09031282126102

ABSTRACT

Transportation and food delivery services are one of the driving sectors of the digital economy in Indonesia. The e-Conomy SEA 2023 report shows that the transportation and food delivery services sector experienced a decrease in GMV in 2023 by 8% from the previous year. The decline in GMV indicates a decrease in transaction value in the transportation and food delivery service sector. GoPartner is an application developed by GoTo to assist driver partners in carrying out various services in the gojek application which is one of the applications engaged in the transportation sector and food delivery services. Drivers as people who provide services directly to consumers are certainly one of the factors that influence customer behavior in using services. To find out the problems faced by drivers, this research conducts knowledge discovery through topic modeling on GoPartner application reviews using BERTopic, LDA, and NMF, each of these methods has a different approach. Based on the research results and the quality of the topics generated, BERTopic and LDA have better quality in analyzing GoPartner user reviews.

Keywords: Topic Modeling, BERTopic, LDA, NMF, Knowledge Discovery

DAFTAR ISI

LEMBAR PENGESAHAN	ii
HALAMAN PERSETUJUAN	iii
HALAMAN PERNYATAAN BEBAS PLAGIAT	iv
HALAMAN PERSEMBAHAN DAN MOTTO	v
KATA PENGANTAR	vi
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	xi
DAFTAR GAMBAR	xiii
DAFTAR TABEL	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	5
1.3 Tujuan Penelitian.....	6
1.4 Manfaat Penelitian.....	6
1.5 Batasan Masalah.....	6
BAB II TINJAUAN PUSTAKA	8
2.1 <i>Knowledge Discovery</i>	8
2.2 GoPartner.....	9
2.3 <i>Text Mining</i>	10
2.4 <i>Natural Language Processing (NLP)</i>	10
2.5 <i>Web Scraping</i>	11
2.6 <i>Topic Modeling</i>	11
2.6.1 BERTopic	12
2.6.2 <i>Latent Dirichlet Allocation (LDA)</i>	12
2.6.3 <i>Non-Negative Matrix Factorization (NMF)</i>	13
2.7 <i>Coherence Measure</i>	13
2.8 Penelitian Terdahulu.....	14
BAB III METODOLOGI PENELITIAN	17
3.1 Objek Penelitian	17
3.2 Tahapan Penelitian	17

3.2.1 Pengumpulan Data	17
3.2.2 Pembersihan Data.....	18
3.2.3 <i>Preprocessing Data</i>	18
3.2.3.1 <i>Case Folding</i>	18
3.2.3.2 <i>Tokenizing</i>	19
3.2.3.3 <i>Stopword</i>	20
3.2.3.4 <i>Stemming</i>	21
3.2.4 Penerapan <i>Topic Modeling</i>	22
3.2.4.1 BERTopic	22
3.2.4.2 LDA.....	23
3.2.4.3 NMF.....	24
3.2.5 Evaluasi	25
BAB IV HASIL DAN PEMBAHASAN	26
4.1 Hasil Proses Pengumpulan Data.....	26
4.2 Hasil Proses Pembersihan Data.....	26
4.3 Hasil <i>Preprocessing Data</i>	27
4.3.1 Hasil <i>Case Folding</i>	27
4.3.2 Hasil <i>Tokenizing</i>	27
4.3.3. Hasil <i>Stopword</i>	28
4.3.4 Hasil <i>Stemming</i>	28
4.4 Hasil Evaluasi.....	29
4.5 Hasil Implementasi Pemodelan Topik (<i>Topic Modeling</i>)	29
4.5.1 Hasil Pemodelan Topik Menggunakan BERTopic.....	30
4.5.2 Hasil Pemodelan Topik Menggunakan LDA.....	32
4.5.3 Hasil Pemodelan Topik Menggunakan NMF	34
BAB V KESIMPULAN DAN SARAN	36
5.1 Kesimpulan.....	36
5.2 Saran	37
DAFTAR PUSTAKA	38

DAFTAR GAMBAR

Gambar 1. 1	Laporan e-Conomy SEA 2023.....	3
Gambar 1. 2	Rating Aplikasi GoPartner di Google Play Store	4
Gambar 2. 1	Lima Langkah Utama KDD (Głowania et al., 2023)	8
Gambar 2. 2	Logo Aplikasi GoPartner	9
Gambar 2. 3	Perhitungan Nilai Koherensi (Syahrial & Afidh, 2024)	14
Gambar 3. 1	Alur Penelitian	17
Gambar 3. 2	<i>Library Python Google Play Scaper</i>	18
Gambar 3. 3	Kode Program <i>Case Folding</i>	19
Gambar 3. 4	Instruksi Penginstalan <i>Library nltk</i>	19
Gambar 3. 5	Kode Program <i>Tokenizing</i>	19
Gambar 3. 6	Instruksi Penginstalan <i>Library Sastrawi</i>	20
Gambar 3. 7	Kode Program <i>Stopword</i>	20
Gambar 3. 8	Kode Program <i>Stemming</i>	21
Gambar 3. 9	Instruksi Penginstalan <i>Library BERTopic</i>	22
Gambar 3. 10	Kode Program <i>BERTopic</i>	23
Gambar 3. 11	Parameter <i>Gridsearch LDA</i>	23
Gambar 3. 12	Parameter <i>Gridsearch NMF</i>	24
Gambar 3. 13	Kode Program <i>Coherence Model</i>	25
Gambar 4. 1	Hasil <i>Scraping Data</i>	26
Gambar 4. 2	Hasil Tahapan <i>Case Folding</i>	27
Gambar 4. 3	Hasil Tahapan <i>Tokenizing</i>	28
Gambar 4. 4	Hasil Tahapan <i>Stopword</i>	28
Gambar 4. 5	Hasil Tahapan <i>Stemming</i>	28
Gambar 4. 6	Perbandingan Nilai Koherensi.....	29

DAFTAR TABEL

Tabel 2. 1 Penelitian Terdahulu.....	14
Tabel 3. 1 Contoh <i>Case Folding</i>	19
Tabel 3. 2 Contoh <i>Tokenizing</i>	20
Tabel 3. 3 Contoh <i>Stopword</i>	21
Tabel 3. 4 Contoh <i>Stemming</i>	21
Tabel 4. 1 Hasil BERTopic dengan <i>Tokenizing Data</i>	30
Tabel 4. 2 Hasil BERTopic dengan <i>Stemming Data</i>	31
Tabel 4. 3 Hasil LDA dengan <i>Tokenizing Data</i>	33
Tabel 4. 4 Hasil LDA dengan <i>Stemming Data</i>	34
Tabel 4. 5 Hasil NMF dengan <i>Tokenizing Data</i>	35
Tabel 4. 6 Hasil NMF dengan <i>Stemming Data</i>	35

DAFTAR LAMPIRAN

Lampiran 1 Surat Kesiediaan Membimbing.....	A-1
Lampiran 2 Form Pengajuan Topik Skripsi	B-1
Lampiran 3 Surat Keputusan Tugas Akhir (SK TA).....	C-1
Lampiran 4 <i>Dataset</i> Awal Penelitian	D-1
Lampiran 5 <i>Dataset</i> Setelah <i>Preprocessed</i>	E-1
Lampiran 6 Bukti Hasil Penerapan <i>Topic Modeling</i>	F-1
Lampiran 7 Bukti Proses Pengumpulan Artikel Jurnal	G-1
Lampiran 8 Bukti <i>Letter of Acceptance</i> (LoA) Jurnal.....	H-1
Lampiran 9 Hasil Pengecekan Turnitin.....	I-1
Lampiran 10 Form Nilai <i>Desk Evaluation</i>	J-1
Lampiran 11 Form Perbaikan Ujian Komprehensif	K-1
Lampiran 12 Kartu Konsultasi	L-1

BAB I

PENDAHULUAN

1.1 Latar Belakang

Beragam terobosan dalam ilmu pengetahuan, teknologi informasi dan komunikasi telah memberikan dorongan terhadap laju pertumbuhan ekonomi digital di Indonesia. Mengacu pada laporan e-Conomy SEA 2023, nilai ekonomi digital Indonesia mencapai USD 82 miliar atau meningkat sebesar 8% dari tahun sebelumnya dan diperkirakan akan terus mengalami kenaikan hingga mencapai *Gross Merchandise Value* (GMV) sekitar USD 110 miliar pada tahun 2025. Kenaikan angka tersebut berasal dari lima sektor yaitu *e-commerce*, perjalanan *online*, transportasi dan layanan pengantaran makanan, media *online*, serta layanan keuangan digital. Berkembangnya sektor-sektor penggerak ekonomi digital tersebut tentu tidak terlepas oleh pengaruh penggunaan internet dan smartphone di kalangan masyarakat yang telah membuat permintaan akan layanan berbasis aplikasi online menjadi meningkat dan terciptanya lapangan kerja (Iqbal & Bhatti, 2020).

Salah satu entitas yang menjadi penggerak dalam ekonomi digital Indonesia adalah PT GoTo Gojek Tokopedia Tbk (GoTo) yang merupakan hasil merge dari dua perusahaan teknologi di Indonesia, Gojek dan Tokopedia. GoTo menyediakan layanan berbasis aplikasi termasuk *e-commerce*, transportasi, dan pesan-antar makanan. Aplikasi Gojek adalah platform yang dimiliki GoTo untuk memenuhi kebutuhan masyarakat Indonesia terhadap layanan transportasi dan pengantaran makanan ataupun barang. Adanya Gojek memberikan lapangan pekerjaan bagi masyarakat yang belum memiliki pekerjaan tetap untuk memperoleh penghasilan

dengan menjadi driver Gojek melalui aplikasi GoPartner (Fakhriyah, 2020). GoPartner adalah aplikasi yang dikelola oleh GoTo untuk memudahkan mitra driver gojek dalam menerima pesanan, memantau pendapatan, dan mengakses informasi. GoPartner dan Gojek sangat terhubung, GoPartner dirancang sebagai alat komunikasi mitra driver Gojek yang menghubungkannya dengan pelanggan dari aplikasi Gojek. Melalui Gojek pengguna dapat melakukan transaksi dengan berbagai macam layanan yang ditawarkan seperti GoRide, GoCar, GoFood, GoMart, dan sebagainya.

Keselarasan kedua aplikasi tersebut dalam menghubungkan driver Gojek dan pelanggan berpengaruh pada kepuasan terhadap penggunaan aplikasi dan layanan yang diberikan. Kualitas pelayanan yang diberikan oleh driver Gojek menjadi salah satu faktor yang mempengaruhi kepuasan pelanggan terhadap aplikasi Gojek (Purwanti et al., 2022). Terpenuhinya kepuasan pelanggan pengguna Gojek dapat memberikan nilai jangka panjang bagi perusahaan. Di sisi lain, kepuasan driver Gojek juga menjadi hal yang tidak kalah penting. Keluhan para driver Gojek terhadap aplikasi GoPartner sebagai fasilitas yang disediakan GoTo bagi driver untuk melakukan pekerjaannya adalah aspek yang perlu dikaji. Hal ini didukung dengan data pada laporan e-Conomy SEA 2023 yang menunjukkan bahwa dari beberapa sektor yang ada, sektor transportasi dan layanan pengantaran makanan mengalami penurunan GMV di tahun 2023 sebesar 8% dari tahun sebelumnya. Fakta ini mengindikasikan adanya perubahan dalam perilaku pembelian konsumen. Perubahan tersebut tentunya melibatkan segala faktor yang berhubungan dengan ekosistem di dalam layanan dan driver adalah salah satunya. Hambatan atau kemudahan yang dirasakan driver Gojek ketika menggunakan aplikasi GoPartner

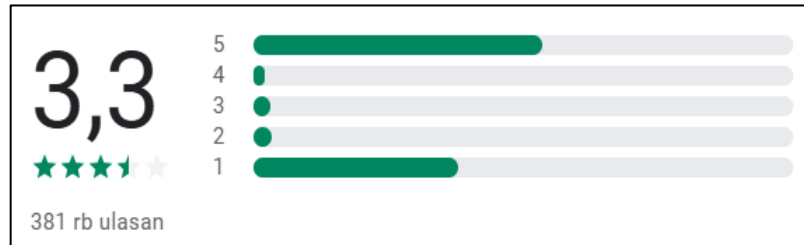
tentu akan mempengaruhi pengalaman pengguna dalam bertransaksi menggunakan aplikasi Gojek.



Gambar 1.1 Laporan e-Conomy SEA 2023

Umumnya tanggapan driver Gojek mengenai aplikasi GoPartner dapat diberikan melalui kolom ulasan aplikasi GoPartner di platform *Google Play Store*. Ulasan yang diberikan dapat menjadi informasi yang mendeskripsikan pengalaman sesungguhnya dari pengguna saat menggunakan aplikasi (Khoirul et al., 2023). Ulasan pengguna juga menjadi sumber data yang penting dalam penemuan pengetahuan atau *knowledge discovery*. *Knowledge discovery* adalah proses sistematis untuk mengekstraksi informasi dan wawasan yang bermakna dari data besar dan kompleks. Proses ini melibatkan berbagai tahapan, mulai dari pengumpulan data, preprocessing data, penerapan teknik analitik, *machine learning*, atau eksplorasi visualisasi. Selain ulasan, *rating* juga merupakan hal yang berpengaruh dalam banyak aspek. *Rating* pada aplikasi menggambarkan nilai kepuasan yang diberikan pengguna selama menggunakan aplikasi (Kapoor & Vij,

2020). Aplikasi GoPartner telah diunduh lebih dari 5 juta kali dengan *rating* 3,3 dan 381 ribu lebih ulasan.



Gambar 1. 2 *Rating* Aplikasi GoPartner di *Google Play Store*

Seiring dengan banyaknya jumlah ulasan dan perbedaan jumlah rating antara bintang 5 dan bintang 1 yang tidak terlalu jauh, diperlukan teknik atau metode yang mampu untuk mengorganisasikan informasi yang terdapat dalam ulasan-ulasan tersebut. Kata kunci yang sering muncul dari ulasan pengguna menjadi kata-kata yang mewakili sebuah topik (Maarif, 2022). Salah satu teknik yang bisa digunakan adalah pemodelan topik atau *topic modeling* yang dianggap sebagai metodologi untuk menyajikan volume data yang sangat besar, konsep yang tersembunyi dan fitur yang menonjol atau laten variabel data. Melalui pemodelan topik dapat dihasilkan topik-topik tersembunyi yang ada pada ulasan pengguna aplikasi GoPartner. Beberapa metode seperti *Latent Dirichlet Allocation* (LDA) yang merupakan model statistik generatif, *Non-Negative Matrix Factorization* (NMF) yang melakukan ekstraksi topik dengan pendekatan aljabar linier, serta BERTopic dan Top2Vec dengan pendekatan *embedding* dapat digunakan untuk melakukan pemodelan topik (Egger & Yu, 2022). Pada LDA yang merupakan sebuah model probabilistik, topik-topik laten ditemukan di dalam koleksi dokumen dengan cara menganalisa distribusi kata (Ying, 2024). NMF adalah model statistik yang memfaktorkan atau menguraikan matriks input non-negatif menjadi dua sub-

matriks non-negatif (Meaney et al., 2022). Sementara itu, BERTopic menggunakan *Bidirectional Encoder Representation from Transformers* (BERT) *embedding* dan algoritma *clustering* untuk menghasilkan representasi topik yang bermakna secara semantik (A. Kumar et al., 2024).

Pada beberapa penelitian sebelumnya, LDA merupakan metode yang paling sering digunakan untuk pemodelan topik, seperti pada penelitian yang dilakukan oleh Mutmainah *et al* (2023) menerapkan BiLSTM dan LDA untuk analisis sentimen dan pemodelan topik pada aplikasi telemedicine di Google Play Store yang menghasilkan topik dengan dengan nilai koherensi 0.6437 pada sentimen positif, 0.6132 pada sentimen negatif, dan 0.6296 pada sentimen netral (Mutmainah et al., 2023). Penelitian lainnya yang dilakukan oleh (Egger & Yu, 2022) membandingkan beberapa metode pemodelan topik seperti LDA, NMF, BERTopic, Top2Vec dengan hasil keefektifan penggunaan BERTopic dan NMF dalam menganalisis data Twitter. Oleh karena itu berdasarkan uraian di atas, penulis tertarik untuk mengangkat permasalahan tersebut sebagai tugas akhir yang berjudul **“*Knowledge Discovery Melalui Pemodelan Topik pada Ulasan Pengguna Aplikasi GoPartner Menggunakan BERTopic, LDA, dan NMF*”**.

1.2 Rumusan Masalah

Masalah yang akan dibahas berdasarkan latar belakang sebelumnya dapat dirumuskan sebagai berikut:

1. Bagaimana perbandingan nilai coherence dari topik yang dihasilkan oleh BERTopic, LDA, dan NMF?

2. Bagaimana kualitas dari *terms* yang dihasilkan oleh masing-masing metode?
3. Apa saja topik dan pengetahuan yang dapat diambil dan disimpulkan dari masing-masing metode?

1.3 Tujuan Penelitian

Penelitian ini diarahkan untuk dapat mencapai tujuan sebagai berikut:

1. Menerapkan metode BERTopic, LDA, dan NMF untuk melakukan pemodelan topik pada data ulasan pengguna aplikasi GoPartner.
2. Mengetahui perbandingan nilai coherence yang dihasilkan oleh BERTopic, LDA, and LDA.
3. Mengidentifikasi topik dan pengetahuan pada setiap topik yang dihasilkan oleh masing-masing metode.

1.4 Manfaat Penelitian

Adapun dampak positif yang diharapkan dapat tercapai dari penelitian ini adalah sebagai berikut:

1. Hasil dari penelitian ini diharapkan dapat berkontribusi sebagai wawasan mengenai penerapan pemodelan topik pada ulasan suatu aplikasi.
2. Hasil penelitian ini dapat memberikan pemahaman mengenai opini atau pendapat pengguna mengenai aplikasi GoPartner.

1.5 Batasan Masalah

Ruang lingkup yang menjadi batasan masalah dalam penelitian ini sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah ulasan aplikasi GoPartner dalam pada *Google Play Store*.
2. Ulasan yang digunakan adalah ulasan dari tanggal 29 Februari 2024 hingga 30 Agustus 2024.
3. *Tools* yang digunakan adalah *excel* dan *google colaboratory*.

DAFTAR PUSTAKA

- Blair, S. J., Bi, Y., & Mulvenna, M. D. (2020). Aggregated topic models for increasing social media topic coherence. *Applied Intelligence*, 50(1), 138–156. <https://doi.org/10.1007/s10489-019-01438-z>
- Boegershausen, J., Datta, H., Borah, A., & Stephen, A. T. (2022). Fields of Gold: Scraping Web Data for Marketing Insights. *Journal of Marketing*, 86(5), 1–20. <https://doi.org/10.1177/00222429221100750>
- Egger, R., & Yu, J. (2022). A Topic Modeling Comparison Between LDA, NMF, Top2Vec, and BERTopic to Demystify Twitter Posts. *Frontiers in Sociology*, 7. <https://doi.org/10.3389/fsoc.2022.886498>
- Fahlevvi, M. R., & Azhari. (2022). Topic Modeling on Online News Portal Using Latent Dirichlet Allocation (LDA). *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 16(4), 335. <https://doi.org/10.22146/ijccs.74383>
- Fakhriyah, P. (2020). *Pengaruh Layanan Transportasi Online (Gojek) Terhadap Perluasan Lapangan Kerja Bagi Masyarakat Di Kota Cimahi* (Vol. 3, Issue 1). <https://doi.org/https://doi.org/10.22460/comm-edu.v3i1.3719>
- George, L., & Sumathy, P. (2023). An integrated clustering and BERT framework for improved topic modeling. *International Journal of Information Technology (Singapore)*, 15(4), 2187–2195. <https://doi.org/10.1007/s41870-023-01268-w>
- Głowania, S., Kozak, J., & Juszczuk, P. (2023). Knowledge Discovery in Databases for a Football Match Result. *Electronics (Switzerland)*, 12(12). <https://doi.org/10.3390/electronics12122712>
- Gokcimen, T., & Das, B. (2024). Exploring climate change discourse on social media and blogs using a topic modeling analysis. *Heliyon*, 10(11). <https://doi.org/10.1016/j.heliyon.2024.e32464>
- Hutama, L. B., & Suhartono, D. (2022). Indonesian Hoax News Classification with Multilingual Transformer Model and BERTopic. *Informatika (Slovenia)*, 46(8), 81–90. <https://doi.org/10.31449/inf.v46i8.4336>
- Iqbal, S., & Bhatti, Z. A. (2020). A qualitative exploration of teachers' perspective on smartphones usage in higher education in developing countries. *International Journal of Educational Technology in Higher Education*, 17(1). <https://doi.org/10.1186/s41239-020-00203-4>
- Kapoor, A. P., & Vij, M. (2020). How to Boost your app Store Rating? An Empirical Assessment of Ratings for Mobile Banking Apps. *Journal of Theoretical and Applied Electronic Commerce Research*, 15(1), 99–115. <https://doi.org/10.4067/S0718-18762020000100108>
- Khoirul, M., Hayati, U., & Nurdiawan, O. (2023). Analisis Sentimen Aplikasi Brimo Pada Ulasan Pengguna Di Google Play Menggunakan Algoritma Naive Bayes. *Jurnal Mahasiswa Teknik Informatika*, 7(1).

- Khurana, D., Koli, A., Khatter, K., & Singh, S. (2023). Natural language processing: state of the art, current trends and challenges. *Multimedia Tools and Applications*, 82(3), 3713–3744. <https://doi.org/10.1007/s11042-022-13428-4>
- Kumar, A., Karamchandani, A., & Singh, S. (2024). Topic Modeling of Neuropsychiatric Diseases Related to Gut Microbiota and Gut Brain Axis Using Artificial Intelligence Based BERTopic Model on PubMed Abstracts. *Neuroscience Informatics*, 100175. <https://doi.org/10.1016/j.neuri.2024.100175>
- Kumar, S., Kar, A. K., & Ilavarasan, P. V. (2021). Applications of text mining in services management: A systematic literature review. In *International Journal of Information Management Data Insights* (Vol. 1, Issue 1). Elsevier Ltd. <https://doi.org/10.1016/j.jjime.2021.100008>
- Kwon, H. J., Ban, H. J., Jun, J. K., & Kim, H. S. (2021). Topic modeling and sentiment analysis of online review for airlines. *Information (Switzerland)*, 12(2), 1–14. <https://doi.org/10.3390/info12020078>
- Li, P., Tseng, C., Zheng, Y., Chew, J. A., Huang, L., Jarman, B., & Needell, D. (2022). Guided Semi-Supervised Non-Negative Matrix Factorization. *Algorithms*, 15(5). <https://doi.org/10.3390/a15050136>
- Maarif, M. R. (2022). Summarizing Online Customer Review using Topic Modeling and Sentiment Analysis. In *Jurnal Informatika Sunan Kalijaga* (Vol. 7, Issue 3). <https://doi.org/10.14421/jiska.2022.7.3.177-191>
- Matira, Y., & Setiawan, I. (2023). Pemodelan Topik pada Judul Berita Online Detikcom Menggunakan Latent Dirichlet Allocation. *Estimasi: Journal of Statistics and Its Application*, 4(1), 2721–379. <https://doi.org/10.20956/ejsa.vi.24843>
- Meaney, C., Escobar, M., Moineddin, R., Stukel, T. A., Kalia, S., Aliarzadeh, B., Chen, T., O'Neill, B., & Greiver, M. (2022). Non-negative matrix factorization temporal topic models and clinical text data identify COVID-19 pandemic effects on primary healthcare and community health in Toronto, Canada. *Journal of Biomedical Informatics*, 128. <https://doi.org/10.1016/j.jbi.2022.104034>
- Mutmainah, S., Fudholi, D. H., & Hidayat, S. (2023). Analisis Sentimen dan Pemodelan Topik Aplikasi Telemedicine Pada Google Play Menggunakan BiLSTM dan LDA. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 7(1), 312. <https://doi.org/10.30865/mib.v7i1.5486>
- Nurmalasari, S., Hidayanto, A. N., Huwaida, L. A., & Wulandari, H. (2023). Sentiment Analysis And Topic Modeling Of Citizen Satisfaction With The Indonesian Government In Handling A Pandemic. In *American Journal of Open Research* (Vol. 2, Issue 7). <https://opsearch.us/index.php/us/index>
- Ogunleye, B., Maswera, T., Hirsch, L., Gaudoin, J., & Brunson, T. (2023). Comparison of Topic Modelling Approaches in the Banking Context. *Applied*

Sciences (Switzerland), 13(2). <https://doi.org/10.3390/app13020797>

- Purwanti, W., Saifudin, A., & Risnawati, H. (2022). Faktor-Faktor Yang Mempengaruhi Kepuasan Konsumen Pada Pengguna Aplikasi Jasa Gojek di Bekasi. *JOURNAL OF MATERIALS PROCESSING TECHNOLOGY*, 29(1).
- Rocca, R., Tamagnone, N., Fekih, S., Contla, X., & Rekabsaz, N. (2023). *Natural language processing for humanitarian action: Opportunities, challenges, and the path toward humanitarian NLP*. <https://thedeep.io>
- Salsabila, K. D. A., & Trianasari, N. (2021). Analisis Persepsi Produk Kosmetik Menggunakan Metode Sentiment Analysis Dan Topic Modeling (Studi Kasus: Laneige Water Sleeping Mask). *Jurnal Teknologi Dan Manajemen Informatika*, 7(1). <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- Syahrial, S., & Afidh, R. P. F. (2024). Fine-Tuning Topic Modelling: A Coherence-Focused Analysis of Correlated Topic Models. *Infolitika Journal of Data Science*, 2(2), 82–87. <https://doi.org/10.60084/ijds.v2i2.236>
- Tavana, M., Shaabani, A., Vanani, I. R., & Gangadhari, R. K. (2022). A Review of Digital Transformation on Supply Chain Process Management Using Text Mining. In *Processes* (Vol. 10, Issue 5). MDPI. <https://doi.org/10.3390/pr10050842>
- Ying, S. (2024). Guests' Aesthetic experience with lifestyle hotels: An application of LDA topic modelling analysis. *Heliyon*, 10(16). <https://doi.org/10.1016/j.heliyon.2024.e35894>
- Zoya, Latif, S., Shafait, F., & Latif, R. (2021). Analyzing LDA and NMF Topic Models for Urdu Tweets via Automatic Labeling. *IEEE Access*, 9, 127531–127547. <https://doi.org/10.1109/ACCESS.2021.3112620>