

# ANALISIS SENTIMEN BERBASIS *LEXICON* DAN PEMODELAN SISTEM DENGAN LDA TERHADAP PERGURUAN TINGGI UPN “VETERAN” JAWA TIMUR PADA *GOOGLE REVIEW*

<sup>1</sup>Firzannabeel Aqila Rafid, <sup>2</sup>Hilya ‘Zada Mardhatilla Al Haadiy, <sup>3</sup>Azzahra Adelia Sabrina Salsabila, <sup>4</sup>Anggraini Puspita Sari\*

<sup>1,3</sup>Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur, <sup>2,3</sup>Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur, <sup>3,3</sup>Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur, <sup>4,3</sup>Fakultas Ilmu Komputer UPN “Veteran” Jawa Timur

Jl. Rungkut Madya No. 1, Gunung Anyar, Surabaya 60294, Jawa Timur

<sup>1</sup>22081010285@student.upnjatim.ac.id, <sup>2</sup>22081010172@student.upnjatim.ac.id,

<sup>3</sup>22081010272@student.upnjatim.ac.id, <sup>4</sup>anggraini.puspita.if@upnjatim.ac.id

\*) Penulis Korespondensi

## Abstrak

Ulasan menjadi faktor penting bagi para pemilik bisnis dalam meningkatkan kredibilitas layanannya. Tidak hanya para pebisnis, pihak pengelola tempat umum dari pemerintah juga perlu menganalisis ulasan untuk membangun reputasi publik yang lebih baik. Salah satunya adalah Perguruan Tinggi Universitas Pembangunan Nasional (UPN) “Veteran” Jawa Timur. Penelitian ini bertujuan untuk menganalisis ulasan publik mengenai UPN “Veteran” Jawa Timur dengan dua model, yakni analisis sentimen berbasis *Lexicon* dan pemodelan topik menggunakan *Latent Dirichlet Allocation (LDA)*. Analisis sentimen bertujuan untuk mengategorikan ulasan positif, negatif, dan netral dari beberapa ulasan publik di *Google Review*. Sedangkan, pemodelan topik bertujuan untuk mengetahui topik-topik utama dalam pembicaraan yang dibahas pada ulasan mengenai UPN “Veteran” Jawa Timur. Dari hasil penelitian terhadap 525 data ulasan, didapatkan sebanyak 352 ulasan dikategorikan positif, 136 ulasan netral, dan 37 ulasan negatif. Pengujian akurasi menunjukkan bahwa algoritma berbasis *Lexicon* memiliki tingkat akurasi sebesar 80%, yang menegaskan bahwa metode ini efektif dalam mengklasifikasikan data dengan akurat. Sementara itu, dalam pemodelan topik, berhasil diidentifikasi sepuluh topik utama pada setiap kategori sentimen, baik positif, negatif, maupun netral. Temuan ini memberikan wawasan berharga bagi pihak perguruan tinggi untuk lebih memahami permasalahan dan kebutuhan pengguna sehingga dapat mengambil langkah perbaikan yang lebih terarah.

**Kata Kunci:** Analisis Sentimen, LDA, *Lexicon*, Pemodelan Topik, Ulasan

## Abstract

Reviews are an important factor for business owners in increasing the credibility of their services. Not only business people, government public place managers also need to analyze reviews to build a better public reputation. One of them is the East Java "Veteran" National Development University (UPN) College. This research aims to analyze public reviews regarding UPN "Veteran" East Java with two models, namely *Lexicon*-based sentiment analysis and topic modeling using *Latent Dirichlet Allocation (LDA)*. Sentiment analysis aims to categorize positive, negative and neutral reviews from several public reviews on *Google Reviews*. Meanwhile, topic modeling aims to find out the main topics in the discussion discussed in the review of UPN "Veteran" East Java. From the results of research on 525 review data, 352 reviews were categorized as positive, 136 reviews were neutral and 37 reviews were negative. Accuracy testing shows that the *Lexicon*-based algorithm has an accuracy rate of 80%, which confirms that this method is effective in classifying data accurately. Meanwhile, in topic modeling, ten main topics were successfully identified in each sentiment category, both positive, negative and neutral. These

*findings provide valuable insight for universities to better understand user problems and needs so they can take more targeted improvement steps.*

**Keywords:** LDA, Lexicon, Review, Sentiment Analysis, Topic Modelling

## PENDAHULUAN

UPN “Veteran” Jawa Timur merupakan salah satu perguruan tinggi negeri di Indonesia yang memiliki reputasi baik dalam bidang pendidikan dan penelitian. Dalam era digital, reputasi universitas semakin dipengaruhi oleh media sosial dan platform online. Berdasarkan riset Katadata Insight Center (KIC) dan Aku Pintar bertajuk ‘Perilaku Siswa dalam Pemilihan Jurusan Pendidikan’, sebanyak 73% calon mahasiswa di Indonesia menggunakan media sosial untuk mencari sumber informasi terkait perguruan tinggi dan jurusan yang akan mereka pilih [1]. Ini menegaskan pentingnya ulasan online dalam membentuk citra universitas. Dalam konteks UPN “Veteran” Jawa Timur, memahami sentimen umum dalam ulasan *Google Review* menjadi penting untuk mendapatkan gambaran yang komprehensif tentang persepsi publik. Evaluasi sentimen dan topik utama dalam ulasan publik juga penting untuk mengidentifikasi area yang membutuhkan perhatian lebih dalam pengembangan fasilitas dan pelayanan kampus.

Analisis sentimen adalah suatu metode yang digunakan untuk mengolah data teks guna memperoleh pemahaman tentang apakah sentimen yang terkandung didalamnya bersifat positif, netral, atau negatif. Teknik ini digunakan untuk menganalisis penilaian atau

opini pribadi yang diberikan oleh pengguna internet melalui berbagai platform media sosial [2]. Pada umumnya, terdapat dua pendekatan untuk melakukan analisis sentimen, yakni *machine learning* dan *knowledge-based*. Penelitian ini fokus menggunakan pendekatan *knowledge-based* dengan metode berbasis *lexicon* [3]. Adapun untuk melengkapi sentimen analisis dilakukan pemodelan topik, yaitu salah satu teknik yang digunakan untuk *text mining* dalam analisis teks untuk mengeksplorasi data, menemukan pola tersembunyi, dan mengungkap hubungan antara informasi dan teks [4]. Pada penelitian ini menggunakan metode LDA untuk memeringkat topik dan mengklasifikasikan datanya ke sentimen positif, negatif atau netral. Beberapa penelitian terkait mengenai analisis sentimen dan pemodelan topik sudah pernah dibahas sebelumnya, seperti halnya penelitian oleh Rifiana Arief dan Karel Imanuel menganalisis sentimen dengan metode *lexicon based* pada *Twitter* [5]. Berdasarkan hasil penilaian *lexicon* terhadap 1000 kata *tweet* mengenai Desa Penari, ditemukan bahwa 33 *tweet* memiliki sentimen positif, 767 *tweet* bersentimen netral, dan 200 *tweet* bersentimen negatif. Persentase *tweet* dengan komentar positif adalah 3,3%, netral 76,7%, dan negatif 20%.

Studi lainnya yang mengkaji analisis sentimen berbasis *lexicon* dengan

menggunakan python library TextBlob memperoleh akurasi secara keseluruhan sebesar 81.9% [6]. Dari 4130 data *tweet* yang telah dikumpulkan, hasil yang diperoleh, yaitu sentimen positif sebanyak 1942, sentimen negatif sebanyak 1259 dan sentimen netral sebanyak 929. Adapun penelitian lain yang menganalisis pemodelan topik menggunakan metode Latent Dirichlet Allocation untuk mengolah data dengan pre-processing [7]. Hasilnya berupa diagram yang menggambarkan konten berita yang mencakup tiga topik utama yang sering dibahas di akun media sosial Twitter detik.com. Visualisasi pemodelan topik dapat dilakukan menggunakan pustaka PyLDAvis.

Berdasarkan kajian yang telah dilakukan, penelitian ini berfokus pada analisis sentimen berbasis *lexicon* serta pemodelan topik menggunakan metode (LDA). Penelitian ini bertujuan untuk mengklasifikasikan ulasan terhadap UPN "Veteran" Jawa Timur ke dalam kategori sentimen positif, negatif, dan netral. Selain itu, juga menetapkan dan mengurutkan

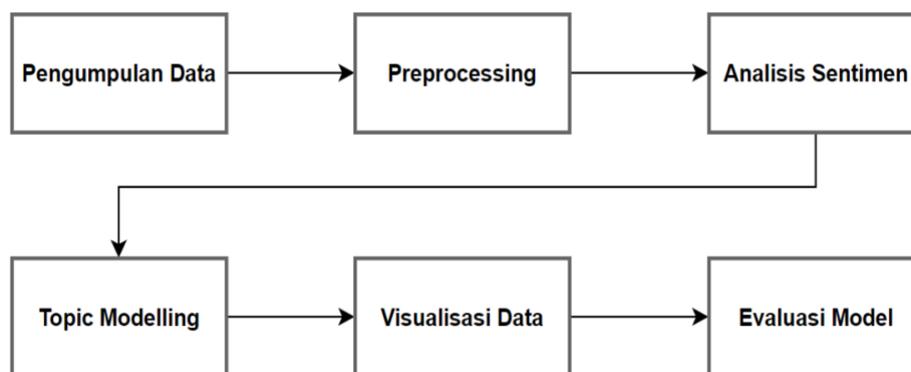
topik-topik utama yang dibahas dalam ulasan Google Review.

## METODE PENELITIAN

Sebuah penelitian memiliki pedoman berupa alur metode penelitian agar hasil yang didapatkan sesuai dengan tujuan penelitian. Penelitian ini menggunakan pendekatan *machine learning* berupa algoritma berbasis *Lexicon* untuk menganalisis sentimen ulasan terhadap UPN "Veteran" Jawa Timur dan algoritma LDA (*Latent Dirichlet Allocation*) untuk menentukan topik dominan dalam ulasan tersebut. Tahapan alur pada penelitian ini dapat dilihat pada Gambar 1.

### Pengumpulan Data

Sumber data yang digunakan dalam dataset pada penelitian ini berasal dari seluruh ulasan tertulis terhadap Perguruan tinggi UPN "Veteran" Jawa Timur pada *Google Review* dalam rentang waktu 11 tahun, yaitu mulai dari tahun 2013 sampai Mei 2024.



**Gambar 1. Alur Metode**

Data yang diperoleh sebanyak 525 ulasan dengan menggunakan teknik *web scraping* pada *Google Review*. *Web scraping* adalah metode otomatis untuk mengumpulkan informasi dari situs web tanpa perlu menyalinnya secara manual [8]. Teknik ini bertujuan untuk mengumpulkan data dari bagian tertentu dari sebuah situs web. Berbeda dengan *web crawling* yang mengunjungi seluruh situs yang berkaitan dengan situs utama, *web scraping* hanya mengekstraksi data tertentu dari situs yang dituju sesuai kebutuhan [9].

### **Preprocessing**

Setelah data terkumpul, dilihat pada Gambar 1, tahap selanjutnya adalah *preprocessing* data. Pada alur *preprocessing* akan dilakukan penyeleksian data. Apabila terdapat data yang tidak relevan, bernilai kosong, dan inkonsistensi akan dihilangkan. Tahapan pada *preprocessing* adalah sebagai berikut:

#### 1. Tokenisasi

Tokenisasi merupakan proses untuk memecah teks menjadi unit yang lebih kecil yaitu berupa kata atau frasa. Pada tahap ini dilakukan menggunakan fungsi `word_tokenize` dari library NLTK (Natural Language Toolkit).

#### 2. Penghapusan Stopwords

Tahap ini merupakan penghapusan kata yang umum dan tidak memberikan makna signifikan, contohnya adalah “dan”, “atau”, dan “saya”. Penghapusan stopwords

dilakukan dengan menggunakan library Sastrawi. Tahap ini efektif dilakukan untuk penyeleksian data karena jumlah kata yang akan diproses akan berkurang.

#### 3. Stemming

Stemming merupakan proses untuk mengubah kata menjadi bentuk dasarnya. Proses ini melibatkan penghilangan imbuhan (*affix*) yang terdiri dari awalan (*prefix*), sisipan (*infix*), akhiran (*suffix*) dan duplikasi.

### **Analisis Sentimen**

Penelitian ini menggunakan metode *Polarity Lexicon* atau *Lexicon Based* untuk melakukan analisis sentimen. *Lexicon based* merupakan kamus atau leksikon yang digunakan untuk memilih kata-kata pada data atau dokumen.

Dalam penerapan metode analisis sentimen, terdapat dua jenis kamus yang sering digunakan yaitu kamus sentimen positif yang mengandung kumpulan kata yang memiliki konotasi positif dan kamus sentimen negatif yang mengandung kumpulan kata yang memiliki konotasi negatif. Kedua kamus ini digabungkan menjadi *wordlist* yang digunakan dalam analisis sentimen berbasis leksikon (*lexicon based*). Dari kamus sentimen (*wordlist*) didapatkan sebanyak 132 kata positif dan 55 kata negatif yang menjadi panduan untuk mengkategorikan sentimen positif dan negatif dari sebuah ulasan berdasarkan kata-kata yang terkandung di dalamnya.

**Tabel 1. Contoh Kategorisasi Sentimen**

Contoh Data Ulasan	Bobot
"Kampus ini memiliki lingkungan yang luas"	1
"MINUS BINTANG 0 buat satpam yang di parkir motor."	-1
"28 Oktober 2023 tepat di Hari Sumpah Pemuda"	0

**Tabel 2. Contoh Data Ulasan yang Dikumpulkan**

Contoh Data Ulasan
"Salah satu Kampus Negeri terbesar di Jawa Timur"
"Salah satu perguruan tinggi negeri di Jawa Timur. Lingkungan cukup asri karena banyak pohon-pohon.."
"butuh peningkatkan kualitas provider dan wifi area upn apa lagi mahasiswa sangat butuh serta memperbarui fasilitas kampus nya"

Tabel 1 menunjukkan beberapa contoh data ulasan yang dikategorikan berdasarkan bobot dari kata yang disampaikan. Setiap ulasan seperti yang tertera pada Tabel 2 akan dikalkulasikan dengan diatas bobot 0 maka diberi label positif, jika dibawah 0 negatif, dan jika 0 netral, tergantung pada kata-kata dan frasa yang digunakan dalam ulasan tersebut. Penggunaan *wordlist* yang berisi kata-kata positif dan negatif membantu dalam menentukan kategori sentimen masing-masing ulasan.

### **Pemodelan Topik**

Pemodelan topik dilakukan untuk mengidentifikasi topik-topik dominan yang dibahas dalam ulasan.

Penelitian ini menggunakan metode LDA (*Latent Dirichlet Allocation*), yang mempelajari hubungan antara kata-kata, topik, dan dokumen dengan mengasumsikan dokumen dihasilkan oleh model probabilistik tertentu. LDA termasuk model tanpa pengawasan (*unsupervised*) yang

menggunakan pendekatan statistik untuk menemukan topik dalam kumpulan dokumen [10].

Metode ini digunakan untuk memahami topik-topik yang paling sering dibahas dalam ulasan dan mendapatkan gambaran yang lebih menyeluruh tentang ulasan terhadap UPN "Veteran" Jawa Timur.

### **Visualisasi Data**

Langkah berikutnya setelah semua tahap dan proses selesai adalah visualisasi. Dalam penelitian ini, visualisasi dilakukan menggunakan library Matplotlib.

Visualisasi ini menghasilkan diagram lingkaran yang menunjukkan persentase akurasi dari polaritas setiap sentimen yang dihasilkan. Penelitian ini menggunakan grafik yang menunjukkan jenis kata dan jumlah kemunculannya dalam setiap topik untuk memvisualisasikan hasil *topic modelling*. Grafik ini digunakan untuk menunjukkan distribusi kata dalam setiap topik yang diidentifikasi oleh model.

## Evaluasi Model

Teknik *confusion matrix* digunakan dalam penelitian ini sebagai evaluasi model. *Confusion matrix* merupakan tabel yang sering digunakan untuk mendeskripsikan performa model *machine learning* karena memberikan gambaran menyeluruh tentang prediksi dan kondisi aktual dari data yang diolahnya.

*Confusion matrix* terdiri dari empat istilah sebagai representasi hasil proses klasifikasi. Keempat istilah tersebut di antaranya :

1. TP (*True Positive*) adalah jumlah prediksi benar untuk kelas positif
2. FN (*False Negative*) adalah jumlah prediksi salah di mana kelas sebenarnya positif tetapi diprediksi negatif
3. TN (*True Negative*) adalah jumlah prediksi benar untuk kelas negatif
4. FP (*False Positive*) adalah jumlah prediksi salah di mana kelas sebenarnya negatif tetapi diprediksi positif.

Tabel 3 menunjukkan Confusion Matrix yang digunakan dalam penelitian ini untuk mengevaluasi hasil prediksi. Tolak ukur hasil perhitungan metode *confusion matrix* dalam penelitian ini yaitu *Accuration*, *Precision*, *Recall*, dan *F1-Score*.

1. *Accuration* mengukur seberapa besar proporsi prediksi yang benar dibandingkan dengan seluruh data. Persamaan akurasi ditunjukkan sebagai berikut :

$$Accuration = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \quad (1)$$

2. *Precision* menunjukkan rasio prediksi benar positif terhadap seluruh hasil yang diprediksi positif. Persamaan presisi ditunjukkan sebagai berikut :

$$Precision = \frac{TP}{TP+FP} \times 100 \quad (2)$$

3. *Recall* menunjukkan rasio prediksi benar positif terhadap seluruh data yang benar-benar positif. Persamaan recall ditunjukkan sebagai berikut :

$$Recall = \frac{TP}{TP+FN} \times 100 \quad (3)$$

4. *F1-Score* menghitung perbandingan antara rata-rata nilai presisi dan recall dari hasil pengujian. Persamaan *F1-Score* ditunjukkan sebagai berikut :

$$F1-Score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \quad (4)$$

Evaluasi model dilakukan dengan memisahkan dataset menjadi data latih dan data uji melalui proses split data. Split data menjadi dua bagian ini memungkinkan evaluasi model dengan objektif. Data uji yang tidak pernah digunakan saat latihan digunakan untuk mengukur kemampuan model dalam menangani data baru dengan lebih akurat.

**Tabel 3. Confusion Matrix**

Kelas Sebenarnya	Kelas Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

## HASIL DAN PEMBAHASAN

Hasil dari *web scraping* ulasan terhadap Perguruan Tinggi UPN “Veteran” Jawa Timur pada *Google Review* dari tahun 2013 hingga Mei 2024 mendapatkan data sebanyak 525 ulasan. Tahap selanjutnya adalah *preprocessing* di mana data akan diolah melalui proses tokenisasi, penghapusan *stopwords*, dan *stemming*. Hasil dari *preprocessing* dilakukan proses analisis sentimen yang hasil dari pemrosesan akan disimpan ke file *google\_labeled.csv* yang kemudian dilakukan *split data* untuk dilakukan evaluasi model setelahnya.

### Analisis Sentimen

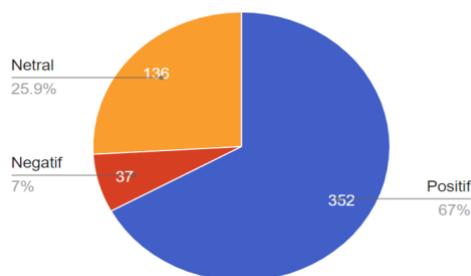
Analisis sentimen dari kamus sentimen yang disimpan di *wordlist*. Dihasilkan 352 ulasan yang tertera pada Gambar 2 memiliki sentimen positif, 136 ulasan bersentimen netral, dan 37 ulasan bersentimen negatif. Dengan total data ulasan sebanyak 525 data yang diproses, persentase yang diperoleh merupakan positif sebanyak 64% (352 data). Kemudian kedudukan kedua 25.9% (136 data) kategori netral, terakhir merupakan negatif dengan persentase 7% (37 data). Sentimen

publik terhadap kampus Universitas Pembangunan Nasional “Veteran” Jawa Timur mayoritas mendapatkan mayoritas positif dan berbanding jauh dengan persentase negatif.

### Pemodelan Topik

Setelah data melalui proses analisis sentimen, maka hasil pelabelan yang ada akan diproses untuk melakukan pemodelan topik yang kemudian menghasilkan tiga keluaran masing-masing yaitu topik positif, netral, dan negatif. Topik bahasan akan dikelompokkan dalam 6 topik keluaran yang menunjukkan kata-kata. Semakin tinggi persentase suatu topik itu muncul, menandakan bahwa topik tersebut banyak dibahas atau diperbincangkan dalam ulasan tersebut.

Pada sumbu vertikal (*y-axis*) grafik, ditampilkan jenis-jenis kata atau istilah yang paling sering muncul dalam setiap topik positif. Kata-kata ini merupakan kata kunci atau frasa yang relevan dengan topik positif dalam ulasan tersebut. Pada sumbu horizontal (*x-axis*), ditampilkan jumlah kemunculan atau frekuensi setiap kata dalam topik tersebut. Dalam hal ini menunjukkan seberapa sering kata tersebut muncul dalam topik suatu ulasan positif.



Gambar 2. Hasil Analisis Sentimen

### Pemodelan Topik Positif

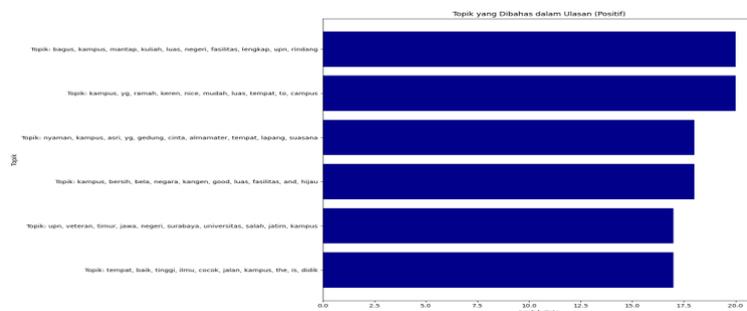
Hasil yang didapat pada ulasan positif sebagaimana tertera pada Gambar 3 bahwa perguruan tinggi UPN “Veteran” Jawa Timur merupakan 6 topik yang dapat disimpulkan bahwa:

1. Topik 1 = Sentimen merasa fasilitas perguruan tinggi yang lengkap dan baik dengan suasana yang rindang.
2. Topik 2 = Sentimen mengapresiasi masyarakat perguruan tinggi tersebut ramah, keren, dan baik.
3. Topik 3 = Sentimen merasakan kecintaannya terhadap almamater perguruan tinggi tersebut dengan lapang.
4. Topik 4 = Sentimen berpendapat bahwa perguruan tinggi tersebut bersih dan hijau yang menciptakan kerinduan.
5. Topik 5 = Sentimen mengapresiasi bahwa perguruan tinggi tersebut merupakan PT negeri di Jawa Timur.
6. Topik 6 = Sentimen berpendapat bahwa perguruan tinggi tersebut cocok untuk dijadikan tempat menimba ilmu.

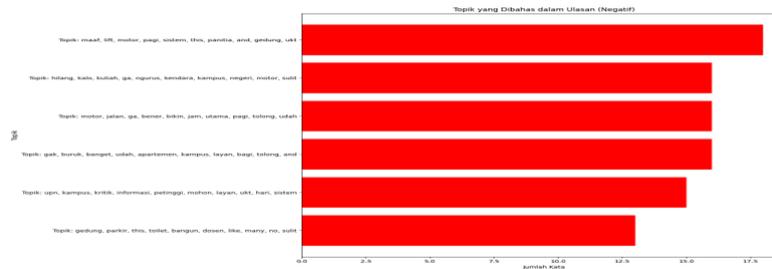
### Pemodelan Topik Negatif

Hasil yang didapat pada ulasan negatif sebagaimana tertera pada Gambar 4 bahwa perguruan tinggi UPN “Veteran” Jawa Timur merupakan 6 topik yang dapat disimpulkan bahwa:

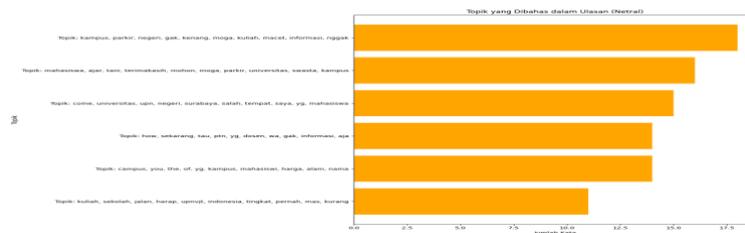
1. Topik 1: Sentimen mengecewakan fasilitas gedung seperti lift dan sistem yang disuguhkan oleh perguruan tinggi tersebut.
2. Topik 2: Sentimen merasa bahwa pelayanan saat kehilangan kurang terurus dan cenderung sulit.
3. Topik 3: Sentimen mengeluhkan keriuhan motor saat pagi hari karena sistem keluar masuk yang tidak efisien.
4. Topik 4: Sentimen merasa bahwa pelayanan perguruan tinggi yang buruk.
5. Topik 5: Sentimen mengkritik para petinggi perguruan tinggi pada pemanfaatan ukt.
6. Topik 6: Sentimen menyayangkan fasilitas parkir, toilet, dan pelayanan dosen yang dirasa kurang baik.



Gambar 3. Diagram Pemodelan Topik Positif



**Gambar 4. Diagram Pemodelan Topik Negatif**



**Gambar 5. Diagram Pemodelan Topik Netral**

**Tabel 4. Hasil Percobaan Evaluasi Model**

Split Data	Akurasi	Presisi	Recall	<i>F1-score</i>
15%	80%	74%	77%	67%
20%	76%	26%	33%	29%
10%	72%	42%	45%	44%

### Pemodelan Topik Netral

Hasil yang didapat pada ulasan netral yang tertera pada Gambar 5 bahwa perguruan tinggi UPN “Veteran” Jawa Timur merupakan 6 topik yang dapat disimpulkan bahwa sentimen tidak berbobot kepada suatu kebaikan ataupun keburukan, contoh masyarakat berpendapat bahwa perguruan tinggi tersebut merupakan kenangannya ataupun menuliskan pelaksanaan kegiatan kemerdekaan dilaksanakan pada PT tersebut.

### Evaluasi Model

Dataset melalui tahap analisis sentimen yang disimpan ke file *google\_labeled.csv* maka akan dilanjutkan dengan melakukan proses evaluasi model dengan menggunakan *split data* dalam beberapa percobaan.

Evaluasi model diperlukan untuk mengetahui hasil performa dari metode yang digunakan.

Ketiga percobaan *split data* seperti yang tercantum pada Tabel 4, percobaan telah memberikan hasil secara keseluruhan terbesar didapat pada 15% dengan tingkat akurasi 80%, maka metode *Lexicon Based* sudah dapat diimplementasikan sebagai metode yang akurat untuk mengklasifikasikan data dengan benar.

Kesalahan 20% yang ada merupakan kemungkinan penggunaan kata pada kalimat ulasan yang tidak sesuai, kesalahan ketik, ataupun adanya kata selain abjad dan angka seperti *emoticon* yang membuat kata dalam kalimat tersebut tidak terbaca oleh sistem.

## KESIMPULAN DAN SARAN

Berdasarkan hasil pengujian analisis, ulasan positif mendominasi pada *Google Review* terhadap kampus UPN “Veteran” Jawa Timur. Dari 525 ulasan, diperoleh 352 ulasan dengan sentimen positif, 136 ulasan dengan sentimen netral, 37 ulasan dengan sentimen negatif. Data ini menunjukkan tingkat kepuasan yang tinggi dari pengunjung terhadap lingkungan sekitar perguruan tinggi yang masih terasa alami dan sejujunya. Analisis pemodelan topik menunjukkan bahwa terdapat enam topik utama dalam setiap kategori sentimen pada ulasan *Google Review* terhadap kampus UPN “Veteran” Jawa Timur. Ulasan positif menyoroti lingkungan perguruan tinggi yang hijau, asri, dan nyaman, serta mengapresiasi UPN “Veteran” Jawa Timur menjadi salah satu Perguruan Tinggi Negeri (PTN) di Jawa Timur yang cocok menjadi tempat untuk menimba ilmu. Ulasan netral tidak dapat disimpulkan karena tidak memiliki kecenderungan kepada suatu kebaikan ataupun keburukan. Adapun ulasan negatif menyoroti keluhan terkait fasilitas parkir dan toilet, volume motor, masalah pembangunan, serta pelayanan administrasi dan sistem yang dianggap kurang efisien oleh masyarakat.

Dibandingkan dengan penelitian-penelitian sebelumnya yang umumnya hanya fokus pada analisis sentimen tanpa mengidentifikasi topik utama dalam setiap kategori sentimen, penelitian ini memberikan pendekatan yang lebih komprehensif dengan

mengidentifikasi topik utama yang muncul dalam ulasan *Google Review* terhadap kampus UPN “Veteran” Jawa Timur. Selain itu, penelitian ini juga menekankan penggunaan metode berbasis *lexicon*, yang terbukti memberikan akurasi sebesar 80%, menunjukkan bahwa metode ini dapat diandalkan dalam mengklasifikasikan data sentimen secara tepat. Sementara itu, penelitian terdahulu lebih sering menggunakan pendekatan berbasis *machine learning* yang memerlukan data pelatihan dalam jumlah besar dan sering kali memerlukan penyesuaian yang lebih kompleks. Penelitian ini juga menawarkan wawasan yang lebih mendalam terkait kepuasan pengunjung terhadap lingkungan kampus, yang belum banyak dibahas secara detail dalam studi-studi sebelumnya.

## DAFTAR PUSTAKA

- [1] C. M. Annur, “Mayoritas Calon Mahasiswa Cari Informasi Universitas dan Jurusan Lewat Sosial Media,” *databoks.katadata.co.id*. [Online]. Available: <https://databoks.katadata.co.id/datapublish/2021/06/21/media-sosial-sumber-informasi-perguruan-tinggi-jurusan-mayoritas-siswa>
- [2] F. V. Sari and A. Wibowo, “Analisis Sentimen Pelanggan Toko Online Jd.Id Menggunakan Metode Naïve Bayes Classifier Berbasis Konversi Ikon

- Emosi,” *J. SIMETRIS*, vol. 10, no. 2, pp. 681–686, 2019.
- [3] S. M. Isa, “Sentiment Analysis Approaches and Methods,” *binus.ac.id*. Accessed: Jun. 12, 2024. [Online]. Available:<https://mti.binus.ac.id/2017/10/04/1900>
- [4] H. Jelodar et al., “Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey,” *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, 2019, doi: 10.1007/s11042-018-6894-4.
- [5] R. Arief and K. Imanuel, “Analisis Sentimen Topik Viral Desa Penari Pada Media Sosial Twitter Dengan Metode Lexicon Based,” *J. Ilm. Matrik*, vol. 21, no. 3, pp. 242–250, 2019, doi: 10.33557/jurnalatrik.v21i3.727.
- [6] A. Syakur, “Implementasi Metode Lexicon Base Untuk Analisis Sentimen Kebijakan Pemerintah Dalam Pencegahan Penyebaran Virus Corona Covid-19 Pada Twitter,” *J. Ilm. Inform. Komput.*, vol. 26, no. 3, pp. 247–260, 2021, doi:10.35760/ik.2021.v26i3.4720.
- [7] N. Cahyono and Angga Reni Dwi Astuti, “Analisis Topic Modelling Persepsi Pengguna Internet Menggunakan Metode Latent Dirichlet Allocation,” *Indones. J. Comput. Sci.*, vol. 12, no. 1, pp. 326–334, 2023, doi: 10.33022/ijcs.v12i1.3155.
- [8] D. D. A. Yani, H. S. Pratiwi, and H. Muhandi, “Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace,” *J. Sist. dan Teknol. Inf.*, vol. 7, no. 4, p. 257, 2019, doi: 10.26418/justin.v7i4.30930.
- [9] A. Setiawan, L. W. Santoso, and R. Adipranata, “Klasifikasi Artikel Berita Bahasa Indonesia Dengan Naive Bayes Classifier,” *J. Infra*, vol. 8, no. 1, pp. 146–151, 2020.
- [10] Wahyudin, “Aplikasi Topic Modeling Pada Pemberitaan,” *Semin. Nas. Off. Stat. Pengemb. Off. Stat. dalam mendukung Implementasi SDG’s*, pp. 309–318, 2020.
- [11] A. P. Sari, A. N. Sihananto, D. A. Prasetya, M. M. A. Haromany, “Pengelompokan Tingkat Penyebaran Covid 19 Pada Kabupaten Jombang Dengan Menggunakan Algoritma K-Means,” *J. Teknol. Inf. dan Komunikasi*, vol. XVII, pp. 1–6, 2022, [Online]. Available:<http://www.ejournal.upnjatim.ac.id/index.php/scan/article/view/3514%0Ahttp://www.ejournal.upnjatim.ac.id/index.php/scan/article/download/3514/2099>
- [12] T. D. Dikiyanti, A. M. Rukmi, and M. I. Irawan, “Sentiment analysis and topic modeling of BPJS Kesehatan based on twitter crawling data using Indonesian Sentiment Lexicon and Latent Dirichlet Allocation algorithm,” *J. Phys. Conf. Ser.*, vol. 1821, no. 1, 2021, doi: 10.1088/1742-6596/1821/1/012054.

- [13] A. R. Dzibrillah and D. O. Dyvia Oliviani, "Analisis Sentimen Dan Topik Perbincangan Netizen Indonesia Mengenai Kebijakan Vaksin Covid-19," *Pseudocode*, vol. 9, no. 2, pp. 106–113, 2022, doi:10.33369/pseudocode.9.2.106-113.
- [14] A. A. R. M. N. Asti, Ismamiaty, "Lexicon Based Sentiment Analysis pada Trending Topics di Nusa Tenggara Barat," *J. Inform. dan Teknol. Komput.*, vol. 03, pp. 93–98, 2022.
- [15] K. Dheanis, A. Salsabila, N. Trianasari, R. Artikel, K. Kunci, and P. Konsumen, "Jurnal Teknologi dan Manajemen Informatika Analisis Persepsi Produk Kosmetik Menggunakan Metode Sentiment Analysis Dan Topic Modeling (Studi Kasus: Laneige Water Sleeping Mask) Info Artikel ABSTRAK," *J. Teknol. dan Manaj. Inform.*, vol. 7, no. 1, pp. 1–9, 2021, [Online]. Available: <http://http://jurnal.unmer.ac.id/index.php/jtmi>
- [16] F. Amaliah and I. K. Dwi Nuryana, "Perbandingan Akurasi Metode Lexicon Based Dan Naive Bayes Classifier Pada Analisis Sentimen Pendapat Masyarakat Terhadap Aplikasi Investasi Pada Media Twitter," *J. Informatics Comput. Sci.*, vol. 3, no. 03, pp. 384–393, 2022, doi: 10.26740/jinacs.v3n03.p384-393.
- [17] Y. Matira, Junaidi, and I. Setiawan, "Pemodelan Topik pada Judul Berita Online Detikcom Menggunakan Latent Dirichlet Allocation," *Estimasi J. Stat. Its Appl.*, vol. 4, no. 1, pp. 2721–379, 2023, doi: 10.20956/ejsa.vi.24843.
- [18] D. Ridhwanulah and D. H. Fudholi, "Pemodelan Topik pada Cuitan tentang Penyakit Tropis di Indonesia dengan Metode Latent Dirichlet Allocation," *J. Ilm. SINUS*, vol. 20, no. 1, p. 11, 2022, doi: 10.30646/sinus.v20i1.589.
- [19] Aditya Quantano Surbakti, Regiolina Hayami, and Januar Al Amien, "Analisa Tanggapan Terhadap Psbb Di Indonesia Dengan Algoritma Decision Tree Pada Twitter," *J. CoSciTech (Computer Sci. Inf. Technol.)*, vol. 2, no. 2, pp. 91–97, 2021, doi:10.37859/coscitech.v2i2.285.
- [20] Arif Widiasan Subagio, Anggraini Puspita Sari, and Andreas Nugroho Sihananto, "Klasifikasi Lexicon-Based Sentiment Analysis Tragedi Kanjuruhan pada Twitter Menggunakan Algoritma Convolutional Neural Network," *J. Ilm. Sist. Inf. dan Ilmu Komput.*, vol. 4, no. 1, pp. 166–177, 2024, doi: 10.55606/juisik.v4i1.759.
- [21] M. Y. Febrianta, S. Widiyanesti, and S. R. Ramadhan, "Analisis Ulasan Indie Video Game Lokal pada Steam Menggunakan Analisis Sentimen dan Pemodelan Topik Berbasis Latent Dirichlet Allocation," *J. Animat. Games Stud.*, vol. 7, no. 2, pp. 117–144, 2021, doi: 10.24821/jags.v7i2.5162.