

**Analisis Sentimen Judul Berita Pada Portal Berita *Online* di
Dinas Kominfotik Provinsi Gorontalo**

Oleh

RISKI YUSUF ZAKARIA

T3121151

USULAN PENELITIAN



**PROGRAM SARJANA
FAKULTAS ILMU KOMPUTER
UNIVERSITAS ICHSAN GORONTALO
GORONTALO 2024**

PERSETUJUAN USULAN PENELITIAN

**Analisis Sentimen Judul Berita Pada Portal Berita *Online* di
Dinas Kominfotik Provinsi Gorontalo**

Oleh

RISKI YUSUF ZAKARIA

T3121151

USULAN PENELITIAN

Telah disetujui dan siap untuk diseminarkan
Gorontalo, 2024

Pembimbing I

Pembimbing II

Irvan Abraham Salihi, M.Kom

Ir. Misrawaty Aprilyana Puspa, M.Kom

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT yang telah memberikan rahmat dan karunianya, shalawat serta salam semoga tercurah kepada Rasulullah SAW, sehingga penulis dapat menyelesaikan Proposal yang berjudul “**Analisis Sentimen Judul Berita Pada Portal Berita *Online* di Dinas Kominfotik Provinsi Gorontalo.**”.

Adapun tujuan dari penyusunan Proposal ini adalah untuk memenuhi salah satu syarat dalam menyelesaikan studi jenjang Strata satu (S1) di Program Studi Teknik Informatika, Universitas Ichsan Gorontalo.

Dengan keterbatasan ilmu dan pengetahuan yang dimiliki oleh penulis, maka penulis membutuhkan peran serta dari pihak lain dalam proses penyelesaian proposal ini.

Penulis juga mengucapkan terima kasih kepada pihak-pihak yang telah membantu terselesainya proposal penelitian ini :

1. Ibu Dr. Dra. Juriko Abdussamad, M.Si, selaku Ketua Yayasan Pengembangan Ilmu Pengetahuan dan Teknologi (YPIPT) Ichsan Gorontalo;
2. Bapak Dr. Abdul Gaffar La Tjokke, M.Si, selaku Rektor Universitas Ichsan Gorontalo;
3. Bapak Irvan A. Salihi, S.Kom, M.Kom, selaku Dekan Fakultas Ilmu Komputer Universitas Ichsan Gorontalo Sekaligus Dosen Pembimbing Utama;
4. Bapak Sudirman Melangi, M.Kom, selaku Wakil Dekan I Bidang Akademik Fakultas Ilmu Komputer;
5. Ibu Irma Surya Kumala, M.Kom, selaku Wakil Dekan II bidang Administrasi Umum dan Keuangan Fakultas Ilmu Komputer;
6. Bapak Sudirman S. Panna, M.Kom, selaku Ketua Jurusan Teknik Informatika Fakultas ilmu Komputer Universitas Ichsan Gorontalo;

7. Ibu Ir. Misrawaty Aprilyana Puspa, M.Kom, selaku Dosen Pembimbing Kedua;
8. Bapak dan Ibu Dosen Universitas Ichsan Gorontalo yang sudah mendidik dan mengajarkan berbagai ilmu disiplin kepada penulis;
9. Orang Tua yang selalu memberi dukungan kepada saya sehingga dapat fokus mengerjakan tugas akhir ini.

Didalam penulisan Proposal ini, penulis telah berusaha semaksimal mungkin, walaupun demikian penulis menyadari bahwa Proposal ini jauh dari sempurna. Untuk itu penulis akan selalu menerima segala masukan yang ditujukan untuk menyempurnakan Proposal ini. Akhir kata, penulis berharap semoga Proposal ini dapat bermanfaat bagi penulis pada khususnya dan pembaca pada umumnya.

Gorontalo, 24 Oktober 2024

(Riski Yusuf Zakaria)

DAFTAR ISI

KATA PENGANTAR	iii
DAFTAR ISI.....	v
DAFTAR GAMBAR	vii
DAFTAR TABEL.....	viii
DAFTAR LAMPIRAN.....	ix
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Identifikasi Masalah	2
1.3. Rumusan Masalah	2
1.4. Tujuan Penelitian.....	2
1.5. Manfaat Penelitian.....	3
BAB II LANDASAN TEORI	4
2.1. Tinjauan Studi	4
2.2. Tinjauan Pustaka	5
2.3. Kerangka Pikir.....	6
BAB III METODE PENELITIAN.....	7
3.1. Jenis, Objek, Waktu dan Lokasi Penelitian	7
3.1.1. Jenis penelitian.....	7
3.1.2. Objek, Waktu dan Lokasi	7
3.2. Teknik Pengumpulan Data	7
3.3. Pengembangan Sistem.....	Error! Bookmark not defined.
3.3.1. Sistem yang Diusulkan	Error! Bookmark not defined.
3.3.2. Analisis Sistem	Error! Bookmark not defined.

3.3.3. Desain Sistem	Error! Bookmark not defined.
3.3.4. Pengujian Sistem.....	Error! Bookmark not defined.
DAFTAR PUSTAKA	16
LAMPIRAN.....	17

DAFTAR GAMBAR

Gambar 3.1 Sistem Yang Diusulkan.....	Error! Bookmark not defined.
Gambar 3.2 Use Case Diagram.....	Error! Bookmark not defined.
Gambar 3.3 Class Diagram	Error! Bookmark not defined.
Gambar 3.4 Halaman Home.....	Error! Bookmark not defined.
Gambar 3.5 Halaman Toko	Error! Bookmark not defined.
Gambar 3.6 Halaman Wilayah.....	Error! Bookmark not defined.

DAFTAR TABEL

Tabel 2. 1 Tinjauan Studi	4
Tabel 2. 2 Kerangka Pikir	6

DAFTAR LAMPIRAN

Lampiran 1 Timeline Penelitian.....	17
-------------------------------------	----

BAB I PENDAHULUAN

1.1. Latar Belakang

Portal berita *online* merupakan situs atau halaman web mengenai berbagai jenis berita : politik, ekonomi, sosial, budaya bahkan hiburan yang bersifat *hard news* maupun *soft news*. [1] Pada portal beritu *online* tentu menampilkan berbagai macam postingan yang terususun mulai dari judul, isi, hingga tanggal dan penulis berita. Judul berita sendiri merupakan representasi dari isi berita yang dituliskan. Penulisan judul dalam sebuah berita harus menggunakan bahasa yang mudah dipahami oleh pembaca. [2] Sehingga, pada penelitian kali ini judul berita dipilih menjadi objek yang akan dianalisa.

Analisis sentimen judul berita yang akan dilakukan pada penelitian kali ini, sebelumnya juga sudah pernah diteliti dengan metode atau teknik analisis yang berbeda. Seperti pada penelitian dengan judul Analisis Sentimen Berita Artis Dengan Menggunakan Algoritma Support Vector Machine Dan Particle Swarm Optimization, yang disusun oleh Norma Yunita [3]. Dimana pada algoritma Support Vector Machine dapat digunakan untuk klasifikasi dan regresi. Sedangkan algoritma Particle Swarm Optimization digunakan untuk bisa meningkatkan akurasi pada Support Vector Machine. Sedangkan pada penelitian lainnya, yaitu pada penelitian yang dilakukan oleh Ary Suryadi beserta teman-teman dengan judul Implementasi Web Scraping dan Sentiment Analysis Terhadap Berita Menggunakan Machine Learning. [4] Pada penelitian kali ini, menggunakan metode Aspect-Based Sentiment Analysis yang dimana fokusnya adalah untuk mengidentifikasi aspek-aspek tertentu dalam sebuah teks dan menentukan sentimen terhadap aspek-aspek tersebut.

Dan pada penelitian kali ini, peneliti akan melakukan analisis sentimen dengan menggunakan metode yang sedikit berbeda dari penelitian sebelumnya, yaitu metode Lexicon dan Algoritma umum dari Support Vector Machine. Dimana

pada teknik lexicon ini menggunakan kamus kata yang berisi daftar kata dengan label sentimen tertentu seperti positif atau negatif. Sedangkan Support Vector Machine digunakan untuk klasifikasi dan regresi atau urutan berbalik kebelakang. Dengan menggabungkan kedua metode tersebut, diharapkan dapat menghasilkan tingkat akurasi dari analisis sentimen yang dilakukan.

Oleh karena itu peneliti mengajukan penelitian ini untuk mendapatkan hasil analisis sentimen berita dengan tingkat akurasi yang lebih baik. Dan kemudian hasil dari analisis sentimen tersebut diharapkan dapat menggambarkan opini publik terhadap sebuah isu yang diangkat pada sebuah judul berita sehingga diharapkan dapat membantu untuk pengambilan sebuah keputusan oleh Pimpinan Dinas Kominfo dan Organisasi Perangka Daerah Lainnya.

1.2. Identifikasi Masalah

Berdasarkan Uraian Tersebut dapat disimpulkan beberapa identifikasi masalah sebagai berikut:

1. Volume Berita yang Besar.
2. Memakan waktu dan tenaga untuk melakukannya secara manual.
3. Kemungkinan adanya judul ambigu, sarkastik, atau tidak langsung.

1.3. Rumusan Masalah

Berdasarkan identifikasi masalah diatas, maka penulis dapat merumuskan masalah sebagai berikut:

1. Bagaimana cara melakukan analisis sentimen?
2. Bagaimana meningkatkan akurasi dari analisis sentimen?
3. Bagaimana mengimplementasikan metode lexicon dan support vector machine pada analisis sentimen?

1.4. Tujuan Penelitian

1. Melakukan analisis sentimen judul berita dari portal berita *online*.

2. Menggunakan metode lexicon dan support vector machine pada analisis sentimen untuk meningkatkan akurasi dari analisis sentimen judul berita.

1.5. Manfaat Penelitian

Penelitian ini dibuat dengan harapan dapat memberikan manfaat secara teoritis dan praktis kepada pihak terkait, sebagai pertimbangan, masukan dan pedoman serta evaluasi meningkatkan pemahaman tentang pola sentimen dalam berita *online*.

1. Secara teoritis, penelitian ini diharapkan dapat memperkaya literatur terkait analisis sentimen judul berita *online*.
2. Adapun manfaat secara praktis dari penelitian ini, Diharapkan dapat membantu menganalisis tingkat sentimen judul berita *online*.

BAB II LANDASAN TEORI

2.1. Tinjauan Studi

Tabel 2. 1 Tinjauan Studi

No	Peneliti	Judul	Hasil
1	Ary Suryadi, Wahid Andika Syb'an, Nazzala Alfa'inna, Eni Henri Hermaliani	Implementasi Web Scraping dan Sentiment Analysis Terhadap Berita Menggunakan Machine Learning[4]	Penelitian ini membahas teknik pengumpulan data menggunakan webscraping dan menganalisis judul berita dengan metode Aspect-Based Sentiment Analysis.
2	Amalia Rachmadana Ismail dan Raden Bagus Fajriya Hakim	Implementasi Lexicon Based Untuk Analisis Sentimen Dalam Mengetahui Trend Wisata Pantai Di DI Yogyakarta Berdasarkan Data Twitter	Dalam makalah ini penulis menyajikan metode analisis sentimen dengan metode lexicon based mengenai opini masyarakat pada media sosial twitter tentang wisata pantai yang sedang trend di DI Yogyakarta.
3	Gian Kristian Haroni, Ferry V.I.A Koagouw dan Meity D. Himpong	Analisis Sentimen Review Aplikasi Media Berita Online Pada Google Play menggunakan Metode Algoritma Support	Penelitian ini menggunakan proses pembagian terstruktur atau klasifikasi mengenai machine learning dengan metode algoritma support vector machine dan naive bayes. Dataset yang diujikan

		Vector Machine (SVM) dan Naïve Bayes	diambil dari opini pengguna Aplikasi Media Berita <i>Online</i> pada kolom komentar digoogle play.
--	--	--------------------------------------	--

2.2. Tinjauan Pustaka

2.2.1. Portal Berita Online

Portal berita online merupakan situs atau halaman web mengenai berbagai jenis berita : politik, ekonomi, sosial, budaya bahkan hiburan yang bersifat hard news maupun soft news. Di Dinas Komunikasi, Informasi dan Statistika Provinsi Gorontalo sendiri bekerja sama dengan beberapa portal berita online yang ada di Gorontalo dengan tujuan untuk memberikan informasi kepada masyarakat yang dapat dipercaya. Adapun beberapa portal online yang bekerjasama dengan Dinas Komunikasi, Informasi, dan Statistika adalah sebagai berikut:

2.2.2. Metode Lexicon Based

Metode Lexicon adalah salah satu pendekatan dalam analisis sentimen yang menggunakan daftar kata atau frasa yang telah dikelompokkan berdasarkan sentimen tertentu (positif atau negatif) untuk menilai polaritas teks. Metode lexicon-based bekerja dengan cara membuat kamus kata opini (lexicon) terlebih dahulu. Kata-kata yang terdapat pada kamus tersebut digunakan untuk mengidentifikasi apakah suatu kalimat mengandung opini atau tidak.[7] Secara sederhana, Metode ini menggunakan aturan pencocokan kata (rule-based) yang relatif mudah untuk diimplementasikan dibandingkan dengan model pembelajaran mesin atau deep learning. Dimana setiap kata atau frasa dalam teks memiliki nilai polaritas tertentu yang jelas, sehingga hasil analisis dapat dengan mudah dipahami.

2.2.3. Support Vector Machine

Support Vector Machine (SVM) adalah metode pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. SVM bekerja dengan mencari hyperplane

terbaik yang dapat memisahkan data ke dalam kategori yang berbeda. Hyperplane ini adalah garis (dalam data dua dimensi) atau bidang (dalam data tiga dimensi) yang memaksimalkan margin antara dua kelas data.

2.3. Kerangka Pikir

Tabel 2. 2 Kerangka Pikir

Masalah		
<ol style="list-style-type: none"> 1. Bagaimana cara menerapkan teknik web <i>scraping</i> untuk mengumpulkan data judul berita dari situs berita <i>online</i>? 2. Bagaimana analisis bigram dapat diterapkan untuk mengidentifikasi pola dalam judul berita yang diambil? 3. Bagaimana visualisasi data dapat membantu dalam mempresentasikan hasil analisis secara informatif? 		
System Development		
1	Analisis Kebutuhan	<ol style="list-style-type: none"> 1. <i>Functional Modeling (Use Case)</i> 2. <i>Structural Modeling (Class Diagram)</i>
2	Pengujian Sistem	<ol style="list-style-type: none"> 1. <i>Unit Testing</i> 2. <i>Integration Testing</i> 3. <i>Performance Testing</i>
Hasil		
<p>Tujuan penelitian yaitu merancang arsitektur berdasarkan kebutuhan dari sistem yang akan dibuat atau dikembangkan sehingga memiliki visualisasi data yang diinginkan.</p>		

BAB III METODE PENELITIAN

3.1. Jenis, Objek, Waktu dan Lokasi Penelitian

3.1.1. Jenis penelitian

Jenis Penelitian yang dipilih pada penelitian kali ini adalah penelitian kualitatif. penelitian kualitatif dapat menjadi metode yang relevan. Penelitian kualitatif bertujuan untuk memahami dan menjelaskan fenomena yang kompleks melalui analisis mendalam tentang persepsi, pandangan, pengalaman, dan konteks sosial yang terlibat.

3.1.2. Objek, Waktu dan Lokasi

Objek pada penelitian kali ini adalah judul berita yang akan divisualisasikan berdasarkan kebutuhan. Waktu penelitian dilakukan sejak tanggal 25 Oktober 2024 dan Berlokasi di Dinas Komunikasi, Informasi, dan Statistik Provinsi Gorontalo.

3.2. Teknik Pengumpulan Data

1. Observasi

Observasi adalah teknik pengumpulan data yang dilakukan dengan cara mengamati secara langsung proses atau objek yang menjadi fokus penelitian. Dalam konteks penelitian ini, observasi dapat dilakukan untuk memahami pola penulisan berita. Proses observasi akan melibatkan Pengamatan terhadap judul berita.

2. Wawancara

Wawancara adalah teknik pengumpulan data kualitatif yang dilakukan dengan berinteraksi langsung dengan narasumber untuk mendapatkan informasi lebih mendalam. Dalam penelitian ini, wawancara akan dilakukan untuk mendapatkan pemahaman lebih lanjut tentang Kebutuhan dan harapan Dinas

Komunikasi, Informatika, dan Statistik Provinsi Gorontalo dan sumber data yang dibutuhkan berupa media-media yang menjadi mitra oleh Dinas Komunikasi, Informatika, dan Statistik Provinsi Gorontalo.

3. Studi Literatur

Teknik studi literatur digunakan untuk mengumpulkan informasi sekunder yang relevan dengan topik penelitian. Penelitian ini akan mengacu pada literatur sebelumnya terkait Analisis sentimen yang menggunakan berbagai macam metode analisis terutama metode yang digunakan pada penelitian ini, yaitu Metode Lexicon dan Support Vector Machine.

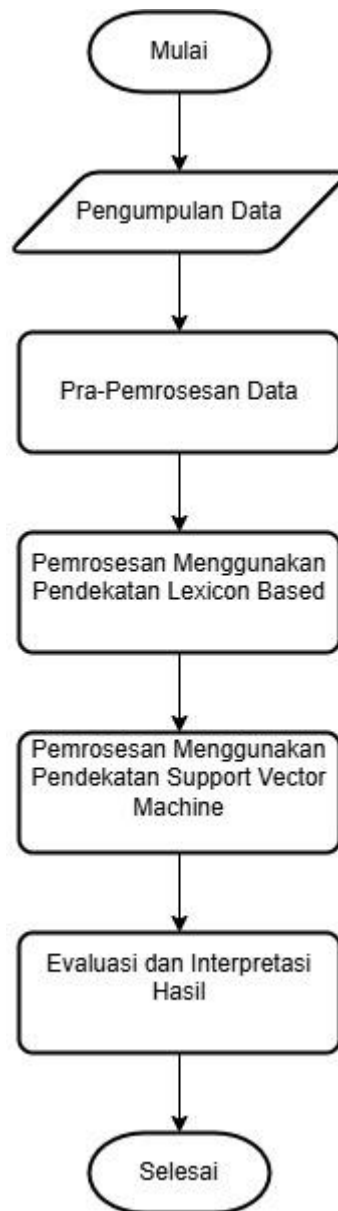
4. Web Scraping

Teknik web *scraping* digunakan sebagai salah satu metode pengumpulan data utama dalam penelitian ini. Dengan memanfaatkan skrip otomatis yang dirancang untuk mengekstraksi data dari halaman web, web *scraping* memungkinkan pengumpulan data judul berita secara sistematis dari berbagai situs berita *online* yang bekerja sama dengan Dinas Kominfotik.

3.3. Pemodelan

3.3.1 Model yang diusulkan

Untuk model yang akan diusulkan, akan digambarkan melalui flowchart berikut:



Pada gambar diatas, proses dimulai dari pengumpulan data yang kemudian masuk pada proses pra-pemrosesan. Pada proses pra-pemrosesan ini terdapat empat tahapan. Tahapan pertama adalah metode *Case folding* yang berarti menyamaratakan huruf besar dan kecil sehingga pencocokan kata dalam lexicon lebih konsisten, proses ini berguna untuk mengurangi kesalahan dalam pencocokan kata akibat perbedaan kapitalisasi. Kemudian Tokenization yang berarti memecah teks menjadi unit-unit kata sehingga setiap kata dapat dianalisis secara terpisah, proses ini sendiri berguna untuk mempermudah pencocokan setiap kata dengan daftar kata sentimen dalam leksikon (positif atau negatif). Tahap selanjutnya adalah

Stopword Removal yang bermaksud untuk membersihkan kata-kata umum yang tidak membawa sentimen, seperti "dan," "di," "yang," dll, tahap ini sendiri berguna agar Fokus hanya pada kata-kata yang relevan untuk analisis sentimen dan Mengurangi gangguan dari kata-kata netral yang tidak memiliki nilai sentimen. Dan tahap pra-pemrosesan yang terakhir adalah Stemming yang bermaksud untuk Mengubah kata-kata ke bentuk dasarnya agar lebih mudah dicocokkan dengan leksikon, tahap ini berguna untuk menyamakan berbagai bentuk kata sehingga pencocokan dengan leksikon lebih akurat dan Memastikan semua varian kata dikenali sebagai satu entitas yang sama dalam analisis sentimen.

Setelah melalui berbagai tahap pra-pemrosesan, kita akan masuk pada bagian pemrosesan menggunakan pendekatan Lexicon Based. Pada proses ini akan Menghitung skor sentimen setiap judul berita berdasarkan kosakata sentimen dan Memberikan label awal: Positif atau Negatif. Setelah pemrosesan menggunakan pendekatan Lexicon Based, Proses selanjutnya adalah *Machine learning* dengan *Support Vector Machine*. Pemrosesan ini menggunakan dataset terlabel dari hasil pemrosesan Lexicon Based untuk melatih model SVM. Kemudian pada pemrosesan ini juga Fitur diekstraksi menggunakan metode *word embedding*. Metode ini adalah metode representasi kata dalam bentuk vektor dengan dimensi tetap yang menangkap makna semantik dan hubungan antar kata. Kata-kata dengan makna atau konteks serupa akan memiliki vektor yang dekat secara geometris. Spesifik metode yang digunakan pada metode *word embedding* pada penelitian kali ini adalah *Pre-trained Models*. Model dari metode ini adalah Model seperti BERT atau *word embedding* yang sudah dilatih pada dataset besar.

Dan tahap terakhir adalah Evaluasi dan Interpretasi Hasil. Evaluasi dilakukan untuk mengukur kualitas model analisis sentimen dalam melakukan klasifikasi (positif atau negatif). Ini dilakukan dengan membandingkan hasil prediksi model dengan label data sebenarnya pada dataset uji (*test set*). Evaluasi ini sendiri melibatkan beberapa metrik utama:

a. *Accuracy* (Akurasi)

Metrik akurasi berfungsi untuk mengukur persentase prediksi yang benar dari keseluruhan prediksi. Metrik ini dapat dilakukan menggunakan rumus berikut:

$$\text{Accuracy} = \frac{\text{Jumlah Prediksi Benar}}{\text{Total Prediksi}}$$

b. *Precision* (Presisi)

Metrik presisi ini berfungsi untuk mengukur akurasi prediksi positif. Cara mengukur menggunakan metrik ini adalah dengan menggunakan rumus berikut:

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

Misalnya, presisi tinggi berarti model jarang salah mengklasifikasikan sentimen positif.

c. *Recall* (Sensitivitas)

Metrik ini berfungsi untuk mengukur sejauh mana model dapat menemukan semua data yang relevan. Cara menggunakan metrik ini dapat memperhatikan rumus berikut:

$$\text{Recall} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

Recall tinggi berarti model berhasil menangkap semua sentimen positif yang ada.

d. F1-Score

Dan metrik yang terakhir adalah metrik F1-Score yang merupakan rata-rata harmonis dari presisi dan *recall*.

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Digunakan untuk menyeimbangkan presisi dan recall, terutama jika dataset tidak seimbang (contoh: lebih banyak data netral dibandingkan positif/negatif).

Setelah evaluasi, selanjutnya adalah interpretasi hasil. Interpretasi dilakukan untuk memberikan makna dari hasil evaluasi dan analisis model, sehingga dapat

digunakan untuk pengambilan keputusan. Contoh interpretasi hasil untuk judul berita adalah sebagai berikut:

Judul Berita	Prediksi Sentimen	Interpretasi
Program bantuan sosial disambut baik oleh masyarakat	Positif	Program pemerintah mendapat tanggapan positif.
Kenaikan harga bahan pokok menimbulkan keresahan	Negatif	Masyarakat cenderung khawatir terkait isu ekonomi.
Proyek infrastruktur di Gorontalo menuai banyak kritik	Negatif	Kritik terhadap proyek perlu dievaluasi pemerintah.
Layanan digital Gorontalo semakin mempermudah warga	Positif	Layanan digital dihargai oleh masyarakat.

3.3.2 Pra Pengolahan Data

Pra Pengolahan Data adalah salah satu langkah penting dalam analisis teks, termasuk pada analisis sentimen pada penelitian ini. Langkah ini bertujuan untuk membersihkan dan mempersiapkan data teks agar lebih sesuai untuk analisis atau pelatihan model. Adapun tahapan-tahapan yang dapat dilakukan pada Pra Pengolahan Data adalah sebagai berikut:

a. Case Folding

Case folding adalah proses mengubah semua huruf dalam teks menjadi huruf kecil (*lowercase*). Langkah ini dilakukan untuk memastikan konsistensi, karena huruf besar dan kecil dianggap sama dalam analisis teks. Berikut adalah contoh bagaimana case folding bekerja: Dimana terdapat berita dengan judul “Pemerintah Gorontalo meningkatkan Layanan Digital” dan judul tersebut akan di proses menggunakan case folding sehingga akan menjadi “pemerintah gorontalo meningkatkan layanan digital”.

b. Tokenization

Tokenization adalah proses memecah teks menjadi unit-unit kecil yang disebut *token* (biasanya kata atau kalimat). Langkah ini penting untuk mempermudah analisis teks pada tingkat kata. Berikut adalah contoh dimana berita dengan judul “Pemerintah Gorontalo meningkatkan layanan digital”, setelah dilakukan proses Tokenization hasilnya menjadi ["pemerintah", "gorontalo", "meningkatkan", "layanan", "digital"].

c. Stopword Removal

Stopword removal adalah proses menghapus kata-kata umum yang sering muncul tetapi tidak memiliki makna signifikan dalam analisis teks, seperti "dan", "di", "yang".

d. Stemming

Stemming adalah proses mengubah kata-kata menjadi bentuk dasarnya (akar kata). Misalnya, "berjalan", "berjalanlah", "berjalannya" akan diubah menjadi "jalan". Setelah melalui beberapa proses diatas, berikut adalah data dari proses stemming yang kita lakukan: ["pemerintah", "tingkat", "layan"].

3.3.3 Validasi

Pada penelitian ini, validasi digunakan untuk mengevaluasi keakuratan model analisis sentimen yang dibangun. Validasi dilakukan dengan membandingkan hasil prediksi sentimen menggunakan metode *Support Vector Machine* (SVM) dengan hasil analisis sentimen berbasis lexicon. Pendekatan ini bertujuan untuk memastikan bahwa model dapat mengidentifikasi polaritas sentimen secara akurat sesuai dengan referensi yang telah ditentukan. Adapun prosedur validasi yang dilakukan pada penelitian ini adalah sebagai berikut:

a. Analisis Sentimen Berbasis Lexicon

Pada prosedur ini, Sentimen dihitung dengan menggunakan lexicon sentimen Bahasa Indonesia. Dimana setiap kata dalam judul berita dibandingkan dengan daftar kata polaritas pada lexicon, dan nilai sentimen dihitung sebagai:

- Positif jika skor > 0
- Negatif jika skor < 0

b. Prediksi Model Menggunakan SVM

Judul berita yang telah diproses diklasifikasikan menggunakan model SVM untuk menghasilkan prediksi sentimen (positif atau negatif).

c. Perbandingan Hasil

Hasil sentimen yang diperoleh dari metode lexicon dibandingkan dengan hasil prediksi model. Jika hasil prediksi model sesuai dengan analisis lexicon, maka validasi dianggap valid.

3.3.4 Evaluasi Model

Untuk mengevaluasi kinerja model, kita dapat menggunakan beberapa metrik evaluasi yaitu:

a. Akurasi

Akurasi akan mengukur persentase prediksi yang benar (baik positif maupun negatif) dibandingkan dengan total jumlah prediksi. Kombinasi dari *True Positives* (TP) dan *True Negatives* (TN).

Contoh jika model memprediksi 90 dari 100 data dengan benar, maka akurasi adalah:

$$\text{Akurasi} = \frac{90}{100} = 90\%$$

b. *Precision*

Precision dapat mengukur proporsi prediksi benar pada kelas positif dibandingkan dengan total prediksi positif (baik benar maupun salah).

Contoh jika model memprediksi 50 data sebagai positif, dan 45 di antaranya benar, maka precision adalah:

$$\text{Precision} = \frac{45}{50} = 90\%$$

c. *Recall*

Recall dapat mengukur proporsi data positif yang terdeteksi dengan benar oleh model dibandingkan dengan total data positif sebenarnya. Contoh jika

terdapat 60 data positif, dan model berhasil mendeteksi 45 di antaranya, maka recall adalah:

$$\text{Recall} = \frac{45}{60} = 75\%$$

d. F1-Score

F1-score adalah harmonic mean antara precision dan recall. Metrik ini memberikan keseimbangan antara precision dan recall. Sebagai contoh, Jika precision = 90% dan recall = 75%, maka F1-score adalah:

$$\text{F1-Score} = 2 \cdot \frac{0.9 \cdot 0.75}{0.9 + 0.75} \approx 81.8\%$$

DAFTAR PUSTAKA

- [1] W. H. S. I. V. O. M. R. K. J. N. H. Kencana, “Penggunaan Media Sosial dalam Portal Berita Online,” Jul 2022.
- [2] oleh Nurmaida dan dan Dosen, “ANALISIS KEEFEKTIFAN JUDUL BERITA SERAMBI INDONESIA,” 2023.
- [3] N. Yunita, “ANALISIS SENTIMEN BERITA ARTIS DENGAN MENGGUNAKAN ALGORITMA SUPPORT VECTOR MACHINE DAN PARTICLE SWARM OPTIMIZATION,” no. AGUSTUS, 2016, [Daring]. Tersedia pada: www.tribunnews.com
- [4] A. Suryadi, W. A. Syb'an, N. Alfa'inna, E. H. Hermaliani, dan U. N. Mandiri, “Implementasi Web Scraping dan Sentiment Analysis Terhadap Berita Menggunakan Machine Learning,” *JURNAL SWABUMI*, vol. 11, no. 1, hlm. 2023, 2023.
- [5] Y. A. Hafiz dan E. Sudarmilah, “IMPLEMENTASI WEB SCRAPING PADA PORTAL BERITA ONLINE,” 2023.
- [6] G. Kristian Haroni, F. VIA Koagouw, dan M. D. Himpong, “ANALISIS ISI BERITA COVID-19 MEDIA ONLINE DI KOTA MANADO.”
- [7] Y. Azhar, “METODE LEXICON-LEARNING BASED UNTUK IDENTIFIKASI TWEET OPINI BERBAHASA INDONESIA,” 2017.

LAMPIRAN

Lampiran 1 Timeline Penelitian

No.	Nama Pekerjaan	Tanggal							
		25	26	27	28	29	30	1	2
1	Pengambilan Data								
2	Analisis Kebutuhan								
3	Pembuatan Sistem								
4	Pengujian Sistem								
5	Perbaikan Sistem								