ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES PADA BERITA KEPOLISIAN POLDA KALIMANTAN TENGAH

TUGAS AKHIR

Disusun Untuk Memenuhi Syarat Kelulusan Program Strata I
pada Sekolah Tinggi Manajemen Informatika Dan Komputer
(STMIK) Palangkaraya

OLEH

DHEA AULIDNASARI
C2255201131
PROGRAM STUDI TEKNIK INFORMATIKA

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
2025

ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES PADA BERITA KEPOLISIAN POLDA KALIMANTAN TENGAH

TUGAS AKHIR

Disusun Untuk Memenuhi Syarat Kelulusan Program Strata I
pada Sekolah Tinggi Manajemen Informatika Dan Komputer
(STMIK) Palangkaraya

OLEH

DHEA AULIDNASARI
C2255201131
PROGRAM STUDI TEKNIK INFORMATIKA

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
2025

PERSETUJUAN

ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES PADA BERITA KEPOLISIAN POLDA KALIMANTAN TENGAH

Tugas Akhir Ini Telah Diujikan dan Disahkan

Pembimbing I,

Lili Rusdiana, M.Kom
NIK. 198707282011007

Pembimbing II,

Sam'ani, S.T., M.Kom
NIK. 197703252005105

Mengetahui
Ketua STMIK Palangkaraya,

Ir. Hj. Siti Maryamah, M.M
NIK. 196902021995004

Mengesahkan
Ketua Prodi Teknik Informatika,

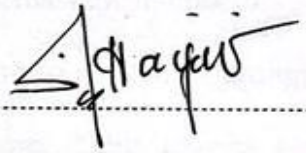
Ir. Abdul Hadi, S.T., M.Kom
NIK. 198505102021101

PENGESAHAN

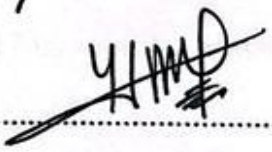
ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES PADA BERITA KEPOLISIAN POLDA KALIMANTAN TENGAH

Tugas Akhir ini telah Diujikan, Dinilai, dan Disahkan
Oleh Tim Seminar pada Tanggal 20 Desember 2025

Tim Sidang Tugas Akhir :

1. Dr. Maura Widyaningsih, S.Kom., M.Cs. ..... 
   Ketua

2. Ir. Abdul Hadi, S.T., M.Kom. ..... 
   Sekretaris

3. Veny Cahya Hardita, M.Kom. ..... 
   Anggota

4. Lili Rusdiana, M.Kom. ..... 
   Anggota

5. Catharina Elmayantie, M.Pd. ..... 
   Anggota

PERNYATAAN KEASLIAN

Yang bertanda tangan di bawah ini:

Nama Mahasiswa : Dhea Aulidnasari

N I M : C2255201131

menyatakan bahwa Tugas Akhir dengan judul :

ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES PADA BERITA KEPOLISIAN POLDA KALIMANTAN TENGAH

adalah hasil karya saya dan bukan merupakan duplikasi sebagian atau seluruhnya dari karya orang lain, kecuali bagian yang sumber informasi dicantumkan.

Pernyataan ini dibuat dengan sebenar-benarnya secara sadar dan bertanggungjawab dan saya bersedia menerima sanksi pembatalan Tugas Akhir apabila terbukti melakukan duplikasi terhadap Tugas Akhir atau karya ilmiah lain yang sudah ada.

Palangka Raya, 15 Desember 2025

Yang Membuat Pernyataan,



DHEA AULIDNASARI

MOTTO DAN PERSEMBAHAN

Boleh jadi kamu tidak menyenangi sesuatu padahal itu baik bagimu, dan boleh jadi kamu menyukai sesuatu padahal itu tidak baik bagimu. Allah SWT mengetahui, sedang kamu tidak mengetahui.

Allah SWT Maha Mengetahui segala takdir terbaik untuk manusia, meskipun terkadang terasa berat atau tidak sesuai keinginan, sementara manusia seringkali tidak mengetahui hikmah di baliknya (QS. Al-Baqarah; 216)

Tugas Akhir Ini Kupersembahkan Untuk :

1. Terimakasih untuk Suami dan Orang Tua yang sudah selalu mensupport dan memberikan yang terbaik selama menempuh pembelajaran.
2. Terimakasih untuk semua Dosen yang ikut serta membantu dan memberikan dorongan agar selalu berani dalam berproses selama menempuh Pendidikan di Universitas STMIK Palangkaraya.

INTISARI

Dhea Aulidnasari, C2255201131, 2025. ***Analisis Sentimen Menggunakan Naïve Bayes Pada Berita Kepolisian Polda Kalimantan Tengah,*** Pembimbing I Lili Rusdiana, M.Kom., Pembimbing II Sam'ani, S.T., M.Kom.

Perkembangan media *online* dan media sosial telah mengubah cara masyarakat memperoleh serta menyampaikan informasi, termasuk informasi terkait institusi kepolisian. Pemberitaan dan komentar masyarakat di dunia maya berperan penting dalam membentuk citra dan kepercayaan publik terhadap kepolisian. Oleh karena itu, diperlukan suatu pendekatan analitis yang mampu mengidentifikasi dan memahami sentimen masyarakat secara sistematis terhadap informasi kepolisian berbasis online.

Penelitian ini bertujuan untuk menganalisis sentimen masyarakat terhadap Polda Kalimantan Tengah berdasarkan komentar pada media sosial *Instagram* menggunakan algoritma *Naïve Bayes*. Jenis penelitian yang digunakan adalah penelitian deskriptif kuantitatif dengan pendekatan *text mining*. Teknik pengumpulan data dilakukan dengan mengambil komentar masyarakat dari akun *Instagram* resmi Polda Kalimantan Tengah sebanyak 50 komentar. Data kemudian diberi label sentimen positif dan negatif secara manual. Tahapan penelitian meliputi *preprocessing teks*, ekstraksi fitur menggunakan *Term Frequency–Inverse Document Frequency* (TF-IDF), pembagian data menjadi data latih dan data uji, pelatihan model *Naïve Bayes*, serta evaluasi kinerja model menggunakan *Confusion Matrix* dan metrik evaluasi.

Hasil penelitian menunjukkan bahwa algoritma *Naïve Bayes* mampu mengklasifikasikan sentimen komentar masyarakat dengan tingkat akurasi yang sangat baik. Mayoritas komentar yang dianalisis menunjukkan sentimen positif terhadap kinerja dan pelayanan Polda Kalimantan Tengah. Kesimpulan dari penelitian ini adalah bahwa analisis sentimen berbasis *Naïve Bayes* efektif digunakan untuk memantau opini publik di media *online* dan dapat dijadikan sebagai alat bantu evaluasi dalam meningkatkan kualitas pelayanan serta komunikasi publik institusi kepolisian.

Kata Kunci : *Naïve Bayes*, Sistem Pakar, Polda Kalteng, TF-IDF, Sentimen

ABSTRACT

Dhea Aulidnasari, C2255201131, 2025. *Sentiment Analysis Using Naïve Bayes On Police News From The Central Kalimantan Regional Police*, Advisor I Lili Rusdiana, M.Kom., Advisor II Sam'ani, S.T., M.Kom.

The development of online and social media has changed the way the public obtains and conveys information, including information related to the police. Online news and public comments play a crucial role in shaping the public's image and trust in the police. Therefore, an analytical approach capable of systematically identifying and understanding public sentiment regarding online police information is needed.

This study aims to analyze public sentiment toward the Central Kalimantan Regional Police (Polda Kalimantan Tengah) based on comments on Instagram using the Naïve Bayes algorithm. This study employed a quantitative descriptive approach with a text mining approach. Data collection involved collecting 50 comments from the official Instagram account of the Central Kalimantan Regional Police. The data were then manually labeled with positive and *negative* sentiments. The research steps included text preprocessing, feature extraction using Term Frequency–Inverse Document Frequency (TF-IDF), data division into training and test data, training a Naïve Bayes model, and evaluating model performance using a Confusion Matrix and evaluation metrics.

The results showed that the Naïve Bayes algorithm was able to classify public comment sentiment with a very good level of accuracy. The majority of comments analyzed indicated positive sentiment toward the performance and services of the Central Kalimantan Regional Police. The conclusion of this study is that Naïve Bayes-based sentiment analysis is effective for monitoring public opinion in online media and can be used as an evaluation tool to improve the quality of services and public communication of police institutions..

Keywords: Naïve Bayes, Expert System, Central Kalimantan Regional Police, TF-IDF, Sentiment

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR

Sebagai sivitas akademik Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya, saya yang bertanda tangan dibawah ini:

Nama : Dhea Aulidnasari

NIM : C2255201131

Program Studi : Teknik Informatika

Jenis Karya : Laporan Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya, Hak Bebas Royalti Non Eksklusif (*Non-exclusive Royalty-Free Right*) atas karya ilmiah saya yang berjudul:

ANALISIS SENTIMEN MENGGUNAKAN NAÏVE BAYES PADA BERITA KEPOLISIAN POLDA KALIMANTAN TENGAH

Dengan Hak Bebas Royalti Non Eksklusif ini Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya berhak menyimpan, mengalih media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Palangka Raya
Pada tanggal :
Yang menyatakan



Dhea Aulidnasari

KATA PENGANTAR

Puji dan syukur kepada Allah SWT berkat Rahmat, Hidayah, dan Karunia-Nya sehingga saya sebagai penulis dapat menyelesaikan skripsi dengan judul "Analisis Sentimen Menggunakan Naïve Bayes Pada Berita Kepolisian Polda Kalimantan Tengah" dengan lancar dan tepat waktu. Oleh karena itu, penulis menyampaikan banyak terima kasih kepada segenap pihak yang telah berkontribusi secara maksimal dalam penyelesaian skripsi ini.

Tidak lupa dalam kesempatan ini saya sebagai penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Ir. Hj. Siti Maryamah, M.M selaku Ketua STMIK Palangka Raya.
2. Ir. Abdul Hadi, S.T., M.Kom selaku Ketua Prodi Teknik Informatika
3. Lili Rusdiana, M.Kom, selaku Dosen Pembimbing I.
4. Sam'ani, S.T., M.Kom. selaku Dosen Pembimbing II.

Akhir kata dari saya semoga hasil penelitian ini dapat memberikan manfaat bagi kita semua khusus untuk lingkungan program studi Teknik informatika STMIK Palangka Raya.

Palangka Raya, Desember 2025

Dhea Aulidnasari
C2255201131

DAFTAR ISI

PENGESAHAN iv
PERNYATAAN KEASLIAN v
MOTTO DAN PERSEMBAHAN vi
INTISARI vii
ABSTRACT viii
PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR ix
KATA PENGANTAR x
DAFTAR ISI xi
DAFTAR TABEL xii
DAFTAR GAMBAR xiv
DAFTAR LAMPIRAN xv
BAB I PENDAHULUAN 1
1.1 Latar Belakang Masalah 1
1.2 Rumusan Masalah 4
1.3 Batasan Masalah 4
1.4 Tujuan dan Manfaat 5
1.5 Sistematika Penulisan 6
BAB II TINJAUAN PUSTAKA 8
2.1 Kajian Teori 8
2.2 Penelitian Yang Relevan 24
BAB III METODOLOGI PENELITIAN 26
3.1 Tinjauan Umum 26
3.2 Jenis penelitian 26
3.3 Desain Penelitian 27
3.4 Instrumen Penelitian 33
3.5 Analisis Kebutuhan 34
BAB IV PEMBAHASAN 39
4.1 Hasil 39
4.2 Implementasi 60
BAB V KESIMPULAN DAN SARAN 64
5.1 KESIMPULAN 64
5.2 SARAN 65
DAFTAR PUSTAKA

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu 11
Tabel 3.1 Contoh Dataset 16
Tabel 3.2 Struktur Tabel Komentar 21
Tabel 3.3 Struktur Tabel *Processing* 22
Tabel 3.4 Struktur Tabel Sentimen 22
Tabel 3.5 Jadwal penelitian Tugas Akhir 23
Tabel 4.1 Dataset Sampel 41
Tabel 4.2 Document Frequency (DF) 45
Tabel 4.3 Perhitungan IDF 46
Tabel 4.4 Perhitungan IDF Komentar 1 46
Tabel 4.5 Perhitungan IDF Komentar 2 46
Tabel 4.6 Perhitungan IDF Komentar 3 46
Tabel 4.7 Perhitungan IDF Komentar 4 46
Tabel 4.8 Perhitungan IDF Komentar 5 47
Tabel 4.9 Perhitungan IDF Komentar 6 47
Tabel 4.10 Perhitungan IDF Komentar 7 47
Tabel 4.11 Perhitungan IDF Komentar 8 47
Tabel 4.12 Perhitungan IDF Komentar 9 48
Tabel 4.12 Perhitungan IDF Komentar 10 48
Tabel 4.13 Perhitungan IDF Komentar 11 48
Tabel 4.14 Perhitungan IDF Komentar 12 48
Tabel 4.15 Perhitungan IDF Komentar 13 49
Tabel 4.16 Perhitungan IDF Komentar 14 49
Tabel 4.17 Perhitungan IDF Komentar 15 49
Tabel 4.18 Perhitungan IDF Komentar 16 49
Tabel 4.19 Perhitungan IDF Komentar 17 50
Tabel 4.20 Perhitungan IDF Komentar 18 51
Tabel 4.21 Perhitungan IDF Komentar 19 51
Tabel 4.22 Perhitungan IDF Komentar 20 51
Tabel 4.23 Perhitungan IDF Komentar 21 51
Tabel 4.24 Perhitungan IDF Komentar 22 51
Tabel 4.25 Perhitungan IDF Komentar 23 52
Tabel 4.26 Perhitungan IDF Komentar 24 52
Tabel 4.27 Perhitungan IDF Komentar 25 52
Tabel 4.28 Perhitungan IDF Komentar 26 52
Tabel 4.29 Perhitungan IDF Komentar 27 52
Tabel 4.30 Perhitungan IDF Komentar 28 54
Tabel 4.31 Perhitungan IDF Komentar 29 54
Tabel 4.32 Perhitungan IDF Komentar 30 54
Tabel 4.33 Perhitungan IDF Komentar 31 54
Tabel 4.34 Perhitungan IDF Komentar 32 55
Tabel 4.35 Perhitungan IDF Komentar 33 55
Tabel 4.36 Perhitungan IDF Komentar 34 55
Tabel 4.37 Perhitungan IDF Komentar 35 55

Tabel 4.38 Perhitungan IDF Komentar 36 56
Tabel 4.39 Perhitungan IDF Komentar 37 56
Tabel 4.40 Perhitungan IDF Komentar 38 56
Tabel 4.41 Perhitungan IDF Komentar 39 56
Tabel 4.42 Perhitungan IDF Komentar 40 57
Tabel 4.43 Perhitungan IDF Komentar 41 57
Tabel 4.44 Perhitungan IDF Komentar 41 57
Tabel 4.45 Perhitungan IDF Komentar 42 57
Tabel 4.46 Perhitungan IDF Komentar 43 58
Tabel 4.47 Perhitungan IDF Komentar 45 58
Tabel 4.48 Perhitungan IDF Komentar 46 58
Tabel 4.49 Perhitungan IDF Komentar 47 58
Tabel 4.50 Perhitungan IDF Komentar 48 59
Tabel 4.51 Perhitungan IDF Komentar 49 59
Tabel 4.52 Perhitungan IDF Komentar 50 59

DAFTAR GAMBAR

Gambar 3.1 Analisis Proses 18
Gambar 3.2 Desain *Flowchart Naïve Bayes* 20
Gambar 4.1 Tahapan Data Mining 39
Gambar 4.1 Stopword Removal 1 42
Gambar 4.2 Stopword Removal 2 42
Gambar 4.3 Stopword Removal 3 42
Gambar 4.4 Stopword Removal 4 42
Gambar 4.5 Stopword Removal 5 43
Gambar 4.6 Stopword Removal 6 43
Gambar 4.7 Stopword Removal 7 43
Gambar 4.8 Mulai Perhitungan TF IDF 46
Gambar 4.9 Memanggil Komentar TF IDF 47
Gambar 4.10 Memanggil Fungsi Pencarian Komentar TF IDF 47
Gambar 4.11 Hasil Komentar TF IDF 48
Gambar 4.12 Hasil Komentar TF IDF 61
Gambar 4.13 Hasil Komentar TF IDF 61

DAFTAR LAMPIRAN

Lampiran 1. Surat Tugas Pembimbing Tugas Akhir
Lampiran 2. Surat Tugas Penguji Seminar Proposal Tugas Akhir
Lampiran 3. Kartu Kegiatan Konsultasi Tugas Akhir
Lampiran 4. Berita Acara Seminar Proposal Tugas Akhir

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Perkembangan teknologi informasi yang pesat telah mengubah cara masyarakat memperoleh dan menyebarkan informasi. Akses terhadap berita kini tidak lagi terbatas pada media cetak atau siaran televisi, melainkan telah beralih ke platform digital yang dapat diakses kapan saja melalui internet. Fenomena ini memunculkan berbagai portal berita *online* yang secara aktif menyajikan informasi dari berbagai bidang kehidupan, termasuk berita terkait kepolisian. Dalam konteks tersebut, berita kepolisian sering menjadi sorotan publik karena berhubungan langsung dengan keamanan, keadilan, dan kepercayaan masyarakat terhadap institusi penegak hukum.

Namun, tidak semua berita yang beredar di dunia maya disampaikan secara objektif. Banyak berita yang disertai opini, narasi emosional, bahkan terkadang mengandung bias tertentu. Hal ini dapat memengaruhi persepsi masyarakat terhadap institusi kepolisian, baik dalam bentuk sentimen positif dan negatif. Oleh karena itu, diperlukan sebuah pendekatan analitis yang mampu memetakan opini publik berdasarkan isi berita, sehingga dapat memberikan gambaran yang lebih jelas tentang bagaimana masyarakat menilai kinerja kepolisian melalui pemberitaan *online*.

Menurut Effendy (2011), opini publik terbentuk sebagai hasil dari interaksi masyarakat dengan informasi yang disampaikan melalui media massa maupun media online. Dalam era digital, media online dan media sosial menjadi sarana

utama pembentukan opini publik karena memungkinkan masyarakat menyampaikan pendapat secara terbuka dan cepat melalui kolom komentar, unggahan, maupun diskusi daring.

Analisis sentimen merupakan salah satu cabang dari text mining yang berfungsi untuk mengidentifikasi dan mengklasifikasikan opini atau perasaan seseorang terhadap suatu objek. Dalam konteks penelitian ini, analisis sentimen dapat digunakan untuk mengetahui kecenderungan opini publik terhadap berita-berita yang memuat topik kepolisian. Melalui proses ini, informasi yang semula bersifat kualitatif dapat diubah menjadi data kuantitatif yang mudah diukur dan dianalisis secara sistematis.

Menurut Han et all (2012), salah satu metode yang populer dan efektif dalam penerapan analisis sentimen adalah algoritma *Naïve Bayes*. Algoritma ini termasuk ke dalam kategori supervised learning yang bekerja berdasarkan prinsip teorema *Bayes*, dengan asumsi bahwa setiap fitur atau kata bersifat independen. Keunggulan *Naïve Bayes* terletak pada kemampuannya dalam mengolah data teks dengan cepat, sederhana, namun tetap memiliki tingkat akurasi yang kompetitif dibandingkan dengan algoritma lain. Karena itulah, metode ini banyak digunakan dalam berbagai penelitian terkait klasifikasi teks seperti spam detection, analisis opini produk, hingga analisis berita *online*.

Dalam konteks berita kepolisian, penerapan analisis sentimen berbasis *Naïve Bayes* menjadi penting untuk memahami persepsi publik yang terbentuk di dunia digital. Pemberitaan mengenai kegiatan kepolisian, penegakan hukum, kasus kriminal, atau tindakan aparat sering kali menimbulkan beragam opini masyarakat.

Dengan menggunakan analisis sentimen, dapat diketahui apakah berita-berita tersebut cenderung memunculkan citra positif dan negatif terhadap institusi kepolisian. Informasi ini dapat menjadi dasar evaluasi bagi pihak kepolisian dalam memperbaiki strategi komunikasi publik dan meningkatkan kepercayaan masyarakat.

Selain itu, analisis sentimen juga dapat berperan sebagai alat pemantau (*monitoring tool*) terhadap dinamika opini publik di media *online*. Melalui klasifikasi sentimen secara otomatis, pihak kepolisian maupun instansi terkait dapat mengetahui isu-isu yang paling banyak dibicarakan serta memahami kecenderungan reaksi masyarakat terhadap suatu peristiwa tertentu. Hal ini dapat membantu dalam pengambilan keputusan strategis, terutama dalam hal penyusunan kebijakan komunikasi dan publikasi agar lebih responsif terhadap opini publik.

Dari sisi akademik, penelitian mengenai analisis sentimen pada berita kepolisian juga memiliki kontribusi penting dalam pengembangan ilmu pengetahuan, khususnya di bidang *natural language processing* (NLP) dan kecerdasan buatan (*artificial intelligence*). Dengan meningkatnya volume data teks di internet, kebutuhan akan sistem analisis otomatis yang mampu memahami konteks emosional dan opini manusia menjadi semakin mendesak. Penelitian ini diharapkan dapat memperkaya pemahaman tentang bagaimana algoritma *Naïve Bayes* dapat diterapkan secara efektif pada data berita berbahasa Indonesia.

Meskipun telah banyak penelitian mengenai analisis sentimen di berbagai bidang seperti *e-commerce*, politik, dan pariwisata, namun penerapan metode ini pada konteks berita kepolisian di Indonesia masih relatif terbatas. Padahal, media

online memainkan peran besar dalam membentuk citra institusi kepolisian di mata publik. Dengan menganalisis sentimen berita kepolisian secara sistematis, dapat diperoleh wawasan yang lebih objektif mengenai bagaimana kinerja dan tindakan aparat diterima oleh masyarakat.

Berdasarkan latar belakang tersebut, penelitian ini berfokus pada "Peran Analisis Sentimen Berbasis *Naïve Bayes* pada Berita Kepolisian Berbasis *Online*." Tujuan utama penelitian ini adalah untuk menganalisis sentimen berita kepolisian menjadi kategori positif dan negatif menggunakan algoritma *Naïve Bayes*, serta menganalisis bagaimana hasil klasifikasi tersebut dapat berperan dalam memahami persepsi publik terhadap kinerja kepolisia.

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas peneliti dapat merumuskan masalah yaitu bagaimana cara metode *Naïve Bayes* melakukan analisis sentiment pada berita kepolisian yang diakses secara *online*?

1.3 Batasan Masalah

Agar pembahasan tugas akhir yang dilakasakan lebih terfokus dan terarah ke permasalahan yang ada, maka penulis membatasi ruang lingkup permasalahan yang dibahas yaitu :

a. Analisis sentimen dilakukan dengan menggunakan algoritma *Naïve Bayes* sebagai metode utama untuk mengklasifikasikan sentimen ke dalam dua kategori, yaitu positif dan negatif.
b. Analisis sentiment ini hanya bersumber dari data yang diakses secara *online* melalui komentar disosial media seperti *instagram* dan portal website.

c. Hasil penelitian ini difokuskan untuk memberikan gambaran umum tentang persepsi publik terhadap berita kepolisian, bukan untuk menilai kebenaran isi berita atau melakukan verifikasi fakta.
d. Analisis sentiment ini menggunakan *google colab* dalam menganalisis secara sistem.
e. Hasil penelitian ini hanya menganalisis sentiment

1.4 Tujuan dan Manfaat

a. Tujuan

Penelitian ini diharapkan dapat memberikan manfaat bagi mahasiswa sebagai tambahan referensi ilmiah, khususnya terkait penerapan metode *Naïve Bayes* dalam analisis sentimen, sehingga dapat digunakan sebagai acuan dalam penelitian sejenis. Bagi pihak kepolisian, hasil penelitian ini diharapkan dapat menjadi bahan evaluasi dan acuan dalam memahami sentimen serta persepsi masyarakat terhadap kinerja dan pelayanan kepolisian di media online. Selain itu, bagi penulis atau peneliti, penelitian ini diharapkan dapat menambah wawasan dan pengetahuan, serta menjadi dasar pengembangan penelitian selanjutnya yang berkaitan dengan analisis sentimen dan pemanfaatan data media sosial.

b. Manfaat

1) Bagi Mahasiswa

Penelitian ini dapat menjadi referensi tambahan bagi mahasiswa yang ingin melakukan kajian serupa di bidang sistem cerdas.

2) Bagi Kepolisian Polda Kalteng

Hasil penelitian ini diharapkan dapat memberikan gambaran nyata mengenai kondisi masyarakat yang menilai kepolisian dari berbagai aspek.

3) Bagi Penulis/Peneliti

Penelitian ini memberikan pengalaman langsung dan pemahaman mendalam mengenai proses analisis sentimen berbasis *Naïve Bayes*. Khususnya terkait analisis negatif dan positif dalam hal berbasis digital di instansi pemerintahan.

1.5 Sistematika Penulisan

Adapun sistematika penulisan dalam penulisan tugas akhir ini terdiri dari bebrapa bab dan masing-masing bab membahas dan menguraikan pokok permasalahan yang berbeda, sebagai gambaran disini penulis menyertakan garis-garis besarnya yaitu :

BAB I PENDAHULUAN

Bab ini berisikan latar belakang masalah, rumusan masalah, Batasan masalah, tujuan dan manfaat, serta sistematika penulisan.

BAB II LANDASAN TEORI

Bab ini berisikan tentang tinjauan pustaka yang diambil dari penelitian yang relevan beserta susunan kajian teori yang disesuaikan dengan tema Tugas Akhir.

BAB III METODOLOGI PENELITIAN

Bab ini berisikan tentang tahapan yang dilakukan peneliti dalam mengumpulkan informasi atau data yang dibutuhkan.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisikan tentang hasil dari penelitian dan pembahasan dari hasil yang didapat.

BAB V KESIMPULAN DAN SARAN

Bab ini berisikan tentang kesimpulan dan saran yang diberikan untuk pengembangan selanjutnya

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Teori

2.1.1 Analisis Sentimen (*Sentiment Analysis*)

Analisis sentimen atau opinion mining merupakan cabang dari pengolahan bahasa alami *Natural language Processing* (NLP) yang berfokus pada identifikasi dan klasifikasi opini, sikap, atau emosi yang terkandung dalam sebuah teks terhadap suatu entitas, topik, atau isu.

Dalam konteks ini, entitas bisa berupa produk, layanan, individu, institusi, atau peristiwa tertentu. Analisis sentimen tidak hanya sekadar menentukan apakah teks bersifat positif atau negatif, tetapi juga dapat memperhatikan tingkat subjektivitas (seberapa banyak opini dalam teks), target opini (apakah opini ditujukan pada keseluruhan entitas atau aspek tertentu), serta konteks atau aspek yang dipersepsikan.

Oleh karena itu, penelitian sering membedakan berbagai level analisis: level dokumen, level kalimat, level aspek/fitur. Dari sisi praktis, analisis sentimen sangat berguna dalam berbagai bidang seperti pemasaran, manajemen reputasi, politik, analisis kepercayaan publik, hingga riset sosial.

Melalui analisis sentimen, organisasi atau institusi dapat memahami bagaimana opini publik terbentuk dan berkembang, serta bagaimana reaksi masyarakat terhadap suatu isu atau kebijakan.

Menurut Gunasekaran (2023) metodologi yang umum digunakan dalam analisis sentimen meliputi pendekatan berbasis leksikon (*lexicon-based*),

pendekatan berbasis pembelajaran mesin (*machine learning*), serta kombinasi atau hibrida antara keduanya.

Menurut Azza (2025) pendekatan leksikon menggunakan kamus kata yang telah diberi skor sementara *machine learning* melatih model berdasarkan data berlabel. Tahapan umum dalam proses analisis sentimen meliputi pengumpulan data teks (misalnya komentar, artikel berita, posting media sosial), pra-pengolahan teks (tokenisasi, penghilangan *stopword*, *stemming*), ekstraksi fitur, pelatihan model (jika *machine learning*), klasifikasi, dan evaluasi kinerja model (misalnya akurasi, presisi, *recall, F-measure*).

Proses ini memastikan bahwa data teks yang sangat tidak terstruktur dapat diubah menjadi hasil yang bermakna dalam bentuk kategori sentimen. Dalam penelitian atau aplikasi yang relevan, analisis sentimen sering digunakan untuk memantau perubahan opini publik secara real-time dalam media *online* atau media sosial, termasuk untuk isu institusi, kebijakan publik, atau layanan masyarakat.

Misalnya, penelitian tentang sentimen terhadap institusi kepolisian dapat mengambil data berita *online* atau komentar publik untuk mengetahui kecenderungan opini masyarakat.

Meski banyak kegunaan, analisis sentimen menghadapi tantangan yang signifikan seperti: ironi/sarkasme, konteks kultural dan linguistik (termasuk bahasa indonesia), aspek multibahasa, ambiguitas makna, dan kualitas data yang buruk (misalnya teks pendek, slang, *typo*).

Oleh karena itu, adaptasi metode terhadap karakteristik data sangat penting agar hasil klasifikasi sentimen lebih akurat dan bermakna.

Kesimpulannya, analisis sentimen merupakan kerangka teoritik dan teknis yang memungkinkan transformasi data teks subjektif menjadi informasi kuantitatif tentang opini publik atau pengguna. Dalam konteks penelitian mengenai berita kepolisian berbasis *online*, teori ini menjadi landasan untuk memahami bagaimana opini publik dapat diklasifikasikan dan dievaluasi melalui algoritma klasifikasi. Dengan demikian, penguasaan teori analisis sentimen sangat krusial agar penelitian berjalan dengan metodologi yang valid dan relevan.

2.1.2 Media Siber

Menurut (Nasrullah, 2022:2) Media *online* atau media siber (*cybermedia*) merupakan sebuah bentuk media massa yang beroperasi melalui jaringan internet dan situs web, memungkinkan jangkauan yang lebih luas serta interaksi yang lebih dinamis dibandingkan media tradisional seperti koran, majalah, radio, atau televisi. Menurut definisi, media *online* disebut juga *cybermedia*, *internet media* atau "*new media*".

Misalnya, dalam penelitian disebutkan bahwa media *online* adalah media yang tersaji secara *online* di situs web (website) internet.

Menurut (Meisyaroh, 2023:12) *Cyber*/siber adalah sesuatu yang berhubungan dengan sistem komputer dan informasi. Dalam perkembangannya, siber dapat diartikan yang berhubungan dengan internet dan dikenal juga sebagai *cyberspace*. Selanjutnya dalam pembahasan bab ini digunakan istilah media siber. Media siber adalah istilah yang digunakan untuk menggambarkan jenis media yang digunakan dalam dunia maya atau internet. Ini termasuk website, blog, forum, media sosial, aplikasi pesan, dan lain-lain. Media siber memiliki karakter khusus sehingga

memerlukan pedoman agar pengelolaannya dapat dilaksanakan secara profesional, memenuhi fungsi, hak, dan kewajibannya sesuai Undang-Undang Nomor 40 Tahun 1999 tentang Pers dan Kode Etik Jurnalistik

Salah satu aspek khas media *online* adalah kemampuannya untuk menyediakan konten secara real-time dan memperbarui informasi dengan cepat serta fleksibel. Sebagai media generasi ketiga setelah media cetak dan media elektronik, media *online* memiliki karakteristik seperti multimedia (teks + audio + video), update cepat, dan sifat fleksibel dalam penerbitan. Misalnya, disebut bahwa ciri-ciri media *online* meliputi multimedia, aktualitas, kecepatan, dan kapasitas yang luas.

Dari perspektif kajian komunikasi dan media, media siber juga dikenal sebagai "*new media*" karena menandai pergeseran paradigma dari pola satu ke banyak (*one to many*) yang dominan pada media tradisional menjadi pola yang lebih interaktif baik banyak ke banyak (*many to many*) maupun sedikit ke sedikit (*few to few*) atau sedikit ke banyak (*few to many*).

Lebih jauh, media *online* menghadirkan aspek partisipasi pengguna yang lebih besar pengguna dapat memberi komentar, berbagi konten, bergabung komunitas daring, serta membentuk opini kolektif melalui interaksi.

Faktor ini menjadikan media *online* sebagai ruang publik digital yang memungkinkan transfer opini, diskusi terbuka, dan konstruksi persepsi masyarakat secara cepat dan luas.

Dalam konteks jurnalistik, media *online* menuntut adaptasi terhadap kaidah baru seperti fleksibilitas waktu terbit, kapasitas yang besar (halaman tak terbatas),

dan integrasi format multimedia. Sebagai contoh, sebuah kajian menunjukkan bahwa media *online* sebagai produk jurnalistik *online* didefinisikan sebagai pelaporan fakta atau peristiwa yang diproduksi dan didistribusikan melalui internet.

Dengan demikian, aktivitas jurnalistik di media *online* harus memperhatikan kecepatan, multimodalitas, dan konektivitas dengan audiens secara langsung.

Meski memiliki banyak keunggulan, media *online* juga menghadapi tantangan yang khas. Beberapa di antaranya adalah kontrol redaksional yang kadang lebih longgar dibanding media tradisional, potensi penyebaran berita hoaks atau informasi yang kurang terverifikasi, serta tekanan untuk "*click bait*" atau sensasional demi menarik perhatian pengguna. Contoh kajian menyebut bahwa media siber menawarkan ruang terbuka bagi khalayak untuk membuat dan mengonsumsi konten, sehingga muncul isu seperti hoaks atau berita bersarang.

Tantangan-ini penting untuk diperhatikan dalam penelitian yang meninjau pemberitaan media *online*.

2.1.4 Deskriptif Kuantitatif

Menurut Sari (2022:5) Pendekatan deskriptif kuantitatif adalah salah satu jenis penelitian yang bertujuan untuk menggambarkan suatu fenomena secara terukur menggunakan data numerik. Penelitian ini tidak mencari hubungan sebab-akibat secara mendalam, melainkan berfokus pada pemaparan dan interpretasi data apa adanya berdasarkan hasil pengukuran. Dengan demikian, penelitian deskriptif kuantitatif digunakan untuk mengetahui bagaimana suatu variabel muncul, seberapa besar frekuensinya, serta bagaimana distribusi data yang terjadi pada objek penelitian.

Penelitian deskriptif kuantitatif mengandalkan angka sebagai dasar analisis sehingga seluruh prosesnya, mulai dari pengumpulan data hingga penyajian hasil, dilakukan dengan pendekatan statistik. Data yang diperoleh akan dikodekan dalam bentuk angka, kemudian dianalisis menggunakan tabel, grafik, atau perhitungan tertentu sehingga mempermudah peneliti membandingkan, mengelompokkan, dan menggambarkan kondisi variabel yang diteliti. Pendekatan ini sering digunakan untuk penelitian berbasis survei, analisis data komentar, maupun pengamatan yang membutuhkan ukuran pasti.

Dalam konteks analisis sentimen, penelitian deskriptif kuantitatif digunakan untuk menggambarkan kecenderungan sentimen masyarakat berdasarkan data komentar yang dianalisis. Sentimen positif dan *negative* dapat dihitung frekuensinya sehingga menghasilkan gambaran umum persepsi masyarakat terhadap suatu instansi, dalam hal ini Polda Kalteng. Proses ini tidak menganalisis mengapa sentimen tersebut muncul, tetapi hanya menjelaskan berapa banyak dan bagaimana distribusinya.

Pendekatan ini juga memberikan kejelasan dan objektivitas karena data dianalisis berdasarkan angka dan tidak dipengaruhi oleh interpretasi subjektif peneliti. Oleh karena itu, kesimpulan yang dihasilkan lebih transparan dan mudah diuji ulang. Metode ini sangat relevan jika penelitian memerlukan visualisasi data yang terukur seperti diagram batang, diagram *pie*, atau tabel statistik lainnya.

Teknik-teknik yang digunakan dalam penelitian deskriptif kuantitatif meliputi perhitungan persentase, distribusi frekuensi, diagram, dan penyajian data statistik sederhana seperti mean, median, atau modus apabila diperlukan. Namun,

penelitian deskriptif kuantitatif tidak melakukan uji hipotesis yang kompleks karena fokus utamanya adalah pada deskripsi fenomena, bukan pengujian hubungan kausal atau pengaruh antar variabel.

Dalam penelitian ini, pendekatan deskriptif kuantitatif diterapkan untuk menyajikan data hasil klasifikasi sentimen yang telah diproses menggunakan metode *Naïve Bayes*. Hasil analisis dikategorikan ke dalam kelas positif dan *negative* kemudian dihitung jumlahnya untuk menggambarkan bagaimana kecenderungan opini masyarakat terhadap berita dan aktivitas kepolisian di dunia maya. Dengan demikian, metode ini mampu memberikan pemahaman awal yang terukur dan komprehensif mengenai persepsi publik.

2.1.3 Google Colab

Google Colab (singkatan dari *Colaboratory*) adalah sebuah layanan berbasis web yang disediakan oleh *Google Research* yang memungkinkan pengguna menulis dan menjalankan kode *Python* langsung melalui browser, tanpa perlu pemasangan lokal.

Beberapa fitur utama *Google Colab* antara lain integrasi dengan *Google Drive*, akses gratis ke sumber daya komputasi seperti GPU dan TPU, serta kemampuan kolaborasi yang mempermudah pembagian dan pengeditan *notebook* bersama pengguna lain. Platform ini banyak dimanfaatkan dalam penelitian dan pembelajaran, termasuk pengembangan algoritma analisis sentimen, karena mempermudah eksperimen, evaluasi model, dan visualisasi hasil secara interaktif. Meski demikian, *Google Colab* memiliki keterbatasan seperti *runtime* terbatas, ketergantungan pada koneksi internet, dan lingkungan yang dapat berubah,

sehingga peneliti perlu mempertimbangkan strategi mitigasi untuk menjaga konsistensi dan *reproducibility* eksperimen.

Dari sisi perangkat lunak, *Google Colab* telah menyediakan hampir sebagian besar pustaka (*library*) yang dibutuhkan. Dalam penelitian ini pustaka yang dibutuhkan adalahKeras, *TensorFlow*, *NumPy*, *Pandas*, dan pendukung lainnya, misalnya untuk pembuatan grafik lewat Matplotlib. Bahkan seluruhversi disediakan, misalnya *Tensor Flow* versi 1x maupun 2.x tersedia, begitu juga versi *Python* yang mulai dari versi 2.x hingga 3.x Dari sisi perangkat keras, *Google Colab* menyediakan layananberupamedia penyimpan yang terintegrasi dengan *Google Drive*, *prosesor* yang berupa CPU, GPU, dan TPU, serta RAM (Guantara, 2023).

2.1.4 Kepolisian

Kepolisian merupakan institusi penegak hukum yang memiliki peran strategis dalam menjaga keamanan, ketertiban, dan perlindungan masyarakat. Dalam kajian kepolisian modern, terdapat beberapa teori dan model teoretis yang relevan untuk memahami fungsi, strategi, dan persepsi publik terhadap institusi ini.

Kepolisian adalah institusi negara yang memiliki tugas utama menjaga keamanan, ketertiban, dan penegakan hukum dalam masyarakat. Fungsi polisi tidak hanya terbatas pada penindakan kejahatan, tetapi juga mencakup pencegahan, perlindungan, dan pelayanan kepada masyarakat. Kepolisian berperan sebagai pengawas dan pengendali sosial, memastikan norma, aturan, dan hukum dijalankan dengan baik, serta membangun kepercayaan publik melalui pendekatan yang profesional, transparan, dan akuntabel. Keberadaan kepolisian yang efektif sangat

penting untuk menciptakan rasa aman, mencegah tindak kriminal, serta menjaga stabilitas sosial dan ketertiban public (Anshar, 2020:8)

Konsep kepolisian pada setiap negara sangat dipengaruhi oleh nilai-nilai yang berkembang pada masyarakat negara tersebut. Hal inilah yang menyebabkan pembahasan aspek filosofis mengenai tugas dan fungsi kepolisan sebagai penegak hukum menjadi sangat penting, karena "hukum nasional" secara konsensus diakui oleh dunia Internasional, sehingga nilai-nilai yang melatar belakangi peran polisi sebagai penegak hukum menjadi sebab adanya pengakuan dunia Internasional apakah peran tersebut dilakukan secara benar atau tidak.

2.1.5 Algoritma *Naïve Bayes*

Naïve Bayes adalah salah satu algoritma klasifikasi berbasis probabilitas yang menggunakan *Teorema Bayes* untuk menghitung kemungkinan suatu data termasuk ke dalam kategori tertentu. Algoritma ini disebut "*naive*" karena mengasumsikan bahwa setiap fitur atau atribut dalam data bersifat independen satu sama lain, meskipun dalam kenyataannya fitur-fitur tersebut mungkin saling berhubungan. Sifat sederhana dan efisien dari *Naïve Bayes* membuatnya banyak digunakan dalam pengolahan bahasa alami, termasuk analisis teks dan analisis sentimen, karena mampu memproses data dalam jumlah besar dengan cepat.

Menurut (Darwis, 2021:3) Algoritma *Naive Bayes* berdasarkan teorema *Bayes* bahwa semua kegiatan memberikan sebuah kontribusi yang sama penting atau saling bebas pada pemilihan kelas tertentu. Salah satu metode klasifikasi untuk menentukan gambaran persepsi masyarakat di dalam *Text Mining* adalah metode *Naïve Bayes* yang sering disebut dengan *Naive Bayes Classifier*

Tujuan penggunaan algoritma *Naïve Bayes* adalah untuk mengklasifikasikan data secara otomatis berdasarkan probabilitas. Dalam konteks analisis sentimen, algoritma ini digunakan untuk menentukan apakah teks berita atau komentar masyarakat bersifat positif dan negatif. Dengan menggunakan probabilitas, *Naïve Bayes* memungkinkan peneliti memperoleh prediksi yang cepat dan konsisten, meskipun data yang dianalisis memiliki kompleksitas bahasa yang tinggi atau volume data yang besar.

Kelebihan *Naïve Bayes* terletak pada kesederhanaannya, efisiensi dalam komputasi, dan kemampuan bekerja dengan dataset besar. Namun, keterbatasannya muncul karena asumsi independensi fitur tidak selalu sesuai dengan karakter bahasa alami; misalnya kata-kata dalam sebuah kalimat sering saling terkait makna. Selain itu, performa algoritma ini bisa menurun jika data yang digunakan tidak seimbang atau *preprocessing* teks seperti *tokenisasi*, *stemming*, dan penghapusan *stopword* tidak dilakukan dengan tepat (Susana, 2022:6).

Dalam penelitian analisis sentimen berita kepolisian, *Naïve Bayes* digunakan sebagai metode utama untuk mengklasifikasikan opini publik. Algoritma ini memungkinkan peneliti mengukur sentimen masyarakat terhadap tindakan kepolisian secara otomatis dan kuantitatif, sehingga mempermudah pemetaan persepsi publik. Dengan penerapan *Naïve Bayes* di *platform* seperti *Google Colab*, peneliti dapat melakukan *preprocessing*, pelatihan model, dan evaluasi hasil secara interaktif, sekaligus mengatasi keterbatasan sumber daya komputasi lokal. Analisis ini kemudian dapat digunakan untuk mengidentifikasi isu atau respons masyarakat yang penting bagi perbaikan layanan dan transparansi kepolisian.

Rumus probabilitas Naive Bayes:

$$P\ (C|X) = \frac{P\ (X|C).P(C)}{P(X)} \qquad (2.1)$$

Keterangan:

a. *C* = kelas sentimen (Positif, Negatif)

b. *X* = fitur (kata/token dalam komentar)

c. *P(C| X)* = probabilitas komentar termasuk kelas *C*

d. *P(X| C)* = probabilitas kata muncul pada kelas *C*

e. *P(C)* = probabilitas prior kelas *C*

f. *P(X)* = probabilitas kemunculan kata secara keseluruhan

2.1.6 Konsep Pengolahan Data

Menurut Ghirid (2024:2) Pengolahan data adalah proses mengubah data mentah menjadi informasi yang dapat digunakan untuk pengambilan keputusan. Data yang diperoleh dari observasi, wawancara, dokumen, maupun sistem, harus diproses melalui tahapan terstruktur agar menghasilkan informasi yang valid dan dapat dianalisis. Proses pengolahan data dapat dibedakan menjadi tiga tahap utama, yaitu:

1) Input Data

   Data dikumpulkan dari berbagai sumber kemudian dilakukan pembersihan data (*cleaning*) untuk menghilangkan kesalahan, duplikasi, dan kekosongan data.

2) Proses Transformasi

   Data diseleksi dan diubah sesuai format kebutuhan penelitian, misalnya melalui proses klasifikasi, normalisasi, tokenisasi (pada NLP), atau konversi nilai.

3) Output Data

Hasil olahan data disajikan dalam bentuk tabel, grafik, atau nilai terstandarisasi sehingga siap dianalisis lebih lanjut.

Pengolahan data bertujuan memastikan bahwa informasi yang dihasilkan akurat, reliabel, dan dapat dibandingkan sesuai indikator penelitian.

Menurut (Urva et all, 2023:11) Perolehan data dari sebuah asset informasi dapat diperoleh dari beberapa sumber. Untaian data atau data set merupakan representasi informasi pasangan parameter dan nilai. Data set dewasa ini dapat berupa bentuk tabel, matrik, graph dan citra. Pada beberapa studi kasus pengolahan data, asset informasi dapat disajikan dalam pencatatan digital meliputi JSON, CSV, Basis Data dan format data lainya. Bentuk agregasi dokumen data, bergantung dari sumber perolehan data.

2.1.7 Konsep Pembobotan

Pembobotan adalah proses memberikan nilai atau tingkat kepentingan pada setiap variabel atau kriteria dalam penelitian. Bobot diperlukan karena tidak semua indikator memiliki peran yang sama dalam menentukan hasil akhir penelitian atau pengambilan keputusan.

Menurut (Wati, 2023:4) pembobotan kata menggunakan *Term Frequency-Inverse Document Frequency* (TF-IDF), pembobotan kata berfungsi untuk mengubah data berupa teks menjadi numerik. Bidang pembelajaran mesin atau Machine learningdan deep learningakan bekerja dengan baik dan maksimal jika data yang diolah berupa data numerik. Maka setiap penelitian pada bidang ini

khususnya analisis sentimen harus mengubah data berupa kumpulan kata-kata atau kalimat menjadi numerik.

Beberapa metode pembobotan yang umum digunakan antara lain:

Tabel 2.1 Pembobotan

Metode	Penjelasan Singkat
Manual/*Expert Judgment*	Bobot ditentukan berdasarkan wawancara atau penilaian pakar.
AHP (*Analytical Hierarchy Process*)	Bobot diperoleh dari perbandingan berpasangan antar kriteria.
Weighting in Machine Learning	Bobot dihitung otomatis berdasarkan kontribusi fitur pada model.
Profile Matching / Gap Weighting	Bobot diperoleh berdasarkan selisih nilai aktual dan nilai standar.

Pembobotan sangat penting untuk memastikan hasil analisis tidak bias dan mencerminkan kepentingan yang proporsional dari setiap indikator.

Tujuan dari TF-IDF adalah untuk mengidentifikasi kata-kata yang paling penting dalam suatu dokumen atau kumpulan dokumen. *Term Frequency* (TF) adalah nilai frekuensi kemunculan suatu kata dalam sebuah dokumen. Nilai TF dapat dihitung dengan menggunakan rumus

$$TF = \frac{(jumlah\ kemunculan\ kata\ dalam\ dokumen)}{(jumlah\ kata\ dalam\ dokumen)} \qquad (2.2)$$

Namun, nilai TF ini tidak memberikan informasi tentang pentingnya kata tersebut dalam dokumen. Kata-kata yang sering muncul dalam sebuah dokumen seperti kata hubung atau kata umum mungkin memiliki nilai TF yang tinggi, tetapi sebenarnya tidak memiliki makna yang penting dalam dokumen tersebut.Oleh karena itu, dibutuhkan teknik lain, yaitu *Inverse Document Frequency* (IDF) yang

memberikan bobot pada kata-kata yang muncul jarang di seluruh dokumen. Nilai IDF dapat dihitung dengan menggunakan rumus:

$$IDF = log\frac{N}{n} \tag{2.3}$$

Dimana N adalah jumlah dokumen dalam kumpulan dokumen dan n adalah jumlah dokumen yang mengandung kata tersebut. Setelah nilai TF dan IDF diperoleh, nilai TF-IDF dapat dihitung dengan mengalikan nilai TF dengan nilai IDF. Kata-kata yang memiliki nilai TF-IDF yang tinggi dianggap penting dan memberikan kontribusi yang lebih besar dalam menentukan topik dokumen atau kumpulan dokumen tersebut.

A. Normalisasi Data

Normalisasi merupakan proses mengubah data yang memiliki rentang nilai berbeda menjadi skala yang seragam sehingga dapat dibandingkan dan dilakukan perhitungan dengan adil. Normalisasi digunakan pada penelitian kuantitatif, *scoring* sistem pakar, metode *machine learning*, dan pengambilan keputusan.

Contoh persamaan normalisasi:

$$N = \frac{x - xmin}{xmax - xmin} \tag{2.4}$$

2.1.8 Konsep Analisis Hasil

Setelah pembobotan dan normalisasi dilakukan, langkah berikutnya adalah perhitungan nilai akhir berdasarkan formula perhitungan tertentu, misalnya perhitungan skor, ranking, atau klasifikasi.

Menurut (Wardani, 2021:2) Analisis akhir merupakan tahap penutup dalam proses pengolahan data yang bertujuan untuk menarik kesimpulan berdasarkan hasil perhitungan dan klasifikasi sentimen yang telah dilakukan sebelumnya.

Setelah proses *pre-processing* dan pembobotan menghasilkan kategori sentimen berupa positif dan negatif, sistem kemudian melakukan perhitungan komputasi untuk mengetahui kecenderungan sentimen secara keseluruhan.

Contoh umum rumus nilai *terweighted*:

$$\text{Score Akhir} = \sum (\text{bobot indikator} \times \text{nilai indikator}) \tag{2.5}$$

Analisis dilakukan untuk menafsirkan hasil perhitungan guna menjawab rumusan masalah penelitian. Analisis dapat berupa:

a. Analisis Kuantitatif: menggunakan perhitungan statistik atau algoritma numerik.
b. Analisis Kualitatif: berdasarkan interpretasi, pola, dan kategori data.

Jika terdapat beberapa komentar untuk satu objek, skor akhir dapat dihitung menggunakan rata-rata skor numerik:

$$Skor\ Akhir = \frac{\sum_{i=1}^{n} Wi}{n} \tag{2.6}$$

Keterangan:

a. Wi = skor numerik komentar ke-i
b. n = jumlah komentar

2.1.9 Flowchart

Flowchart adalah representasi grafis dari alur proses atau langkah-langkah yang sistematis dalam suatu kegiatan atau sistem. *Flowchart* menggunakan simbol-simbol standar, seperti kotak untuk aktivitas atau proses, belah ketupat untuk pengambilan keputusan, dan panah untuk menunjukkan arah alur, sehingga memudahkan pemahaman urutan kerja atau prosedur yang kompleks. Dalam

penelitian atau pengembangan sistem, *flowchart* berfungsi sebagai alat dokumentasi dan komunikasi visual yang membantu peneliti, pengembang, maupun pembaca memahami bagaimana data atau informasi diproses dari awal hingga akhir (Rosali, 2019:5). Dengan menggunakan *flowchart*, proses analisis, seperti analisis sentimen menggunakan algoritma *Naïve Bayes*, dapat digambarkan secara sistematis, mulai dari pengumpulan data, *preprocessing*, pelatihan model, hingga evaluasi hasil, sehingga mempermudah pemahaman alur kerja dan identifikasi potensi kendala dalam setiap tahap.

Flowchart analisis sentimen berita kepolisian dimulai dari tahap pengumpulan data, di mana berita *online* atau komentar masyarakat dikumpulkan sebagai dataset. Data ini kemudian melewati proses *preprocessing*, termasuk *case folding* untuk menyamakan huruf, tokenisasi untuk memecah teks menjadi kata, penghapusan *stopword*, dan stemming untuk mendapatkan bentuk dasar kata. Tahap ini penting untuk memastikan data siap diolah oleh algoritma dan meminimalkan gangguan dari kata-kata yang tidak relevan.

Setelah *preprocessing*, data diberi label sentimen seperti positif dan negatif. Dataset kemudian dibagi menjadi data training dan data *testing*. Pada tahap pelatihan model, algoritma *Naïve Bayes* menghitung probabilitas kemunculan kata terhadap masing-masing label sentimen dan membangun model klasifikasi..

2.2 Penelitian Yang Relevan

Tabel 2.1. Penelitian yang relevan

No	Penulis / Tahun	Topik Penelitian	Metode	Pembahasan	Hasil
1	Nurpandi et all (2024)	Analisis Sentimen terhadap Kinerja Kepolisian Negara Republik Indonesia dengan data Twitter	*Metode: Multinomial Naïve Bayes + LSTM + Lexicon-Based. Data: ~12.222 tweet setelah pre-processing.*	Membahas persepsi publik terhadap kinerja kepolisian melalui tweet yang mengangkat isu-kasus seperti penembakan Brigadir J, tragedi Kanjuruhan, pengamanan KTT G20 Bali, Natal & Tahun Baru 2023. Klasifikasi sentimen dibagi ke kategori positif & negatif.	Ditemukan bahwa ~53% tweet memiliki sentimen negatif, ~47% positif. Metode Multinomial NB menghasilkan akurasi ~78%, precision ~78.5%, recall ~77%, f1-score ~77%. Sedangkan LSTM mencapai ~87% untuk masing-masing metrik.
2	Rony Oktavaizar (2024)	Penerapan Naïve Bayes untuk Klasifikasi Sentimen Opini Netizen tentang Kinerja Polisi RI	*Metode: Naïve Bayes; Pembobotan Bag of Words; data komentar Facebook dengan kata kunci "percuma lapor polisi"*	Fokus pada opini netizen melalui komentar Facebook, klasifikasi tiga kelas (positif, negatif, netral).	Dihasilkan akurasi ~82% dan rata-rata F1-score ~74%, menunjukkan sistem klasifikasi bekerja dengan baik
3	Muhammad Fauzan Nurfadhillah (2025)	Analisis Sentimen Kinerja Kepolisian di Media YouTube menggunakan Naïve Bayes dan SVM	*Metode: Naïve Bayes + Support Vector Machine; Data komentar YouTube; tahapan preprocessing (normalisasi teks, tokenization, stopword removal, stemming).*	Membahas opini publik yang disalurkan melalui komentar video YouTube terkait kinerja kepolisian, mengklasifikasikan ke dalam kategori sentimen.	Hasil klasifikasi menunjukkan efektivitas metode; namun nilai spesifik akurasi/precision/recall dalam abstrak tidak disebutkan secara lengkap dalam sumber yang saya temukan

No	**Penulis / Tahun**	**Topik Penelitian**	**Metode**	**Pembahasan**	**Hasil**
4	Ilka Zufria, Aidil Halim Lubis, Siti Septia Febiyaula (2024)	Analisis Sentimen Kepercayaan Masyarakat terhadap Kepolisian RI menggunakan SVM	*Metode: Support Vector Machine*	Pembahasan berfokus pada bagaimana kepercayaan masyarakat terhadap institusi kepolisian dipengaruhi oleh berita/isu melalui media internet, menggunakan klasifikasi sentimen untuk mengukur persepsi	Hasil menunjukkan akurasi ~83.95%, precision ~85.71%, recall ~87.50%, f1-score ~86.60% dengan metode SVM
5	Solihin & Prasetyo, E. (2023)	Komunikasi Kesehatan Dalam Isu Kebijakan Publik: Analisis Tayangan Youtube'Polemik Aturan BPJS 2025	*Naïve Bayes Classifier*	Penelitian ini membahas analisis sentimen opini masyarakat terhadap institusi kepolisian berdasarkan komentar di media sosial. Menggunakan algoritma Naïve Bayes untuk mengetahui kecenderungan sentimen positif dan negatif.	Hasil penelitian menunjukkan bahwa algoritma Naïve Bayes mampu mengklasifikasikan sentimen dengan baik, dengan tingkat akurasi sebesar ±86%, sehingga metode ini dinilai efektif untuk analisis sentimen opini publik terhadap institusi kepolisian

Berdasarkan penelitian-penelitian terdahulu yang berfokus pada analisis sentimen menggunakan berbagai metode klasifikasi dan dataset berbeda, penelitian ini mengusulkan pendekatan yang lebih terarah pada konteks sosial tertentu, yaitu sentimen masyarakat terhadap Polda Kalteng melalui komentar di media sosial atau portal berita *online*.

BAB III

METODOLOGI PENELITIAN

3.1 Tinjauan Umum

Penelitian ini berfokus pada penerapan metode analisis sentimen menggunakan algoritma *Naïve Bayes Classifier* terhadap berita kepolisian berbasis *online*. Metode analisis sentimen dipilih karena mampu mengolah data teks dalam jumlah besar dan mengekstrak informasi berupa kecenderungan opini (positif dan negatif). Dalam penelitian ini, tahapan analisis meliputi pengumpulan data berita, *preprocessing* teks, pembangkitan fitur, penerapan algoritma *Naïve Bayes*, hingga evaluasi performa model. Proses ini menjadi landasan dalam membangun sistem cerdas yang dapat membantu pihak terkait dalam memahami persepsi publik secara lebih akurat dan terukur.

Penelitian ini juga menempatkan lokasi kajian secara khusus pada berita-berita *online* yang berkaitan dengan Polda Kalimantan Tengah (Polda Kalteng). Pemilihan lokasi ini didasarkan pada banyaknya pemberitaan digital mengenai kegiatan kepolisian di wilayah Kalimantan Tengah, baik dari media nasional maupun lokal. Berita-berita tersebut memberikan gambaran yang jelas mengenai bagaimana kinerja, kebijakan, serta dinamika kepolisian diliput oleh media, sehingga cocok digunakan sebagai sampel dalam penelitian.

3.2 Jenis penelitian

Jenis penelitian ini adalah penelitian deskriptif kuantitatif yang menggunakan pendekatan sistem cerdas (*machine learning*) untuk menganalisis sentimen masyarakat terhadap pemberitaan kepolisian.

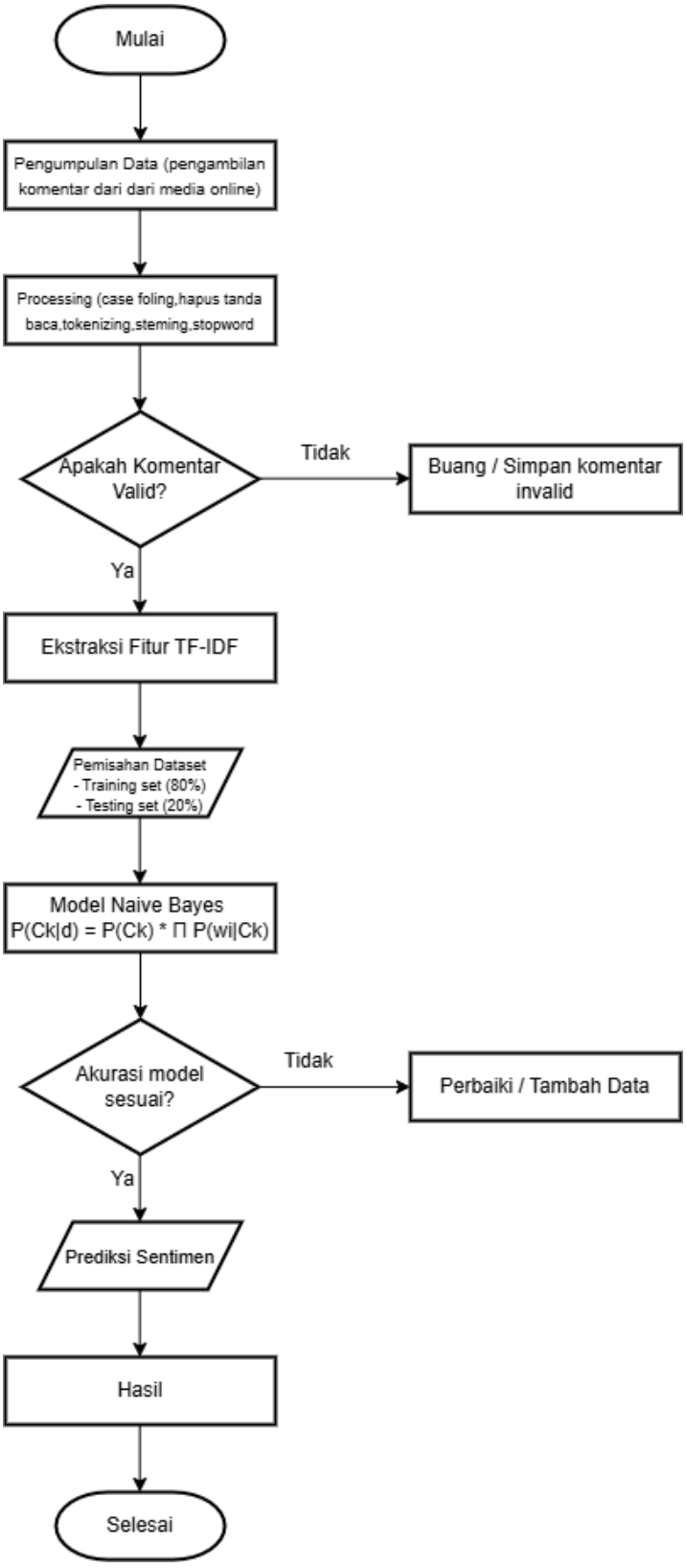
Penelitian ini bersifat deskriptif karena bertujuan untuk menganalisis persepsi masyarakat melalui komentar-komentar di dunia maya, khususnya terkait Polda Kalimantan Tengah. Penelitian ini juga bersifat kuantitatif, karena menggunakan metode *Naïve Bayes* untuk mengklasifikasikan data komentar menjadi kategori sentimen (positif dan negatif) secara numerik.

Penelitian ini menggunakan dataset kecil (sekitar 50 komentar), sehingga lebih fokus pada analisis tren persepsi masyarakat, tanpa perlu membangun sistem besar atau mengolah dataset ribuan data.

3.3 Desain Penelitian

Desain penelitian ini mengikuti alur sederhana dan efisien untuk memudahkan pengerjaan:

Proses pada penelitian ini menggambarkan alur analisis sentimen menggunakan metode *Naïve Bayes* secara sistematis, dimulai dari tahap pengumpulan data komentar hingga menghasilkan visualisasi sentimen masyarakat. Setiap komentar yang diperoleh melewati proses *preprocessing* untuk memastikan teks bersih dan layak diproses. Pada tahap ini terdapat percabangan keputusan untuk memeriksa apakah komentar valid jika tidak valid, komentar akan dibuang.



Gambar 3.2 Desain *Flowchart Naïve Bayes*

Tahapan desain penelitian tersebut dijabarkan sebagai berikut :

A. Pengumpulan Data

Data penelitian yang digunakan berupa komentar masyarakat yang diambil dari akun Instagram resmi Polda Kalteng pada beberapa postingan terkait berita dan informasi pelayanan kepolisian. Jumlah komentar yang dikumpulkan sebanyak kurang lebih 50 data komentar. Proses pengambilan data dilakukan secara manual melalui metode dokumentasi, yaitu dengan menyalin komentar secara langsung dari kolom komentar Instagram ke dalam *software Microsoft Excel*. Selanjutnya, setiap komentar yang terkumpul diberi label sentimen berupa kategori positif dan *negative* sebagai dasar untuk proses analisis lanjutan menggunakan metode klasifikasi sentimen.

B. *Preprocessing* Teks

Tahapan pemrosesan teks (*Natural Language Processing*/NLP) dalam penelitian ini dilakukan menggunakan model IndoBERT Lite, yang mencakup beberapa proses utama, yaitu *case folding*, *tokenizing*, *stopword removal*, dan *stemming*. Tahap *case folding* dilakukan dengan mengubah seluruh huruf dalam teks komentar menjadi bentuk *lowercase* agar konsistensi data terjaga dan tidak terjadi perbedaan makna akibat variasi huruf besar dan kecil. Selanjutnya, dilakukan proses *tokenizing*, yaitu memecah kalimat komentar menjadi potongan kata-kata tunggal sehingga memudahkan analisis. Tahap berikutnya adalah *stopword removal*, yaitu menghapus kata-kata yang dianggap tidak memiliki kontribusi terhadap makna sentimen, seperti “yang”, “dan”, “atau”, serta kata umum lainnya. Tahap terakhir yaitu stemming, yang

bertujuan mengembalikan setiap kata ke bentuk dasar atau akar kata sehingga kata dengan turunan yang berbeda dapat dikenali sebagai satu makna dasar. Seluruh tahapan ini diperlukan untuk menghasilkan data teks yang bersih dan siap digunakan dalam proses analisis sentimen.

C. Ekstraksi Fitur

Tahap selanjutnya ekstraksi fitur dilakukan untuk mengubah data teks hasil *pre-processing* menjadi bentuk numerik yang dapat diproses oleh algoritma *Naïve Bayes*. Metode yang digunakan dalam penelitian ini adalah TF-IDF (*Term Frequency – Inverse Document Frequency*).

Pada tahap *Term Frequency* (TF), dihitung jumlah kemunculan setiap kata pada satu komentar, dengan rumus TF = jumlah kemunculan kata / total kata dalam dokumen; semakin sering sebuah kata muncul, semakin besar nilai TF-nya. Selanjutnya, tahap *Inverse Document Frequency* (IDF) digunakan untuk mengukur seberapa penting sebuah kata dalam keseluruhan dokumen, dengan rumus

IDF = log (total dokumen / jumlah dokumen yang mengandung kata) Nilai TF-IDF diperoleh dengan mengalikan nilai TF dan IDF untuk setiap kata, sehingga menghasilkan representasi numerik yang mempertimbangkan baik frekuensi kata dalam satu komentar maupun kepentingannya dalam seluruh dataset. Hasil akhir dari tahap ini adalah setiap komentar diubah menjadi vektor numerik berdasarkan nilai TF-IDF kata-kata yang muncul, sehingga siap digunakan dalam proses klasifikasi sentimen menggunakan *Naïve Bayes*.

Contoh hasil Ekstraksi:

Tabel 3.1 Tabel Contoh Ekstraksi

No	Komentar Asli	Hasil *Pre-processing*	Label Sentimen	Kata Kunci	TF	IDF	TF-IDF
1	*"Terima kasih atas pelayanan cepatnya pak polisi."*	terima kasih layanan cepat polisi	Positif	layanan	1/5 = 0.20	0.48	0.096
2	*"Pelayanan sangat lambat, laporan saya tidak diproses dengan baik."*	layanan lambat lapor proses baik	Negatif	lambat	1/6 = 0.16	0.70	0.112

D. Pemisahan Dataset

Tahap pembagian dataset dimulai dengan menentukan porsi data yang akan digunakan untuk pelatihan dan pengujian model. Dalam penelitian ini, dataset dibagi dengan proporsi 80% untuk training set dan 20% untuk *testing* set, pembagian yang umum digunakan pada model klasifikasi karena memberikan jumlah data yang cukup bagi model untuk belajar sekaligus menyediakan data uji yang memadai untuk mengevaluasi performa. Selanjutnya, dataset dilakukan pengacakan (*shuffling*) agar distribusi sentimen tidak mengelompok, misalnya agar komentar *negative* dan positif tidak semuanya berada di awal.

Pengacakan ini penting untuk mencegah bias model terhadap pola penempatan data. Setelah data diacak, dataset dibagi sesuai persentase yang telah ditentukan; misalnya, dari 50 komentar, 40 komentar digunakan sebagai data training dan 10 komentar sebagai data *testing*. Data *training* digunakan

untuk membangun model *Naïve Bayes*, sedangkan data *testing* digunakan untuk menghitung akurasi model. Proses pembagian ini diterapkan pada dataset yang telah diekstraksi fiturnya menggunakan TF-IDF dan dilengkapi dengan label sentimen masing-masing komentar.

E. Pelatihan Model *Naïve Bayes*

Model *Naïve Bayes* merupakan algoritma klasifikasi berbasis probabilitas yang banyak digunakan untuk analisis sentimen, karena kemampuannya dalam menangani data teks dengan jumlah fitur yang besar secara efisien. Algoritma ini bekerja dengan prinsip Teorema Bayes, yang menghitung probabilitas suatu komentar termasuk dalam kelas tertentu (Positif dan Negatif) berdasarkan kemunculan kata-kata yang ada dalam komentar tersebut.

Pada penelitian ini, data komentar yang telah melalui tahap *pre-processing* dan ekstraksi fitur menggunakan TF-IDF dimasukkan ke dalam model *Naïve Bayes*. Setiap komentar kemudian diberikan prediksi kelas sentimen dengan menghitung probabilitas.

F. Visualisasi Hasil

Tahap visualisasi hasil dilakukan untuk menampilkan distribusi sentimen yang telah diprediksi oleh model *Naïve Bayes* dalam bentuk grafik. Visualisasi ini bertujuan untuk mempermudah pemahaman mengenai kecenderungan opini masyarakat berdasarkan jumlah komentar dalam masing-masing kategori sentimen, yaitu positif dan negatif.

3.4 Instrumen Penelitian

Instrumen penelitian adalah alat yang digunakan untuk mengumpulkan dan memperoleh data yang diperlukan dalam proses penelitian. Pada penelitian berjudul "Peran Analisis Sentimen Berbasis *Naïve Bayes* Pada Berita Kepolisian Berbasis Online", instrumen penelitian yang digunakan adalah dokumen komentar pengguna media sosial Instagram. Instrumen ini digunakan untuk mengumpulkan data tekstual yang kemudian diproses dan dianalisis menggunakan metode NLP (*Natural Language Processing*).

A. Komentar pada akun Instagram resmi Polda Kalteng yang diambil dari postingan terkait informasi pelayanan dan kegiatan kepolisian. Komentar ini digunakan sebagai sumber utama data sentimen masyarakat.

   Contoh:

Tabel 3.1 Contoh Komentar

Komentar	Label
Polisi cepat menangani laporan warga	Positif
Pelayanan SIM lambat banget	Negatif

B. *Software Microsoft Excel* Instrumen ini digunakan untuk menyalin dan menyimpan komentar yang telah dikumpulkan secara manual serta memberikan label kategori sentimen (Positif dan Negatif) sebelum dilakukan pemrosesan menggunakan *Google Colab*.

C. Setiap komentar diberikan label kategori sentimen berdasarkan isi komentar, yaitu: positif dan negatif, sebagai acuan dalam proses pelatihan dan pengujian model klasifikasi.

D. Teknik pengumpulan data lainnya yaitu berupa dari pustaka. Dimana studi pustaka merupakan suatu kegiatan penelusuran dan kegiatan penelitian literatur. Kegiatan ini dilakukan untuk mengumpulkan data yang berasal dari jurnal untuk mendukung penelitian.

3.5 Analisis Kebutuhan

3.5.1 Analisis Data

Data yang digunakan adalah berita kepolisian berbasis *online*, yang dikumpulkan dari beberapa portal media digital. Data dikumpulkan dalam bentuk teks dan disimpan ke dalam basis data untuk diproses lebih lanjut. Data dikategorikan menjadi tiga label sentimen:

a. Positif
b. Negatif

Penelitian ini dilakukan secara terstruktur untuk memperoleh komentar-komentar masyarakat yang relevan terhadap berita kepolisian, khususnya yang berkaitan dengan aktivitas Polda Kalteng di media *online*. Teknik pengumpulan data dilakukan dengan beberapa tahapan agar data yang dikumpulkan valid, representatif, dan dapat dianalisis menggunakan metode *Naïve Bayes*.

a. Peneliti mengumpulkan data komentar dari media sosial dan portal berita *online* yang menjadi sumber informasi kepolisian, seperti Facebook, Instagram, dan YouTube resmi Polda Kalteng, serta beberapa website berita lokal. Komentar yang diambil dipilih berdasarkan keterkaitannya dengan topik kepolisian, keamanan, pelayanan publik, dan isu kriminalitas. Pengumpulan ini dilakukan secara manual maupun semi-otomatis.

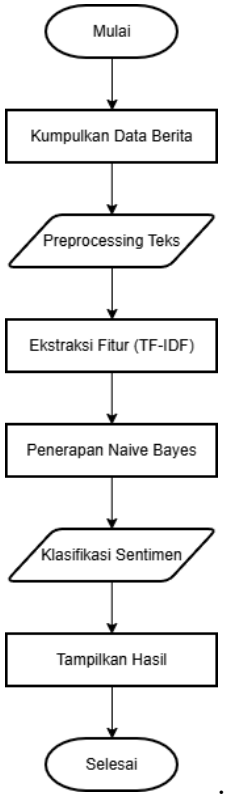
b. Setiap komentar yang berhasil dikumpulkan dicatat ke dalam *tabel database* yang telah disiapkan, meliputi informasi sumber, waktu komentar, dan isi komentar mentah. Data yang tidak relevan, seperti iklan, spam, komentar kosong, atau komentar dengan bahasa tidak jelas, disaring untuk memastikan kualitas data tetap terjaga..
c. Peneliti melakukan pembersihan data (*data cleaning*) sebagai bagian dari proses pengumpulan data lanjutan. Komentar yang memiliki kesalahan penulisan ekstrem, mengandung *emoticon* tanpa teks, ataupun tidak memiliki konteks akan dihapus. Komentar duplikat juga dieliminasi untuk mencegah bias dalam analisis sentimen.
d. Data yang terkumpul kemudian diklasifikasikan berdasarkan kategori sentimen awal (tanpa model), yaitu positif dan *negative* melalui teknik pengkodean manual apabila diperlukan sebagai *dataset training* pendukung. Pengkodean ini dilakukan untuk membantu model *Naïve Bayes* memperoleh contoh data awal yang jelas dan konsisten.

   Seluruh data yang sudah terkumpul dan dibersihkan disimpan dalam basis data yang telah dirancang pada bagian sebelumnya. Penyimpanan ini bertujuan agar data mudah diproses pada tahap *preprocessing*, ekstraksi fitur, pelatihan model, hingga klasifikasi akhir. Dengan prosedur pengumpulan data yang sistematis, penelitian ini mampu menghasilkan data komentar yang terstruktur dan siap dianalisis menggunakan metode analisis sentiment.

3.5.2 Analisis Proses

Flowchart proses penelitian ini menggambarkan alur kerja dari pengumpulan hingga analisis komentar masyarakat terhadap Polda Kalteng. Proses dimulai dengan pengumpulan data komentar dari media sosial atau portal berita *online*, kemudian dilanjutkan dengan *preprocessing* teks, yang mencakup *case folding*, penghapusan tanda baca, *tokenizing, stopword removal*, dan *stemming*.

Setelah teks dibersihkan, dilakukan ekstraksi fitur menggunakan TF-IDF, sehingga data teks dapat diubah menjadi format numerik yang siap dianalisis. Selanjutnya, data dibagi menjadi *training set* dan *testing set*, kemudian dilatih menggunakan model *Naïve Bayes* untuk klasifikasi sentimen menjadi positif atau negatif.



Gambar 3.1 Analisis Proses

Berdasarkan Gambar 3.1 Analisis Proses diatas proses penelitian dimulai dari tahap pengumpulan data, yaitu dengan mengambil komentar masyarakat pada postingan berita dan informasi di akun Instagram resmi Polda Kalteng. Proses ini dilakukan secara manual melalui dokumentasi dan penyalinan komentar ke dalam *Microsoft Excel*. Data komentar yang diperoleh selanjutnya diberikan label sentimen awal berupa positif dan *negative* sebagai dasar proses analisis berikutnya.

Tahap berikutnya adalah *pre-processing text*, yaitu proses pembersihan dan standarisasi data teks agar siap untuk dianalisis. Pada tahap ini dilakukan beberapa proses seperti *case folding* , *tokenizing*, *stopword removal*, serta stemming. Tujuan dari tahap ini adalah menghilangkan noise dan meningkatkan kualitas data sebelum masuk ke tahap perhitungan numerik.

Selanjutnya dilakukan ekstraksi fitur menggunakan metode TF-IDF (*Term Frequency – Inverse Document Frequency*). Tahap ini mengubah teks menjadi bentuk numerik berdasarkan bobot kemunculan kata dalam satu dokumen dan tingkat kepentingan kata dalam keseluruhan dataset. Nilai TF-IDF inilah yang nantinya menjadi representasi vektor numerik dari setiap komentar, dan menjadi input utama untuk proses klasifikasi.

Tahap berikutnya adalah penerapan algoritma *Naïve Bayes*, yaitu metode klasifikasi berbasis probabilitas yang menghitung kemungkinan suatu komentar termasuk ke dalam kategori sentimen tertentu. Berdasarkan vektor TF-IDF, algoritma *Naïve Bayes* menghitung probabilitas posterior untuk kelas sentimen positif dan negatif menggunakan Teorema Bayes, kemudian memilih kelas dengan probabilitas terbesar sebagai hasil prediksi.

Setelah pemodelan, dilakukan tahap klasifikasi sentimen, yaitu menentukan kategori sentimen setiap komentar berdasarkan hasil perhitungan *Naïve Bayes*. Komentar diberi label sentimen sesuai nilai probabilitas tertinggi yang diperoleh dari proses klasifikasi. Pada tahap ini dilakukan pula evaluasi model dengan menggunakan data *testing* untuk mengukur akurasi dan ketepatan prediksi.

3.5.3 Analisis Kelemahan

Sebelum merancang penelitian, perlu diperhatikan beberapa kendala yang mungkin muncul dalam analisis sentimen. Dari sisi sistem, keterbatasan dataset dapat menyebabkan model bias jika data tidak seimbang. Kompleksitas Bahasa Indonesia, termasuk idiom, sarkasme, dan konteks lokal, menyulitkan model untuk memahami makna secara akurat. Selain itu, algoritma *Naïve Bayes* yang berbasis probabilitas mengasumsikan fitur saling independen, sehingga tidak selalu sesuai dengan karakter bahasa alami. Sementara kesalahan dalam *preprocessing*, seperti *stemming* atau *stopword removal*, dapat menurunkan akurasi.

Dari perspektif metode *PIECES*, analisis ini juga menghadapi kelemahan: dari sisi *Performance*, model mungkin kurang cepat atau akurat; pada *Information*, informasi yang diperoleh bisa tidak lengkap atau bias; *Economics* menunjukkan keterbatasan sumber daya untuk pengumpulan dataset besar; *Control* terkait validasi data dan konsistensi preprocessing; *Efficiency* dipengaruhi oleh kompleksitas algoritma dan volume data, sedangkan *Service* berkaitan dengan kemampuan sistem memberikan hasil analisis yang dapat diandalkan bagi pengguna. Dengan memahami kelemahan ini, penelitian dapat merancang strategi mitigasi sebelum pembangunan sistem dilakukan.

BAB IV

PEMBAHASAN

4.1 Hasil

Pada tahap ini disajikan hasil penelitian yang diperoleh setelah melalui proses pengumpulan data, *preprocessing*, ekstraksi fitur, pelatihan model, pengujian, dan evaluasi. Data yang digunakan dalam penelitian ini adalah 50 komentar publik yang diambil dari akun Instagram resmi Polda Kalteng, terutama pada unggahan terkait informasi pelayanan, penegakan hukum, kegiatan sosial, dan berita kriminal. Data tersebut telah diberi label secara manual menjadi sentimen positif dan sentimen negatif.

Ada beberapa tahapan proses dalam data mining. Diagram dibawah ini menggambarkan beberapa tahapan/proses yang berlangsung dalam data mining. Fase awal dimulai dari data sumber dan berakhir dengan adanya informasi yang dihasilkan yaitu:

Langkah-langkah dalam melakukan data mining adalah sebagai berikut :

a. Data *cleaning* yaitu untuk menghilangkan *noise* data yang tidak konsisten.
b. Data *integration* (di mana sumber data yang terpecah dapat disatukan).
c. Data *selection* (di mana data yang relevan dengan tugas analisis dikembalikan ke dalam *database*).
d. Data *transformation* (di mana data berubah atau bersatu menjadi bentuk yang tepat untuk ringkasan performa atau operasi agresi).

4.1.1. Hasil Processing Teks

Pada tahap *preprocessing* teks, seluruh komentar yang telah dikumpulkan dari Instagram diproses melalui beberapa langkah pembersihan agar dapat digunakan sebagai input model analisis sentimen. Tahap pertama yang dilakukan *adalah case folding*, yaitu mengubah seluruh huruf dalam komentar menjadi huruf kecil untuk menyamakan format teks dan menghindari perbedaan makna akibat kapitalisasi. Setelah itu dilakukan *tokenizing*, yaitu memecah setiap komentar menjadi potongan kata (token) sehingga dapat dianalisis secara lebih detail. Langkah selanjutnya adalah *stopword removal*, kata-kata umum seperti "yang", "dan", "di", dan kata tidak bermakna lainnya dihilangkan karena tidak memberikan kontribusi signifikan pada penentuan sentimen. Tahap terakhir adalah stemming, yakni mengubah setiap kata menjadi bentuk dasarnya, misalnya "pelayanannya" menjadi "layanan" atau "kerjanya" menjadi "kerja", sehingga model hanya mempelajari akar kata yang memiliki makna paling relevan. Melalui proses *preprocessing* ini, komentar yang semula tidak terstruktur berubah menjadi teks bersih dan sederhana, sehingga siap digunakan pada tahap ekstraksi fitur menggunakan TF-IDF.

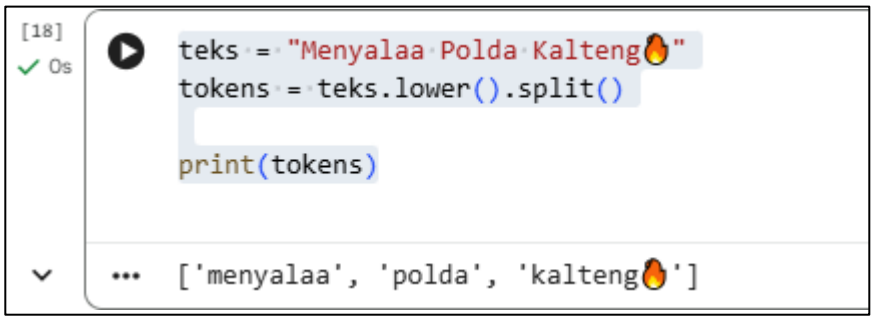
Sebelum melakukan pengolahan data peneliti terlebih dahulu merangkum data agar bisa dilakukan dengan hasil maksimal dengan contoh 5 sampel data sebagai berikut.

Tabel 4.1 Dataset Sampel

No	Username	Id Komen	Komentar	Hasil
1	kyyy_990	18184005499355800	Menyalaa Polda Kalteng 🔥	Menyala polda kalteng
2	dimazsyh	18079766504170600	sehat n suksess selalu jendral...👏 🔥	sehat sukses selalu jendral
3	cymiens_462	18047657978457600	Luar biasa...semangat pak Polisi, jadikan ladang ibadah 🔥 ❤️	luar biasa semangat polisi jadikan ladang ibadah
4	jayyidrabbani	18076531739231900	JAYAA SELALUU POLDA KALTENG 🙌🙌🙌🔥🔥🔥👍🏻👍🏻 👍🏻AAAA🌟🌟🌟 🌟🌟	jayaa selaluu polda kalteng
5	never_knowlann	18083458717862000	sehat n suksess selalu jendral...👏	sehat suksess selalu jendral
6	shhdimm_	18544862977022700	sehat n suksess selalu jendral...👏	sehat suksess selalu jendral
7	riefky37	18098858629692700	jayajayaa	Jayajaya
8	diditfc	18081106073119200	Polda kalteng jaya 👐 🔥 🔥	polda kalteng jaya
9	akbrchyos	18120961918536400	mantap polri jaya 🔥	mantap polri jaya
10	junasptrra	18063912599426300	Jaya terus Polda kaltengA	jaya terus polda kalteng
11	azkayunadaa	18413988055191600	menyalaa polrii 🔥 🔥	menyalaa polrii
12	rnooaazz_	18007288238660100	mantap polri ku 🔥 🔥	menyalaa polrii ku
13	dcyoch_19	17965428984004500	Bravo polri untuk Indonesia ❤️❤️❤️🙌🙌🙌🙌	bravo polri untuk Indonesia
14	riefky37	17912582736249000	@suudhabibiii____umaa jabul nah 🔥	umaa jabul nah
15	romansutaa	17909244768128400	Menyala 🔥 🔥	menyala
16	ddobboys	17870532741398700	Menyala polri 🔥 🔥 🔥	menyala polri
17	dimzdyy_	18066499481568900	@riefky37 berkah 🔥	berkah 🔥
18	wiiirantoo	18078520496132200	menyala jendral 👐	menyala jendral
19	aldiwinaata	17886166353281900	niceee kinggg 🔥 🔥 🔥	niceee kinggg
20	ydaaa_541	18166727749375800	Menyala terus Polda Kalteng 🔥 🔥 🔥	menyala terus Polda Kalteng
21	kikirizqi.22	17933277486109900	Jaya polrii 🔥 🔥	jaya polrii
22	edoorama_	17989426856904900	luar biasa 🔥	luar biasa
23	chocoyou.lite	18184558495349500	Tetap semangat selalu polri 🔥	tetap semangat selalu polri

24	iketutrama18	18539422489057100	Polda kalteng jayaaa 🔥	polda kalteng jayaaa
25	gabriellbatistuta_ _	17913413262251500	Menyala Selalu Polda Kalteng 🔥 🔥 🔥 🔥	menyala Selalu polda kalteng
26	awanmn	18553572490057100	bravoo polri 🔥 🔥	bravoo polri
27	dimzdyy_	17861403288542700	polri untuk masyarakat 🔥	polri untuk masyarakat
28	arffthrrmn	17906154708147200	Jaya selalu polda kalteng 🔥	jaya selalu polda kalteng
29	whyuariii	18199918531323500	Menyala 🔥 🔥	menyala
30	wyan_29	18112775083581500	sukses selalu polda kalteng 🔥 🔥	sukses selalu polda kalteng
31	mhdramlidarwis	17886164379281900	Masih kurang lagi menindak	masih kurang lagi menindak
32	ermu7798	17898947772342700	Kalteng Respek..Bravo Polda Kalteng... AA 🔥	kalteng respek..bravo polda kalteng
33	mirsydrdh	18057682568647500	Terima kasih POLRI slalu mnjadi penolong masyarakat ❤️	terima kasih polri slalu mnjadi penolong masyarakat
34	dimzdyy_	18188535721340200	sehat n suksess selalu jendral...🙌	sehat n suksess selalu jendral
35	ibraaajaaa_	18106152772629500	Menyala polri 🔥 🔥	menyala polri
36	rh.jmll	18039950897711900	Jaya jaya jaya 🔥	jaya jaya jaya
37	gueadit___	18090237815497200	mantap polda kaltengggg 🔥 🔥 🔥	mantap polda kaltengggg
38	jo_za117	17970768473986900	Polri jaya 🔥	polri jaya
39	rrwinnn_j	18062354840443400	Sukses polrii 🔥 🔥 🙌	sukses polrii
40	marlon.joe.90	18083960267091200	Menyalaa 🔥	menyalaa
41	c_ronaldelo	18085124717083600	BRAVO POLDA KALTENG 🔥	bravo polda kalteng
42	yuddee___	18165343819387700	menyala 🔥	menyala
43	junior_sihotang	18544893307044100	Menyala Polda kalteng ❤️ ❤️ 🔥 🔥	menyala polda kalteng
44	made_ken_diva	18064255226425200	Menyala	menyala
45	and.sgnt1990	18073602584253600	Gas terus polri 🔥 🔥 🔥	gas terus polri
46	muhmdammar	18075148751238300	Polri semakin di depan 🔥	polri semakin di depan
47	ahmad_okt_	17867729280499100	Polri jaya 🔥 🔥	polri jaya
48	dywaa80_	18033448949738500	Menyala 🔥 🔥	menyala
49	suudhabibiii____	18052828232652000	Menyala polri 🔥	menyala polri
50	rajarambeee_	18100758238745000	Terlalu lambat	Terlalu lambat

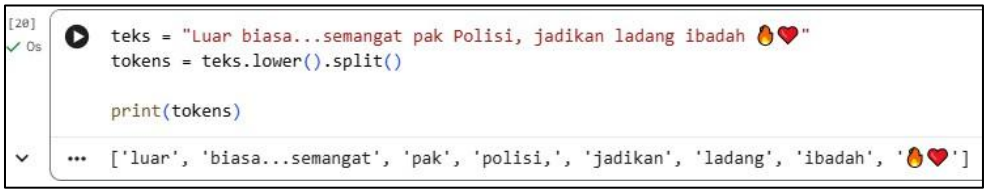
Berikut hasil dari *preprocessing text* yang sudah dilakukan peneliti melalui *google colab.*



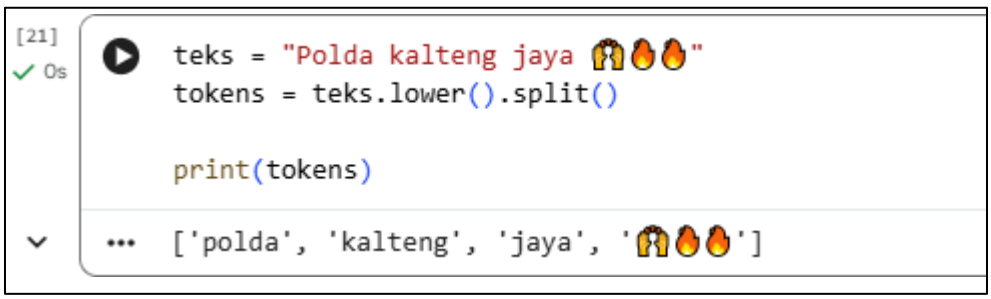
Gambar 4.1 *Stopword Removal* 1



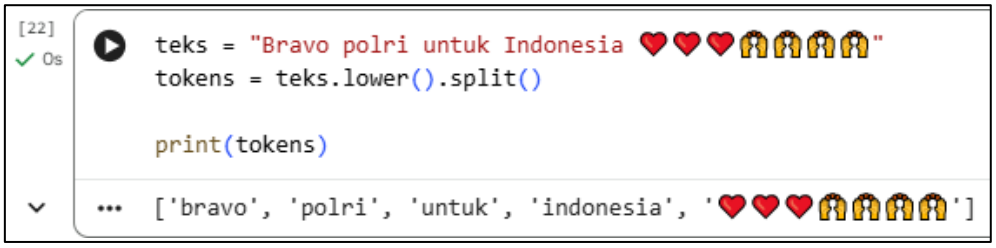
Gambar 4.2 *Stopword Removal* 2



Gambar 4.3 *Stopword Removal* 3



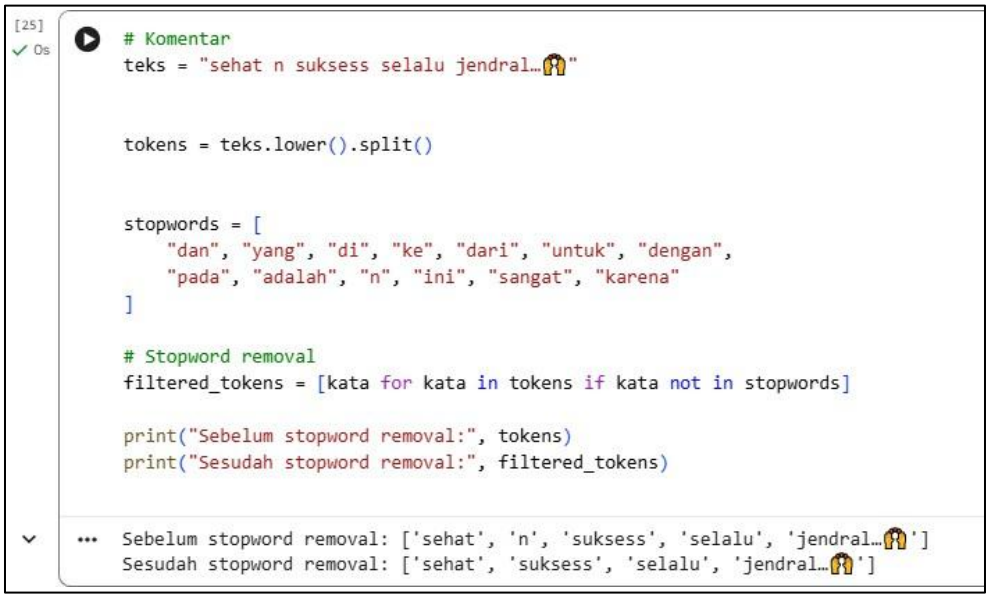
Gambar 4.4 *Stopword Removal* 4



Gambar 4.5 *Stopword Removal* 5



Gambar 4.6 *Stopword Removal* 6



Gambar 4.7 *Stopword Removal* 7



Gambar 4.8 *Stopword Removal* 8

Gambar 4.2 sampai 4.8 tersebut menampilkan hasil *preprocessing text* yang digunakan dalam penelitian, yang terdiri dari dua kolom utama yaitu komentar dan label. Kolom komentar berisi teks komentar masyarakat yang diambil dari media sosial Instagram Polda Kalteng, sedangkan kolom label menunjukkan hasil pelabelan sentimen secara manual dengan kategori positif. Contoh komentar yang ditampilkan mayoritas mengandung unsur apresiasi, dukungan, dan pujian terhadap kinerja kepolisian, seperti penggunaan kata “salut”, “sukses”, “luar biasa”, dan “pengamanan maksimal”. Dataset ini selanjutnya digunakan sebagai data awal sebelum melalui tahap *preprocessing*, ekstraksi fitur TF-IDF, serta pelatihan dan pengujian model *Naïve Bayes*.

4.1.2. Hasil Ekstraksi Fitur TF-IDF

Setelah melalui tahap *preprocessing* teks yang meliputi *case folding*, *tokenizing*, *stopword removal*, dan *stemming*, data komentar selanjutnya diproses

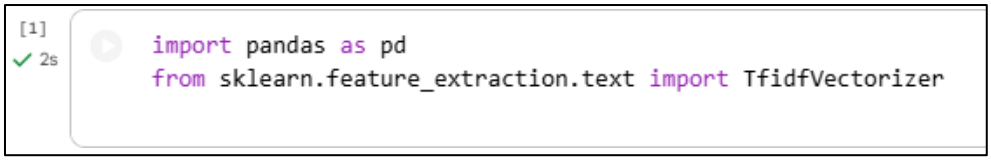
menggunakan metode *Term Frequency–Inverse Document Frequency* (TF-IDF). Tahapan ini bertujuan untuk mengubah data teks menjadi bentuk numerik sehingga dapat digunakan dalam proses klasifikasi menggunakan algoritma *Naïve Bayes*.

Berdasarkan dataset yang digunakan, komentar-komentar masyarakat terhadap akun Instagram Polda Kalteng didominasi oleh kata-kata bernada positif seperti salut, sehat, sukses, luar biasa, semangat, pelayanan, dan pengamanan. Kata-kata tersebut memiliki bobot TF-IDF yang relatif lebih tinggi karena sering muncul pada komentar positif namun tidak terlalu umum pada seluruh dokumen, sehingga dianggap memiliki tingkat kepentingan yang tinggi dalam proses klasifikasi sentimen.

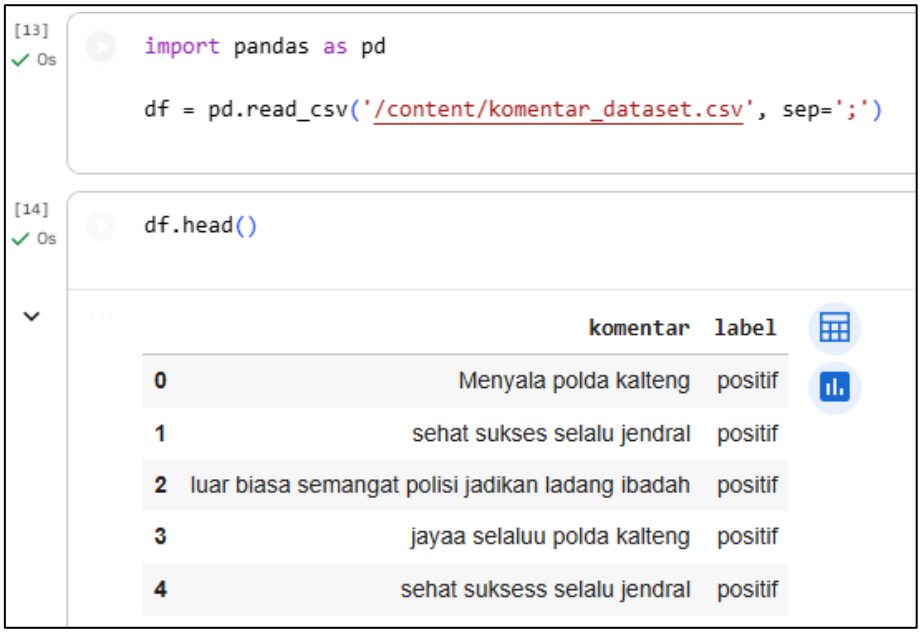
Sebagai contoh, pada komentar “Sip salut untuk tim polda kalteng”, kata salut, tim, dan polda memperoleh bobot TF-IDF yang lebih besar dibandingkan kata sambung, karena kata tersebut mencerminkan ekspresi apresiasi secara langsung. Demikian pula pada komentar “Terima kasih atas pengamanan yang maksimal”, kata terima kasih dan pengamanan memiliki bobot signifikan yang menunjukkan sentimen positif terhadap kinerja kepolisian.

Berikut hasil ekstraksi TF-IDF.

Sebelum melakukan perhitungan TF-IDF manual peneliti melakukan dengan perhitungan dengan *google colabs*.



Gambar 4.8 Mulai Perhitungan TF IDF



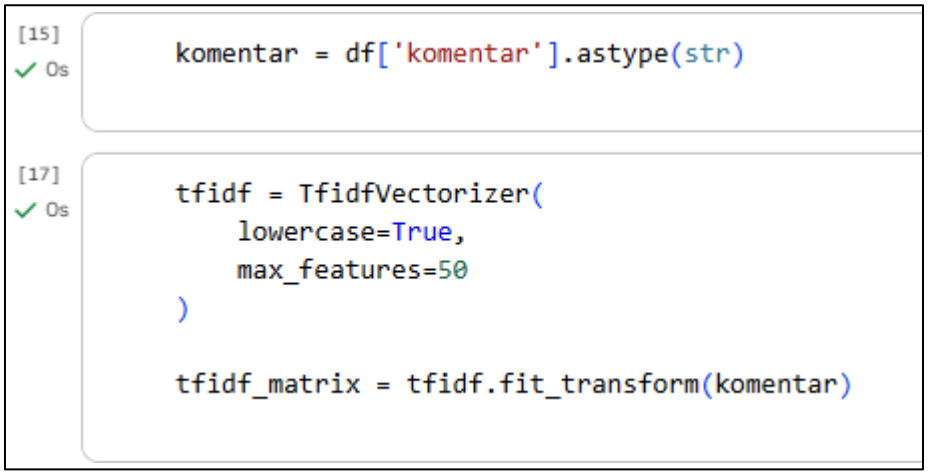
```
[13]
✓ 0s
import pandas as pd

df = pd.read_csv('/content/komentar_dataset.csv', sep=';')
```

```
[14]
✓ 0s
df.head()
```

	komentar	label
0	Menyala polda kalteng	positif
1	sehat sukses selalu jendral	positif
2	luar biasa semangat polisi jadikan ladang ibadah	positif
3	jayaa selaluu polda kalteng	positif
4	sehat suksess selalu jendral	positif

Gambar 4.9 Memanggil Komentar TF IDF

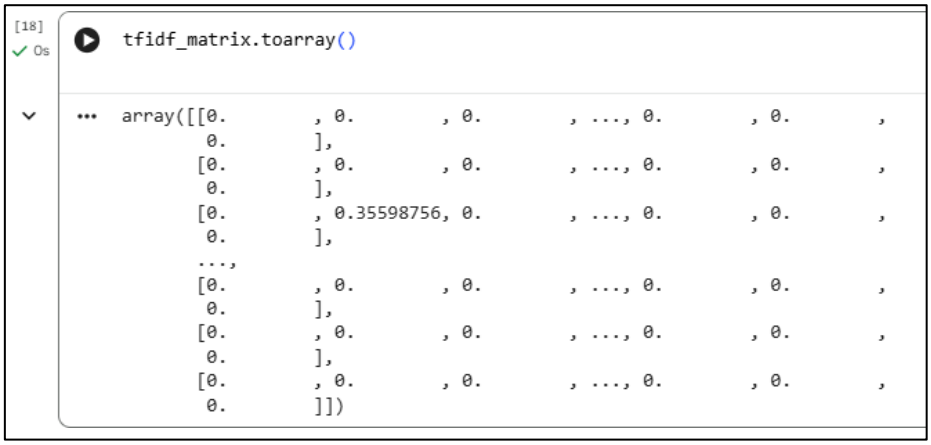


```
[15]
✓ 0s
komentar = df['komentar'].astype(str)
```

```
[17]
✓ 0s
tfidf = TfidfVectorizer(
    lowercase=True,
    max_features=50
)

tfidf_matrix = tfidf.fit_transform(komentar)
```

Gambar 4.10 Memanggil Fungsi Pencarian Komentar TF IDF



Gambar 4.11 Hasil Komentar TF IDF

Tabel 4.2 *Document Frequency* (DF)

Term	**DF**
menyala	18
polda	17
kalteng	16
polri	15
jaya	14
bravo	6
sukses	6
lambat	1

Tabel 4.3 Perhitungan IDF

Term	**DF**
menyala	log(50/18) = 0.44
polda	log(50/17) = 0.47
kalteng	log(50/16) = 0.49
polri	log(50/15) = 0.52
jaya	log(50/14) = 0.55
bravo	log(50/6) = 0.92
sukses	log(50/6) = 0.92
lambat	log(50/1) = 1.69

Komentar 1 : Menyala polda kalteng

Tabel 4.4 Perhitungan IDF Komentar 1

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 2 : sehat sukses selalu jendral

Tabel 4.5 Perhitungan IDF Komentar 2

Term	TF	IDF	TF-IDF
sehat	1	1.0969	1.0969
sukses	1	1.2218	1.2218
selalu	1	0.7959	0.7959
jendral	1	1.0000	1.0000

Komentar 3 : luar biasa semangat polisi jadikan ladang ibadah

Tabel 4.6 Perhitungan IDF Komentar 3

Term	TF	IDF	TF-IDF
luar	1	1.3979	1.3979
biasa	1	1.3979	1.3979
semangat	1	1.3979	1.3979
polisi	1	1.6990	1.6990
jadikan	1	1.6990	1.6990
ladang	1	1.6990	1.6990
ibadah	1	1.6990	1.6990

Komentar 4 : jayaa selaluu polda kalteng

Tabel 4.7 Perhitungan IDF Komentar 4

Term	TF	IDF	TF-IDF
jayaa	1	1.6990	1.6990
selaluu	1	1.6990	1.6990
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 5 : sehat suksess selalu jendral

Tabel 4.8 Perhitungan IDF Komentar 5

Term	TF	IDF	TF-IDF
sehat	1	1.0969	1.0969
suksess	1	1.2218	1.2218
selalu	1	0.7959	0.7959
jendral	1	1.0000	1.0000

Komentar 6 : sehat suksess selalu jendral

Tabel 4.9 Perhitungan IDF Komentar 6

Term	TF	IDF	TF-IDF
sehat	1	1.0969	1.0969
suksess	1	1.2218	1.2218
selalu	1	0.7959	0.7959
jendral	1	1.0000	1.0000

Komentar 7 : Jayajaya

Tabel 4.10 Perhitungan IDF Komentar 7

Term	TF	IDF	TF-IDF
jayajaya	1	1.6990	1.6990

Komentar 8 : polda kalteng jaya

Tabel 4.11 Perhitungan IDF Komentar 8

Term	TF	IDF	TF-IDF
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198
jaya	1	0.7959	0.7959

Komentar 9 : mantap polri jaya

Tabel 4.12 Perhitungan IDF Komentar 9

Term	TF	IDF	TF-IDF
mantap	1	1.3979	1.3979
polri	1	0.5850	0.5850
jaya	1	0.7959	0.7959

Komentar 10 : jaya terus polda kalteng

Tabel 4.12 Perhitungan IDF Komentar 10

Term	TF	IDF	TF-IDF
jaya	1	0.7959	0.7959
terus	1	1.2218	1.2218
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 11 : menyalaa polrii

Tabel 4.13 Perhitungan IDF Komentar 11

Term	TF	IDF	TF-IDF
menyalaa	1	1.2218	1.2218
polrii	1	1.0969	1.0969

Komentar 12 : menyalaa polrii ku

Tabel 4.14 Perhitungan IDF Komentar 12

Term	TF	IDF	TF-IDF
menyalaa	1	1.2218	1.2218
polrii	1	1.0969	1.0969
ku	1	1.6990	1.6990

Komentar 13 : bravo polri untuk Indonesia

Tabel 4.15 Perhitungan IDF Komentar 13

Term	TF	IDF	TF-IDF
bravo	1	1.2218	1.2218
polri	1	0.5850	0.5850
untuk	1	1.3979	1.3979
indonesia	1	1.6990	1.6990

Komentar 14 : umaa jabul nah

Tabel 4.16 Perhitungan IDF Komentar 14

Term	TF	IDF	TF-IDF
umaa	1	1.6990	1.6990
jabul	1	1.6990	1.6990
nah	1	1.6990	1.6990

Komentar 15 : menyala

Tabel 4.17 Perhitungan IDF Komentar 15

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850

Komentar 16 : menyala polri

Tabel 4.18 Perhitungan IDF Komentar 16

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850
polri	1	0.5850	0.5850

Komentar 17 : berkah

Tabel 4.19 Perhitungan IDF Komentar 17

Term	TF	IDF	TF-IDF
berkah	1	1.6990	1.6990

Komentar 18 : menyala jendral

Tabel 4.20 Perhitungan IDF Komentar 18

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850
jendral	1	1.0000	1.0000

Komentar 19 : niceee kinggg

Tabel 4.21 Perhitungan IDF Komentar 19

Term	TF	IDF	TF-IDF
niceee	1	1.6990	1.6990
kinggg	1	1.6990	1.6990

Komentar 20 : menyala terus polda kalteng

Tabel 4.22 Perhitungan IDF Komentar 20

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850
terus	1	1.2218	1.2218
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 21 : jaya polrii

Tabel 4.23 Perhitungan IDF Komentar 21

Term	TF	IDF	TF-IDF
jaya	1	0.7959	0.7959
polrii	1	1.0969	1.0969

Komentar 22 : luar biasa

Tabel 4.24 Perhitungan IDF Komentar 22

Term	TF	IDF	TF-IDF
luar	1	1.3979	1.3979
biasa	1	1.3979	1.3979

Komentar 23 : tetap semangat selalu polri

Tabel 4.25 Perhitungan IDF Komentar 23

Term	TF	IDF	TF-IDF
tetap	1	1.6990	1.6990
semangat	1	1.3979	1.3979
selalu	1	0.7959	0.7959
polri	1	0.5850	0.5850

Komentar 24 : polda kalteng jayaaa

Tabel 4.26 Perhitungan IDF Komentar 24

Term	TF	IDF	TF-IDF
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198
jayaaa	1	1.6990	1.6990

Komentar 25 : menyala selalu polda kalteng

Tabel 4.27 Perhitungan IDF Komentar 25

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850
selalu	1	0.7959	0.7959
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 26 : bravoo polri

Tabel 4.28 Perhitungan IDF Komentar 26

Term	TF	IDF	TF-IDF
bravoo	1	1.6990	1.6990
polri	1	0.5850	0.5850

Komentar 27 : polri untuk Masyarakat

Tabel 4.29 Perhitungan IDF Komentar 27

Term	TF	IDF	TF-IDF
polri	1	0.5850	0.5850
untuk	1	1.3979	1.3979
masyarakat	1	1.3979	1.3979

Komentar 28 : jaya selalu polda kalteng

Tabel 4.30 Perhitungan IDF Komentar 28

Term	TF	IDF	TF-IDF
jaya	1	0.7959	0.7959
selalu	1	0.7959	0.7959
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 29 : menyala

Tabel 4.31 Perhitungan IDF Komentar 29

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850

Komentar 30 : sukses selalu polda kalteng

Tabel 4.32 Perhitungan IDF Komentar 30

Term	TF	IDF	TF-IDF
sukses	1	1.2218	1.2218
selalu	1	0.7959	0.7959
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 31 : masih kurang lagi menindak

Tabel 4.33 Perhitungan IDF Komentar 31

Term	TF	IDF	TF-IDF
masih	1	1.6990	1.6990
kurang	1	1.6990	1.6990
lagi	1	1.6990	1.6990
menindak	1	1.6990	1.6990

Komentar 32 : kalteng respek bravo polda kalteng

Tabel 4.34 Perhitungan IDF Komentar 32

Term	TF	IDF	TF-IDF
kalteng	2	0.6198	1.2396
respek	1	1.6990	1.6990
bravo	1	1.2218	1.2218
polda	1	0.5850	0.5850
kalteng	2	0.6198	1.2396

Komentar 33 : terima kasih polri slalu mnjadi penolong Masyarakat

Tabel 4.35 Perhitungan IDF Komentar 33

Term	TF	IDF	TF-IDF
terima	1	1.6990	1.6990
kasih	1	1.6990	1.6990
polri	1	0.5850	0.5850
slalu	1	1.6990	1.6990
mnjadi	1	1.6990	1.6990
penolong	1	1.6990	1.6990
masyarakat	1	1.3979	1.3979

Komentar 34 : sehat n suksess selalu jendral

Tabel 4.36 Perhitungan IDF Komentar 34

Term	TF	IDF	TF-IDF
sehat	1	1.0969	1.0969
n	1	1.6990	1.6990
suksess	1	1.2218	1.2218
selalu	1	0.7959	0.7959
jendral	1	1.0000	1.0000

Komentar 35 : menyala polri

Tabel 4.37 Perhitungan IDF Komentar 35

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850
polri	1	0.5850	0.5850

Komentar 36 : jaya jaya jaya

Tabel 4.38 Perhitungan IDF Komentar 36

Term	TF	IDF	TF-IDF
jaya	3	0.7959	2.3876
jaya	3	0.7959	2.3876
jaya	3	0.7959	2.3876

Komentar 37 : mantap polda kaltengggg

Tabel 4.39 Perhitungan IDF Komentar 37

Term	TF	IDF	TF-IDF
mantap	1	1.3979	1.3979
polda	1	0.5850	0.5850
kaltengggg	1	1.6990	1.6990

Komentar 38 : polri jaya

Tabel 4.40 Perhitungan IDF Komentar 38

Term	TF	IDF	TF-IDF
polri	1	0.5850	0.5850
jaya	1	0.7959	0.7959

Komentar 39 : sukses polrii

Tabel 4.41 Perhitungan IDF Komentar 39

Term	TF	IDF	TF-IDF
sukses	1	1.2218	1.2218
polrii	1	1.0969	1.0969

Komentar 40 : menyalaa

Tabel 4.42 Perhitungan IDF Komentar 40

Term	TF	IDF	TF-IDF
menyalaa	1	1.2218	1.2218

Komentar 41 : bravo polda kalteng

Tabel 4.43 Perhitungan IDF Komentar 41

Term	TF	IDF	TF-IDF
bravo	1	1.2218	1.2218
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 42 : menyala

Tabel 4.44 Perhitungan IDF Komentar 41

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850

Komentar 43 : menyala polda kalteng

Tabel 4.45 Perhitungan IDF Komentar 42

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850
polda	1	0.5850	0.5850
kalteng	1	0.6198	0.6198

Komentar 44 : menyala

Tabel 4.46 Perhitungan IDF Komentar 43

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850

Komentar 45 : gas terus polri

Tabel 4.47 Perhitungan IDF Komentar 45

Term	TF	IDF	TF-IDF
gas	1	1.6990	1.6990
terus	1	1.2218	1.2218
polri	1	0.5850	0.5850

Komentar 46 : polri semakin di depan

Tabel 4.48 Perhitungan IDF Komentar 46

Term	TF	IDF	TF-IDF
polri	1	0.5850	0.5850
semakin	1	1.6990	1.6990
di	1	1.6990	1.6990
depan	1	1.6990	1.6990

Komentar 47 : polri jaya

Tabel 4.49 Perhitungan IDF Komentar 47

Term	TF	IDF	TF-IDF
polri	1	0.5850	0.5850
jaya	1	0.7959	0.7959

Komentar 48 : menyala

Tabel 4.50 Perhitungan IDF Komentar 48

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850

Komentar 49 : menyala polri

Tabel 4.51 Perhitungan IDF Komentar 49

Term	TF	IDF	TF-IDF
menyala	1	0.5850	0.5850
polri	1	0.5850	0.5850

Komentar 50 : terlalu lambat

Tabel 4.52 Perhitungan IDF Komentar 50

Term	TF	IDF	TF-IDF
terlalu	1	1.6990	1.6990
lambat	1	1.6990	1.6990

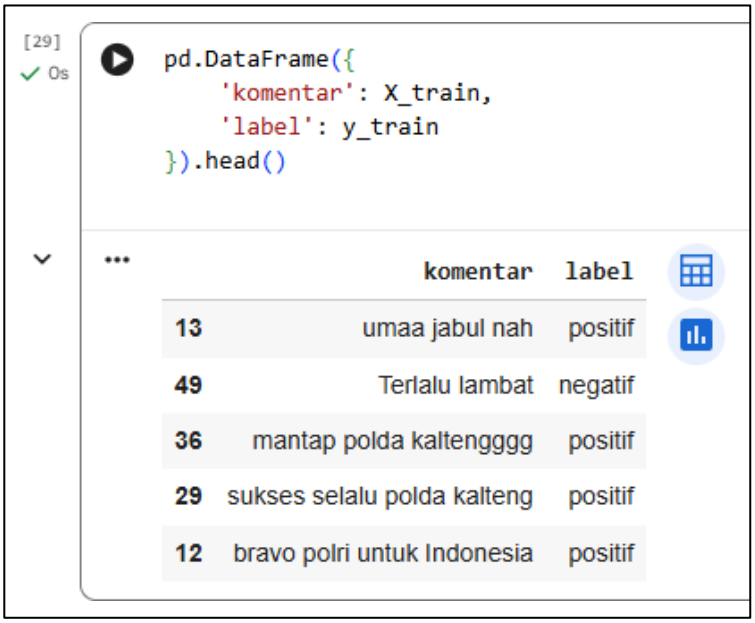
Kata dengan nilai TF-IDF tinggi menunjukkan tingkat kepentingan yang besar dalam menentukan sentimen. Kata 'lambat' memiliki bobot tinggi karena jarang muncul namun bermakna kuat. Kata apresiatif seperti “menyala”, “jaya”, dan “bravo” mendominasi sentimen positif.

4.1.3. Hasil Pembagian Dataset

Setelah proses ekstraksi fitur menggunakan metode TF-IDF, tahap selanjutnya adalah pembagian dataset (*dataset splitting*). Pembagian ini dilakukan untuk memisahkan data yang digunakan dalam proses pelatihan (*training*) dan pengujian (*testing*) model *Naïve Bayes*. Tujuan dari pembagian dataset adalah untuk mengukur kemampuan model dalam mengklasifikasikan data yang belum pernah dilihat sebelumnya.

Pada penelitian ini, dataset yang digunakan berjumlah 50 komentar yang diambil dari kolom komentar akun Instagram Polda Kalteng.

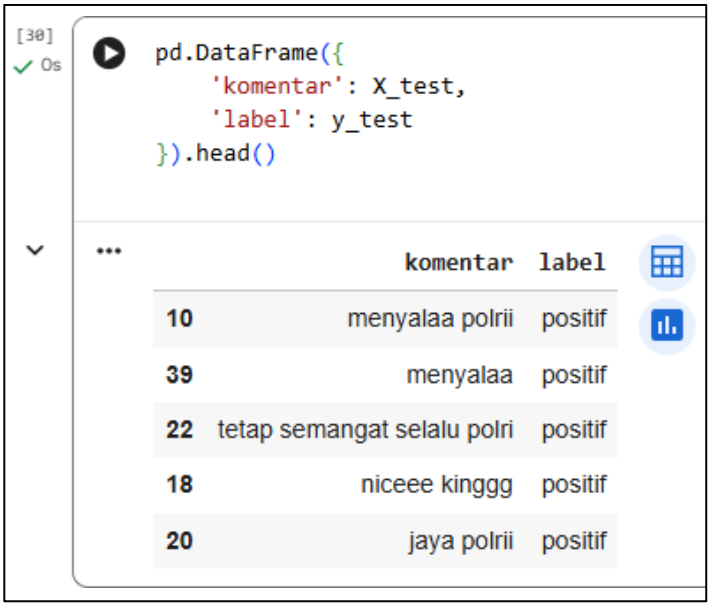
Klasifikasi Pelatihan Model



	komentar	label
13	umaa jabul nah	positif
49	Terlalu lambat	negatif
36	mantap polda kaltengggg	positif
29	sukses selalu polda kalteng	positif
12	bravo polri untuk Indonesia	positif

Gambar 4.12 Hasil Komentar TF IDF 61

Klasifikasi *Testing* Model



	komentar	label
10	menyalaa polrii	positif
39	menyalaa	positif
22	tetap semangat selalu polri	positif
18	niceee kinggg	positif
20	jaya polrii	positif

Gambar 4.13 Hasil Komentar TF IDF

4.2 Implementasi

Berdasarkan hasil pengujian yang telah dilakukan pada 50 data komentar, diperoleh hasil bahwa mayoritas komentar masyarakat terhadap Polda Kalteng bersentimen positif. Hal ini menunjukkan bahwa citra dan kinerja kepolisian di mata masyarakat, khususnya di media sosial Instagram, cenderung mendapatkan respon yang baik. Komentar positif umumnya berkaitan dengan apresiasi terhadap pelayanan, pengamanan, sikap profesional, serta respons cepat aparat kepolisian.

Proses *preprocessing* teks memiliki peran yang sangat penting dalam penelitian ini. Tahapan seperti *case folding*, *tokenizing*, *stopword removal*, dan *stemming* berhasil menghilangkan unsur-unsur yang tidak relevan dalam teks komentar, sehingga kata-kata kunci yang bermakna dapat diproses secara optimal oleh sistem. Dengan *preprocessing* yang baik, *noise* dalam data dapat diminimalkan dan akurasi model meningkat.

Selanjutnya, penerapan TF-IDF (*Term Frequency–Inverse Document Frequency*) sebagai metode ekstraksi fitur terbukti efektif dalam merepresentasikan teks komentar ke dalam bentuk numerik. TF-IDF mampu memberikan bobot lebih besar pada kata-kata yang sering muncul dalam satu komentar namun jarang muncul pada komentar lain, sehingga membantu model *Naïve Bayes* dalam membedakan karakteristik sentimen positif dan negatif. Kata-kata seperti "terima kasih", "pelayanan", "baik", "maksimal" memiliki bobot TF-IDF tinggi pada komentar positif, sedangkan kata seperti "lambat" dan "kurang sigap" lebih dominan pada komentar negatif.

Secara keseluruhan, hasil penelitian ini menunjukkan bahwa analisis sentimen berbasis *Naïve Bayes* dapat menjadi alat bantu yang efektif bagi institusi kepolisian dalam memantau opini publik di dunia maya. Informasi sentimen yang diperoleh dapat digunakan sebagai bahan evaluasi kinerja, peningkatan kualitas pelayanan, serta perumusan strategi komunikasi publik yang lebih baik di masa mendatang.

Berdasarkan hasil pengujian, seluruh komentar dengan sentimen positif berhasil diklasifikasikan dengan benar oleh model ke dalam kelas positif, sedangkan komentar dengan sentimen negatif juga berhasil diprediksi sesuai dengan label aslinya.

BAB V

KESIMPULAN DAN SARAN

5.1 KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan mengenai Peran Analisis Sentimen Berbasis *Naïve Bayes* pada Kepolisian Berbasis Online, dapat disimpulkan:

a. Analisis sentimen berbasis *Naïve Bayes* mampu memberikan gambaran persepsi masyarakat terhadap institusi kepolisian, khususnya Polda Kalimantan Tengah, melalui komentar di media sosial Instagram.
b. Tahapan penelitian yang meliputi pengumpulan data, pelabelan manual, *preprocessing* teks, ekstraksi fitur TF-IDF, serta pelatihan dan pengujian model menunjukkan bahwa metode *Naïve Bayes* efektif dalam mengklasifikasikan sentimen komentar menjadi positif dan negatif.
c. Dominasi sentimen positif pada komentar Instagram Polda Kalimantan Tengah menandakan bahwa persepsi masyarakat terhadap kinerja dan pelayanan kepolisian berada pada kategori baik.
d. Analisis sentimen berbasis *Naïve Bayes* dapat digunakan sebagai alat bantu yang sederhana dan efektif untuk memantau opini publik terhadap institusi kepolisian di media online.

5.2 SARAN

Berdasarkan hasil dan keterbatasan penelitian ini, penulis memberikan beberapa saran sebagai berikut:

a. Pengembangan metode lainnya seperti *Vector Machin*e (SVM) dan *Random Forest*.
b. Sumber Data lebih banyak tidak hanya sosial media tetapi pengumpulan data dan sumber data bersifat *offline*/secara langsung.
c. Dikembangkan menjadi sebuah sistem baru seperti sistem berbasis website menggunakan bahas pemrograman tertentu.

DAFTAR PUSTAKA

Azza, N., & Hadian, N. (2025). Tweet Sentiment Analysis with Support Vector Machine (SVM) Algorithm for PT. XYZ Digital Strategy. G-Tech: Jurnal Teknologi Terapan, 9(4).

Darwis, D., Siskawati, N., & Abidin, Z. (2021). Penerapan Algoritma Naive Bayes Untuk Analisis Sentimen Review Data Twitter Bmkg Nasional. Jurnal Tekno Kompak, 15(1), 131-145.

Gunasekaran, S., Alabed, S., Gulsin, G. S., & Hanneman, K. (2023). Top 2023 Images in Cardiothoracic Imaging. Radiology: Cardiothoracic Imaging, 5(6).

Guantara, R. G. (2023). Pemanfaatan Google Colab Untuk Aplikasi Pendeteksian Masker Wajah Menggunakan Algoritma Deep Learning YOLOv7. *Jurnal Teknologi Dan Sistem Informasi Bisnis*, *5*(1), 55-60.

Meisyaroh, S., Kraugusteeliana, K., Irwanto, I., Dawami, D., Daud, R. F., Khairunnisa, K., ... & Doho, D. B. (2023). MEDIA SIBER.

Nasrullah, R. (2022). Teori dan riset media siber (cybermedia). Prenada Media.

Nurpandi, F., Sulaeman, F. S., & Hermawan, A.(2024). Analisis Sentimen Terhadap Kinerja Kepolisian Indonesia Menggunakan Metode Multinomial *Naïve Bayes*, Long Short-Term Memory, dan Lexicon-Based. *Media Jurnal Informatika*, 16(1), 1-10.

Rony, O. (2024). Penerapan Naïve Bayes Untuk Klasifikasi Sentiment Opini Netizen Tentang Kinerja Polisi Negara Republik Indonesia (*Doctoral dissertation, Universitas Islam Negeri Sultan Syarif Kasim Riau*).

Rosaly, R., & Prasetyo, A. (2019). *Pengertian Flowchart Beserta Fungsi dan Simbol-simbol Flowchart yang Paling Umum Digunakan*.

Sari, L., & Sari siregar, G. yanti kemala. (2021). Perancangan Aplikasi Pendataan Data Kepegawaian Negeri Sipil Pada Dinas Komunikasi Dan Informatika kota Metro. Jurnal Mahasiswa Ilmu Komputer, 2(1), 115–135. https://doi.org/10.24127/.v2i1.1235.

Solihin, O., Widiono, S., & Evanne, L. (2025). Komunikasi Kesehatan Dalam Isu Kebijakan Publik: Analisis Tayangan Youtube'Polemik Aturan BPJS 2025. Jurnal Agregasi: Aksi Reformasi Government dalam Demokrasi, 13(2), 164-180.

Susana, H. (2022). Penerapan Model Klasifikasi Metode *Naïve Bayes* Terhadap Penggunaan Akses Internet. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, *4*(1), 1-8.

Urva, G., Albanna, I., Sungkar, M. S., Gunawan, I. M. A. O., Adhicandra, I., Ramadhan, S., ... & Junaidi, S. (2023). Penerapan Data Mining Di Berbagai Bidang: Konsep, Metode, dan Studi Kasus. PT. Sonpedia Publishing Indonesia.

Wardani, Y. Arum Sari, Indriati. (2021) “Analisis Sentimen menggunakan Metode Naïve Bayes Classifier terhadap Review Produk Perawatan Kulit Wajah menggunakan Seleksi Fitur N-gram dan Document Frequency Thresholding.” Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, Vol. 5, No. 12

Wati, R., Ernawati, S., & Rachmi, H. (2023). Pembobotan TF-IDF Menggunakan Naïve Bayes pada Sentimen Masyarakat Mengenai Isu Kenaikan BIPIH. Jurnal Manajemen Informatika (JAMIKA), 13(1), 84-93.

Zufria, I., Lubis, A. H., & Febiyaula, S. S. (2024). Analisis Sentimen Kepercayaan Masyarakat Terhadap Kepolisian Republik Indonesia Menggunakan Algoritma Svm. *Journal Of Science And Social Research*, 7(3),.

LAMPIRAN

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangka Raya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

SURAT TUGAS PEMBIMBING TUGAS AKHIR

NO. 275/STMIK-3.C.2/AU/XI/2025

Ketua Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya menugaskan nama-nama tersebut di bawah ini :

1. Nama : Lili Rusdiana, M.Kom
   NIK : 198707282011007
   Sebagai Pembimbing I dalam **PEMBUATAN PROGRAM**
2. Nama : Sam'ani, S.T., M.Kom
   NIK : 197703252005105
   Sebagai Pembimbing II dalam **PENULISAN TUGAS AKHIR**

Untuk membimbing Tugas Akhir mahasiswa :

Nama	:	DHEA AULIDNASARI
NIM	:	C2255201131
Program Studi	:	TEKNIK INFORMATIKA (55201)
Tanggal Daftar	:	30/10/2025
Judul Tugas Akhir	:	PERAN ANALISIS SENTIMEN BERBASIS NAÏVE BAYES PADA BERITA KEPOLISIAN BERBASIS ONLINE

Demikian surat ini dibuat agar dapat dipergunakan sebagaimana mestinya, dan dilaksanakan dengan penuh tanggung jawab.

Palangka Raya, 03 November 2025

Ketua Program Studi,

November 2025

Ir. Abdul Hadi, S.T., M.Kom
NIK. 198505102021101

TEMBUSAN :
1. Pembimbing I dan II
2. Mahasiswa yang bersangkutan
3. Arsip Prodi TI

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangka Raya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

SURAT TUGAS PENGUJI TUGAS AKHIR

358/STMIK-3.C.2/AU/XII/2025

Ketua Program Studi Teknik Informatika Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Palangkaraya menugaskan kepada nama- nama berikut :

1. Nama : Dr. Maura. Widyaningsih, S.Kom., M.Cs.
   NIK : 197502232009002
   Sebagai Ketua
2. Nama : Ir. Abdul Hadi, S.T., M.Kom
   NIK : 198505102021101
   Sebagai Sekretaris
3. Nama : Veny Cahya Hardita, M.Kom
   NIK : 199504302020002
   Sebagai Anggota
4. Nama : Lili Rusdiana, M.Kom
   NIK : 198707282011007
   Sebagai Anggota
5. Nama : Catharina Elmayantie, M.Pd
   NIK : 197610252015003
   Sebagai Anggota

Untuk menguji Tugas Akhir mahasiswa :

Nama Mahasiswa	:	DHEA AULIDNASARI
N I M	:	C2255201131
Hari / Tanggal Ujian	:	Sabtu, 20 Desember 2025
Waktu	:	07.00 WIB
Judul TA	:	PERAN ANALISIS SENTIMEN BERBASIS NAÏVE BAYES PADA BERITA KEPOLISIAN BERBASIS ONLINE
Ruang	:	Sidang 1

Demikian surat ini dibuat agar dapat dipergunakan sebagaimana mestinya dan dilaksanakan dengan penuh tanggung jawab.

Palangka Raya, 16 Desember 2025

Ketua Program Studi Teknik Informatika,



16 Desember 2025

Ir. ABDUL HADI, S.T., M.Kom
NIK. 198505102021101

Tembusan :
1. Dosen Penguji
2. Mahasiswa yang Bersangkutan
3. Arsip

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangkaraya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

KARTU KEGIATAN KONSULTASI TUGAS AKHIR

Nama Mahasiswa : DHEA AULIDHASARI
NIM : C2255201131
Tanggal Persetujuan Judul : 3 NOVEMBER 2025
Judul Tugas Akhir : PERAN ANALISIS SENTIMEN BERBASIS NAIVE BAYES PADA BERITA KEPOLISIAN BERBASIS ONLINE

No.	Tanggal Konsultasi		Uraian	Tanda Tangan
	Terima	Kembali		
	3/11 25		- Konsul daring	
	3/11 25		- Revisi Bab I	
	4/11 25		- Revisi bab II	
	4/11 25		- Lengkapi naskah - Acc Daftar Seminar	
	10/10 25		sesuaikan isi bab dengan topik	
	1/11 25		perbaiki bab 1 & 2 menyesuaikan isi teori untuk bab 3	
	6/11 25		-Benahi penulisan bab 1 sampai 3 -lengkapi untuk mendaftar seminar	
	7/11 25		- Acc seminar	
	5/12 25		- Sesuaikan bab 4 dengan penulisan laporan lanjut bab 5	
	8/12 25		Benahi bab 4 dan bab 5	
	11/12 25		lengkapi laporan	
	13/12 25		Acc sidang	

No.	Tanggal Konsultasi		Uraian	Tanda Tangan
	Terima	Kembali		
		1/12	Revisi	
		3/12	Revisi	
		6/12	Revisi	
		9/12	Lengkapi	
		13/12	Acc Daftar Sidang	

Menyetujui :

Dosen Pembimbing I,

LILI RUSDIANA, M.KOM

Dosen Pembimbing II,

SAM'ANI, S.T., M.KOM

SEKOLAH TINGGI MANAJEMEN INFORMATIKA DAN KOMPUTER
(STMIK) PALANGKARAYA
Jl. G. Obos No.114 Telp.0536-3224593, 3225515 Fax.0536-3225515 Palangka Raya
email : humas@stmikplk.ac.id - website : www.stmikplk.ac.id

BERITA ACARA
UJIAN TUGAS AKHIR

Periode : Desember 2025

Pada hari ini Sabtu, 20 Desember 2025 Pukul 07.00 WIB sampai dengan ..... WIB telah dilaksanakan Ujian Tugas Akhir mahasiswa atas nama :

Nama Mahasiswa	:	DHEA AULIDNASARI
N I M	:	C2255201131
Program Studi	:	Teknik Informatika
Tahun Angkatan	:	2022
Judul TA	:	PERAN ANALISIS SENTIMEN BERBASIS NAÏVE BAYES PADA BERITA KEPOLISIAN BERBASIS ONLINE

Dosen Penguji :

	Nama		Nilai	Tanda Tangan
1.	Dr. Maura. Widyaningsih, S.Kom., M.Cs	=	.....	(.....)
2.	Ir. Abdul Hadi, S.T., M.Kom	=	.....	(.....)
3.	Veny Cahya Hardita, M.Kom	=	.....	(.....)
4.	Lili Rusdiana, M.Kom	=	.....	(.....)
5.	Catharina Elmayantie, M.Pd	=	.....	(.....)

Hasil Ujian : **LULUS / TIDAK LULUS *)** NILAI = ... 80,15 .....

Revisi (Tidak Ada / Minor / Mayor *)

Catatan Penting :
1. Lama Perbaikan : ..... hari
2. Jika lebih dari 1 (satu) bulan dikenakan sanksi berupa denda sebesar Rp. 600.000,- (Enam ratus ribu rupiah) per bulan dari tanggal ujian.
3. Jika lebih dari 3 (tiga) bulan dari tanggal ujian maka hasil ujian dibatalkan dan wajib mengajukan judul dan pembimbing baru.

Demikian Berita Acara ini dibuat, agar dapat digunakan sebagaimana semestinya.

Palangka Raya, 18 Desember 2025

Ketua Penguji,

Dr. Maura. Widyaningsih, S.Kom., M.Cs
NIK. 197502232009002

Mahasiswa,

DHEA AULIDNASARI
NIM. C2255201131

Mengetahui :
Ketua Program Studi Teknik Informatika,

Ir. Abdul Hadi, S.T., M.Kom
NIK. 198505102021101

Tembusan:
Mahasiswa yang bersangkutan.
Berita acara ini diberikan setelah menyelesaikan semua revisi dari semua penguji
**) Coret yang tidak perlu*