ANALISIS SENTIMEN TERHADAP OPINI WARGANET TENTANG WISATA D.I YOGYAKARTA PADA PLATFORM INSTAGRAM MENGGUNAKAN NAÏVE BAYES CLASSIFIER

(Studi Kasus : Data Opini Warganet di Instagram Selama Satu Tahun)

TUGAS AKHIR

Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Program Studi Statistika

Disusun Oleh:

Luthfiana Laksmi Larasati

17611113

PROGRAM STUDI STATISTIKA

FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM

UNIVERSITAS ISLAM INDONESIA

YOGYAKARTA

2024

HALAMAN PERSETUJUAN PEMBIMBING

TUGAS AKHIR

Judul	:	Analisis Sentimen Terhadap Opini Warganet Tentang Wisata D I Yogyakarta pada Platform Instagram Menggunakan *Naïve Bayes Classifier*.
Nama Mahasiswa	:	Luthfiana Laksmi Larasati
NIM	:	17611113

TUGAS AKHIR INI TELAH DIPERIKSA DAN DISETUJUI UNTUK DIUJIKAN

Mengetahui	Yogyakarta, 31 Januari 2024
Ketua Prodi Statistika	Dosen Pembimbing
Dr. Atina Ahdika, S.Si., M.Si.	Achmad Fauzan, S.Pd., M.Si.

HALAMAN PENGESAHAN

TUGAS AKHIR

ANALISIS SENTIMEN TERHADAP OPINI WARGANET TENTANG WISATA D.I YOGYAKARTA PADA PLATFORM INSTAGRAM MENGGUNAKAN NAÏVE BAYES CLASSIFIER

(Studi Kasus : Data Opini Warganet di Instagram Selama Satu Tahun)

Nama Mahasiswa : Luthfiana Laksmi Larasati

NIM : 17611113

TUGAS AKHIR INI TELAH DIUJIKAN

PADA TANGGAL : 2 Februari 2024

Nama Penguji	Tanda Tangan
1. Ayundyah Kesumawati, S.Si., M.Si.	.....
2. Sekti Kartika Dini, S.Si., M.Si.	.....
3. Achmad Fauzan, S.Pd., M.Si.	.....

Mengetahui,

Dekan Fakultas Matematika dan Ilmu Pengetahuan Alam

UNIVERSITAS ISLAM INDONESIA YOGYAKARTA FAKULTAS MIPA

(Prof. Riyanto, S.Pd., M.Si., Ph.D.)

KATA PENGANTAR

بِسْمِ اللَّهِ الرَّحْمَنِ الرَّحِيمِ

Assalamu'alaikum Wr.Wb

Alhamdulillah, puji syukur terhaturkan pada Allah SWT atas segala berkah dan rahmatnya sehingga terselesaikannya penyusunan TA/ skripsi ini sebagai syarat kelulusan juga diperolehnya gelar Sarjana Statistika, setelah perjalanan perkuliahan yang panjang dan berliku.

Terimakasih kepada Bapak dan Ibu, sebagai orangtua kandung yang telah sampai pada salah satu tujuan dan doanya, membersamai proses perkuliahan yang tidak mudah hingga mengantarkan putrinya pada tahap kelulusan. Semoga segala doa, perjuangan, dan pengorbanan Bapak dan Ibu berbalas barokah dan ridho dari Allah SWT. Dan segala ilmu bermanfaat yang ditunaikan putrinya, dapat menjadi amal jariyah keduanya, Aamiin…

Terimakasih kepada Bapak dan Ibu Kaprodi Statistika UII, Bapak Dr. RB Fajriya Hakim, S.Si, M.Si., Bapak Dr. Eddy Widodo, S.Si, M.Si., dan Ibu Dr. Atina Ahdika, S.Si, M.Si. Atas kepemimpinannya yang bijak pada setiap masa jabatannya. Semoga segala berkah dan ridho Allah menjadi balasan untuk Bapak dan Ibu, Aamiin…

Terimakasih kepada Bapak dan Ibu dosen serta staff prodi statistika UII yang telah memberikan bekal pengajaran moril dan materiil, yang inshaAllah menjadi amal jariyah bagi Bapak dan Ibu sekalian, terimakasih juga atas segala doa serta arahan yang diberikan selama perjalanan kuliah yang panjang ini, terimakasih telah mengoptimalkan kinerja Bapak dan Ibu, profesional dalam segala aturan yang mendidik saya untuk disiplin serta memahami tanggung jawab. Segala doa baik saya haturkan atas segala yang diberikan Bapak dan Ibu sekalian…

Kepada Ibu Tutik Purwaningsih, S.Stat, M.Stat dan Bapak Achmad Fauzan, S.Pd, M.Si. selaku DPA dan dosbing TA, semoga Allah SWT mengabulkan segala doa serta mempermudah hajat Bapak & Ibu sebagaimana bimbingan yang Bapak&Ibu berikan, yang mempermudah saya dalam menamatkan studi…

Teruntuk Bang Tibi dan Gerryl, terimakasih telah turut membantu proses pengerjaan TA. Terutama untuk Bang Tibi yang siap sedia ditanya kapanpun itu

tentang TA, maaf dan makasih banyak ya Bang, juga Gerryl yang memberikan ide TA, serta membantu dalam proses *running coding* sampai menanyakan pada Jul, terimakasih juga ya Jul. Dhila, terimakasih sempat berkontribusi saat KP. Yuyun dan Udin, makasih sudah membersamai dalam proses perkuliahan panjang. Untuk Udin, *Mauliate Godang*, Din! Selain materiil, support moril yang kamu kasih juga kuat. *Barakallah* ya Din. *I praise Allah for sending all of ya guys, May Allah blesses you always guys as you pray for cause I can't give it back ...*

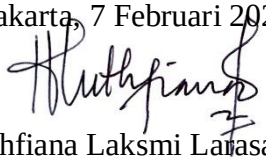
Kepada teman-temanku tersayang, Luthfi, Yuyun, Selvi, Udin, Ompel, terimakasih tidak akan cukup atas motivasi kalian, kebersamaan, kesetiaan bahkan "kesudian" kalian yang terbalut ketulusan untuk menjadi saksi atas segala lemah dan rapuhnya diri ini, berada pada titik terendah yang jauh dari kata mudah tapi atas berkah berlimpahNya, kemurahanNya, dikirimkanNya kalian sebagai anugerah yang menguatkanku agar tidak "menyerah". Atas keterbatasan diri yang tidak mampu membalas segala hal luar biasa yang kalian berikan, biarlah Allah SWT yang membalasnya dengan berkah berlimpah, yang memang pantas kalian dapatkan. Dan semoga kita tetap bersama ya, bahkan sampai nanti kita disana… *It was my blessed mess dears. Blessed me to had you in my messy time…*

Teruntuk Mas Hendra, Bude Api, Pakde Pur, *maturnuwun sanget* untuk semua motivasi, doa, dan bantuan di masa sulit dan terhimpit saat perkuliahan hingga masa revisi. Menjadi jalan keluar atas masalah yang mendesak. Segala kemurahanNya tercurah bagi Pakde, Bude, dan Mas Hendra ya… Aamiin…

Terakhir, terimakasih kepada semua pihak terkait yang tidak mampu saya sebutkan satu persatu, semoga Allah memudahkan dan memberkahi kalian semua. Aamiin….

Sekian penghaturan saya, wassalamualaikum wr wb…

Yogyakarta, 7 Februari 2024



Luthfiana Laksmi Larasati

DAFTAR ISI

DAFTAR TABEL

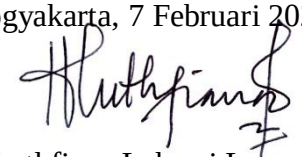
DAFTAR GAMBAR

DAFTAR LAMPIRAN

PERNYATAAN

Dengan ini saya menyatakan bahwa dalam Tugas Akhir ini tidak terdapat karya karya yang sebelumnya pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan orang lain, kecuali yang diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Yogyakarta, 7 Februari 2024



Luthfiana Laksmi Larasati

INTISARI

ANALISIS SENTIMEN TERHADAP OPINI WARGANET TENTANG WISATA D.I. YOGYAKARTA PADA PLATFORM INSTAGRAM MENGGUNAKAN NAÏVE BAYES CLASSIFIER

(Studi Kasus : Data Opini Warganet di Instagram Selama Satu Tahun)

Luthfiana Laksmi Larasati
Program Studi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Islam Indonesia

Sebagai salah satu provinsi yang memiliki potensi wisata yang cukup tinggi di Indonesia, Provinsi Daerah Istimewa Yogyakarta perlu untuk terus meningkatkan kualitas pariwisatanya melalui evaluasi yang diberikan oleh masyarakat mengenai pengalamannya ketika berwisata di provinsi tersebut. Saat ini masyarakat umum sudah menggunakan media sosial sebagai alat untuk menyampaikan tinjauannya terhadap suatu produk maupun jasa yang mereka gunakan, termasuk jasa suatu objek pariwisata. Instagram sendiri saat ini telah menjadi salah satu media sosial yang populer digunakan oleh masyarakat Indonesia dalam berbagi konten maupun tinjauan. Dalam memberikan informasi mengenai opini masyarakat terhadap pariwisata Provinsi Daerah Istimewa Yogyakarta maka perlu dilakukan klasifikasi sehingga dapat dipetakan sebaran komentar positif maupun negatifnya. Penelitian ini bertujuan untuk melakukan klasifikasi sentimen terhadap ulasan dan opini netizen Instagram mengenai pariwisata di Provinsi Daerah Istimewa Yogyakarta. Melalui teknik *scraping* di website Instagram terhadap lima akun informasi pariwisata di Provinsi Daerah Istimewa Yogyakarta, 26406 baris data berhasil diperoleh untuk dianalisis lebih lanjut. Metode yang digunakan untuk mengklasifikasikan sentimen adalah Naïve Bayes Classifier. Klasifikasi akan dibagi menjadi dua kelas yakni positif dan negatif. Proses pembentukan model menggunakan data training yang memiliki proporsi 80-20 dengan data testing. Hasil klasifikasi menunjukkan bahwa Naïve Bayes Classifier mampu memberikan akurasi sebesar 93.4% pada data testing. Adapun penelitian ini turut memberikan gambaran mengenai sebaran istilah-istilah yang masuk dalam kelompok komentar positif dan negatif menggunakan *word cloud.*

Kata Kunci : Analisis Sentimen, Instagram, Naive Bayes Classifier, Pariwisata, Provinsi Daerah Istimewa Yogyakarta

ABSTRACT

SENTIMENT ANALYSIS OF NET CITIZENS' OPINIONS ABOUT TOURISM D.I. YOGYAKARTA ON THE INSTAGRAM PLATFORM USING NAÏVE BAYES CLASSIFIER

(Case Study : *Netizen Opinion Data on Instagram for One Year*)

Luthfiana Laksmi Larasati
Department of Statistics, Faculty of Mathematics and Natural Sciences
Universitas Islam Indonesia

As one of the province that has quite high tourism potential in Indonesia, the Special Region of Yogyakarta needs to continue to improve the quality of its tourism through evaluations given by the public regarding their experiences when traveling in the province. Currently, the general public is using social media as a tool to convey their reviews of the products or services they use, including the services of a tourism object. Instagram itself has now become one of the popular social media used by Indonesian people to share content and reviews. In providing information regarding public opinion regarding the Special Region of Yogyakarta tourism, it is necessary to classify it so that the distribution of positive and negative comments can be mapped. This research aims to classify the sentiments of reviews and opinions of Instagram netizens regarding tourism in the Special Region of Yogyakarta. Through scraping techniques on the Instagram website for five tourism information accounts in the Special Region of Yogyakarta, 26,406 rows of data were obtained for further analysis. The method used to classify sentiment is the Naïve Bayes Classifier. Classification will be divided into two classes, namely positive and negative. The model formation process uses training data which has an 80-20 proportion to testing data. The classification results show that the Naïve Bayes Classifier is able to provide an accuracy of 93.4% on testing data. This research also provides an overview of the distribution of terms included in the positive and negative comment groups using a word cloud.

Keywords: *Instagram, Naive Bayes Classifier, Sentiment Analysis, Special Region of Yogyakarta, Tourism*

BAB I
PENDAHULUAN

1.1. Latar Belakang Masalah

Indonesia menjadi salah satu negara dengan potensi wisata yang sangat besar dan juga mampu memberikan kontribusi nyata bagi seitap pihak. Data dari Kementerian Pariwisata menyebutkan bahwa kontribusi nilai devisa pariwisata lebih dari 6 Miliar USD dan kontribusi terhadap PDB mencapai 3.76% pada Triwulan I Tahun 2023 (indonesia.go.id, 2023). Pada triwulan 1 tahun 2023, terdapat 2.5 juta kunjungan wisata mancanegara atau naik 508,87% dari periode tahun sebelumnya (mediakeuangan, 2023). Adapun secara keseluruhan, BPS mencatat bahwa Indonesia memiliki Objek Wisata Komersial sebanyak 2.563 usaha (dataindonesia.id, 2023). Melalui potensi besar tersebut maka berbagai lokasi di Indonesia harus saling meningkatkan kualitas pariwisatanya sehingga memperbesar kekuatan wisata di Indonesia secara keseluruhan.

Daerah Istimewa Yogyakarta merupakan salah satu Provinsi di Indonesia yang memiliki potensi wisata dalam jumlah besar. Dikutip dari JogjaDataku (2023), jumlah kunjungan wisatawan di tahun 2022 adalah 19.275.989 orang angka ini naik cukup tinggi dari tahun 2021 yang hanya berjumlah 7.590.233. Adapun di tahun 2022 terdapat 148 daya tarik baru di D.I. Yogyakarta. Selain itu kontribusi sektor akomodasi dan makan minum di DI Yogyakarta sendiri menyentuh angka 10 Miliar Rupiah di tahun 2022. Pariwisata di D.I. Yogyakarta pada akhirnya harus terus dikelola dengan baik dan berkelanjutan agar dapat memberikan dampak positif bagi setiap pihak.

Namun begitu, saat ini masih terdapat beberapa permasalahan pada pengelolaan objek pariwisata di wilayah D.I. Yogyakarta. Penelitian dari Purwaningsih et al. (2020) menyoroti bahwa di objek wisata Pantai Parangtritis masih terdapat permasalahan mengenai pengelolaan sampah dimana hal ini dapat terjadi karena kurang tertibnya wisatawan dalam membuang sampah, secara khusus sampah botol plastik. Adapun Kusuma et al. (2020) turut mengamati bahwa tempat berjalan kaki di lokasi sekitar Malioboro kurang tertata dengan rapi dan sempit sehingga menyusahkan para wisatawan untuk berjalan kaki. Selain itu Saputra

(2020) menyoroti bahwa masih terdapat berbagai permasalahan di objek wisata Taman Sari seperti belum adanya sarana dan prasarana yang berkualitas, penyampaian produk-produk unggulan maupun dari segi kelembagaannya. Berbagai permasalahan tersebut tentunya harus diidentifikasi secara menyeluruh sehingga para pemangku kepentingan dapat menentukan kebijakan yang tepat.

Evaluasi terhadap objek wisata di Yogyakarta menjadi suatu hal yang perlu untuk dilakukan karena memberikan manfaat bagi setiap pihak. Bagi perekonomian, sektor pariwisata dapat berkontribusi pada pendapatan daerah, meningkatkan kesejahteraan pelaku pariwisata serta menciptakan lapangan kerja bagi penduduk sekitar (Ahmad, 2022). Dari segi pemasaran, pariwisata juga dapat menjadi sarana *branding* baik bagi suatu daerah maupun bagi negara sehingga harus dikelola dengan baik dan benar. Dalam segi lingkungan, pengelolaan wisata mampu menjadi sarana untuk melestarikan alam dan sumber daya. Melalui pentingnya evaluasi tersebut maka perlu adanya cara untuk mengevaluasi sektor pariwisata D.I. Yogyakarta secara efektif dan efisien.

Saat ini identifikasi terhadap opini-opini masyarakat mengenai suatu objek dapat dilakukan secara cepat melalui pemantauan opini atau komentar di media sosial. Dalam konteks pariwisata, teknologi tersebut memungkinkan penggunanya untuk memberikan opini, komentar maupun menceritakan pengalamannya terhadap suatu objek wisata dengan mudah (Purba & Irwansyah, 2022). Adapun media sosial dapat juga digunakan sebagai media untuk mencari informasi mengenai suatu objek wisata melalui ulasan-ulasan yang disampaikan oleh pengguna lainnya. Konten di media sosial kerap menjadi sebuah referensi bagi seseorang sebagai dasar dalam memutuskan dalam melakukan sebuah kunjungan. Melalui berbagai konten dan data yang ada di media sosial maka dapat diketahui bagaimana opini netizen mengenai objek wisata di Yogyakarta.

Instagram merupakan salah satu media sosial yang saat ini memiliki pengguna relatif banyak di Indonesia sehingga berpengaruh juga terhadap jumlah data yang ada di media sosial tersebut. Berdasarkan statistik dari Databoks (2023), jumlah pengguna instagram di Indonesia hingga Oktober 2023 adalah 104,8 juta pengguna dimana angka itu menjadikan Instagram sebagai media sosial terbesar setelah Facebook, Youtube dan Whatsapp. Adapun melalui statistik tersebut,

Indonesia menjadi negara ke-4 dengan pengguna Instagram terbanyak. Salah satu riset turut menjelaskan bahwa setiap menitnya ada 66.000 foto yang diunggah di Instagram, hal itu tentu belum diikuti dengan berapa komentar yang mengikutinya (Riyanto & Pertiwi, 2023). Melalui besarnya potensi dari pengguna Instagram tersebut maka menjadi peluang untuk melihat bagaimana gambaran opini dan pengalaman masyarakat dalam berwisata di Yogyakarta

Dalam mengevaluasi opini suatu objek pariwisata maka perlu diketahui apakah opini tersebut memiliki sentimen yang positif atau negatif. Salah satu metode yang dapat digunakan untuk mengklasifikasikan sentimen dalam suatu opini adalah analisis sentimen. Fungsi utama dari analisis tersebut adalah mengelompokkan polaritas dari sebuah teks yang berada di dalam sebuah dokumen maupun sebuah kalimat menjadi kelas-kelas tertentu yakni positif, negatif maupun netral (Murnawan, 2017). Proses analisis sentimen berkaitan erat dengan *data mining* yang bertujuan untuk menemukan pola-pola tersembunyi dalam suatu kumpulan data yang besar. Adapun analisis sentimen memerlukan sebuah algoritma atau teknik yang digunakan untuk memodelkan data teks sehingga dapat menjadi pola tertentu. Model yang diperoleh dari analisis sentimen dapat digunakan kembali untuk memprediksi suatu dokumen atau kalimat sehingga diharapkan model yang didapat memiliki performa yang baik. Melalui analisis sentimen tersebut maka penelitian ini akan menganalisis, memodelkan serta mengklasifikasikan opini netizen di media sosial Instagram mengenai pariwisata kota Yogyakarta.

Berbagai penelitian telah melakukan hal serupa dalam menguji efektivitas model dalam analisis sentimen. Penelitian dari Azmi et al. (2020) menganalisis sentimen komentar Instagram terkait objek wisata di Lombok dengan menggunakan teknik *Naïve Bayes Classifier*. (Finandra & Hamami, 2021) turut melakukan penelitian yang sama dengan Azmi et. al. (2020) tetapi mengganti objeknya dengan komentar mengenai objek wisata di Bandung. Adapun penelitian dari Putu et al. (2021) menganalisis komentar di Facebook dan Instagram mengenai objek wisata di Malang menggunakan *Naïve Bayes Classifier* dan *Support Vector Machine*. Melalui penelitian-penelitian terdahulu maka penelitian ini ingin memberikan kebaruan dalam menganalisis secara khusus opini netizen di Instagram mengenai objek pariwisata di Yogyakarta. Dari penelitian – penelitian tersebut,

diketahui bahwa tingkat *Naïve Bayes Classifier* akurasi cukup tinggi. Bahkan pada penelitian yang membandingkan *Naïve Bayes Classifier* dengan teknik lainnya, didapat tingkat akurasi tertinggi didapat saat menggunakan *Naïve Bayes Classifier.* Pertimbangan tersebut yang melatarbelakangi digunakannya *Naïve Bayes Classifier* pada penelitian ini.

Tujuan penelitian ini adalah untuk melakukan analisis sentimen terhadap opini netizen Instagram terhadap objek wisata di D.I. Yogyakarta menggunakan teknik *Naïve Bayes Classifier*. Penentuan model analisis sentimen dengan performa yang tinggi dapat memberikan informasi yang tepat dalam sebagai referensi untuk membuat kebijakan. Adapun melalui penelitian diharapkan dapat diperoleh model data mengenai opini netizen di Instagram sehingga dapat digunakan untuk memodelkan sentimen pada dokumen-dokumen opini objek wisata di D.I. Yogyakarta lainnya. Penelitian ini diharapkan dapat memberikan sebuah referensi bagi pemangku kepentingan dalam mengevaluasi pengelolaan objek wisata di Yogyakarta.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, maka dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana hasil analisis sentimen dari objek data penelitian?
2. Bagaimana evaluasi dari klasifikasi yang diperoleh dengan menggunakan metode *Naive bayes Classifier*?
3. Apa saja istilah-istilah yang paling banyak disampaikan pada masing-masing kelas sentimen?

1.3. Batasan Masalah

Adapun batasan masalah dalam melakukan penelitian ini yaitu :

1. Aplikasi yang digunakan sebagai media promosi online adalah instagram.
2. Periode data yang digunakan terhitung dari 1 Agustus 2022 sampai dengan 31 Agustus 2023.
3. Sumber data penelitian menggunakan lima akun Instagram mengenai informasi wisata Yogyakarta, yaitu @explorejogja, @jogjaviral, @wisatajogja_, @pariwisata.jogjakarta, @wonderfuljogja.

1.4. Tujuan Penelitian

Adapun tujuan dari penelitian ini yang diperoleh berdasarkan rumusan masalah yaitu untuk

1. Mengetahui bagaimana analisis sentimen dari objek data penelitian.
2. Mengetahui konten seperti apa yang paling memikat atensi pengguna instagram dalam mempromosikan wisata Yogyakarta secara online.
3. Mengetahui bagaimana hasil klasifikasi yang diperoleh dengan menggunakan metode *Naive bayes Classifier.*

1.5. Manfaat Penelitian

Adapun dalam penelitian ini terdapat beberapa manfaat adalah sebagai berikut :

1. Untuk Peneliti
   a. Sebagai pengimplementasian keilmuan Statistika, khususnya dalam melakukan analisis sentimen menggunakan pendekatan *Naïve Bayes Classifier.*
   b. Bermanfaat bagi pihak – pihak terkait secara tidak langsung, dengan menyimpulkan hasil analisis penelitian.
2. Untuk Pengelola Wisata

   Mampu mengoptimalkan pengelolaan lokasi wisata melalui pemanfaatan menggunakan model yang mampu memberikan hasil akurasi analisis sentimen yang baik sehingga mengetahui sudut pandang warganet dengan tepat dalam menilai objek – objek wisata.
3. Untuk Pengelola Akun Instagram Wisata Yogyakarta

   Mampu mengoptimalkan pengelolaan akun instagram terkait Wisata Yogyakarta karena mengetahui sudut pandang warganet dalam menilai objek – objek wisata.
4. Untuk Masyarakat

   Dengan adanya penelitian ini, masyarakat dapat memahami bagaimana opini warganet terhadap wisata di D.I Yogyakarta dan juga menjadi wawasan untuk penelitian-penelitian selanjutnya.

BAB II
TINJAUAN PUSTAKA

Bagian ini menyampaikan beberapa penelitian terdahulu yang memiliki kemiripan terkait analisis sentimen terhadap opini terhadap suatu objek di media sosial. Melalui penyampaian penelitian-penelitian sebelumnya maka dapat ditentukan *gap* yang dapat diisi dari penelitian ini. Tabel 2.1. menyajikan berbagai penelitian yang sebelumnya sudah terbit dan memiliki kemiripan dengan penelitian ini.

Tabel 2.1 Tabel Penelitian Sebelumnya

Tahun	Nama	Judul	Hasil Penelitian
2020	Muhammad Azmi, Amiruddin Khairul Huda, dan Arief Setiyanto	Pemanfaatan Data Instagram Untuk Mengetahui Reputasi Tempat Wisata Di Lombok	Penelitian ini melakukan analisis klasifikasi terhadap sentimen terhadap komentar di Instagram mengenai pariwisata di kota Lombok. Adapun yang menjadi kata kunci dalam pengambilan data adalah pantai, gili dan gunung. Melalui kata kunci tersebut diperoleh 600 data. Metode yang digunakan dalam penelitian ini adalah Naïve Bayes Classifier. Terdapat tiga klasifikasi dalam penelitian ini yakni positif, netral dan negatif. Hasil yang diperoleh dari penelitian ini menunjukkan bahwa akurasi dari Naïve bayes adalah 59%
2020	Gunawan Budi Prasetyo, Habibie Ed Dien, Dikta Afif Rahman Prasetyo	Sentimen Analisis Terhadap Objek Wisata Alam Kota Malang Di Instagram Dan Facebook Menggunakan Metode Naive Bayes dan Support Vector Machine	Penelitian ini melakukan klasifikasi komentar Instagram dan Facebook mengenai wisata di Kota Malang. Terdapat pemisahan analisis antara data dari Instagram dan juga Facebook. Melalui hasil scraping data diperoleh data Facebook sebanyak 913 baris dan Instagram sebanyak 998 baris. Jika dibandingkan dengan

Tahun	Nama	Judul	Hasil Penelitian
			Support Vector Machine. Naïve Bayes Classifier memberikan hasil akurasi yang lebih tinggi dalam mengklasifikasikan sentimen baik pada Instagram (75%) maupun pada Facebook (60%)
2021	Sandy Finandra, Murahartawaty ,Faqih Hamami	Penerapan Analisis Sentimen Melalui Data Instagram Untuk Mengetahui Reputasi Wisata Kuliner Di Kota Bandung Menggunakan metode Klasifikasi Naïve Bayes	Penelitian ini bertujuan untuk melakukan klasilifasi komentar instagram mengenai Pariwisata di Kota Bandung. Terdapat 34 akun mengenai wisata Kota Bandung yang dianalisis dalam penelitian ini. Secara keseluruhan terdapat 864 data yang diperoleh dari penelitian ini. Selain menggunakan teknik TF-IDF, penelitian ini menggunakan metode *simple additive weighting* (SAW) untuk mendukung keputusan klasifikasi. Melalui perbandingan berbagai teknik klasifikasi, teknik Naïve Bayes Classifier memberikan nilai akurasi terbesar yakni 68.15%.
2021	Hidayatus Sibyan, Nur Hasanah	Analisis Sentimen Pada Wisata Dieng Dengan Algoritma K-Nearest Neighbor (K-Nn)	Penelitian ini bertujuan untuk melakukan klasifikasi terhadap ulasan pengunjung wisata Dieng Jawa Tengah melalui analisis sentimen. Terdapat 147 ulasan yang diambil dari internet sebagai data yang akan dianalisis. Metode yang digunakan untuk mengklasifikasikan sentimen adalah K-Nearest Neighbor (K-NN). Pada penelitian ini, komentar akan diklasifikasikan menjadi 2 kelas yakni positif dan negatif. Melalui metode K-NN diperoleh akurasi dari

Tahun	Nama	Judul	Hasil Penelitian
			hasil klasifikasi yang cukup tinggi yakni 86%.
2023	Doni Arya Utama	Analisis Sentimen Tempat Wisata Kabupaten Gunungkidul Menggunakan Metode Naive Bayes Berdasarkan Kritik Saran Wisatawan	Penelitian ini bertujuan untuk mengklasifikasikan komentar-komentar para netizen mengenai pariwisata di Gunung Kidul Yogyakarta di situs web *Trip Advisor* dan juga *Google Maps* pada tahun 2022-2023. Melalui teknik *scraping*, terdapat 500 baris data yang berhasil diambil dari penelitian ini. Metode klasifikasi yang digunakan dalam penelitian ini adalah Naïve Bayes Classifier. Penelitian ini membagi analisis menjadi 3 bagian yakni perbandingan data training dan testing 80-20, 70-30 dan 60-40. Akurasi yang didapatkan cenderung sama yakni berada di antara 79 hingga 80% pada masing-masing pengujian.

Tabel 2.1. menyajikan berbagai penelitian terdahulu mengenai analisis sentimen dengan berbagai macam algoritma *machine learning*. Penelitian yang telah dilakukan mengambil konteks-konteks tertentu, baik dalam sektor politik, bisnis maupun pariwisata. Terdapat perbedaan pada penelitian ini dengan penelitian-penelitian sebelumnya yakni menggunakan data mengenai wisata di D.I Yogyakarta pada tahun 2022-2023. Adapun pemilihan yogyakarta sebagai fokus analisis karena Yogyakarta merupakan salah satu provinsi di Indonesia yang dikenal memiliki banyak lokasi wisata dan ragam jenis wisata yang patut dijelajahi.

BAB III
LANDASAN TEORI

3.1. Promosi Wisata di Provinsi D.I. Yogyakarta

Kegiatan berwisata merupakan suatu aktivitas berpegian dimana seseorang maupun suatu kelompok menuju ke suatu tempat di luar tempat tinggalnya, baik didasari pada kepentingan berlibur, pendidikan, ekonomi, sosial, budaya maupun hanya untuk mengeksplorasi (Riswandha et al., 2017). Berwisata dapat dilakukan dengan berbagai macam moda transportasi baik berjalan, bersepeda, mengendarai mobil, menggunakan pesawat terbang dan lain sebagainya. Melalui berwisata, seseorang memiliki peluang untuk mendapatkan pengalaman baru, melihat pemandangan baru, mencoba masakan baru serta memperluas perspektif dari orang-orang di sekitar daerah wisata. Berwisata juga dapat dilakukan menuju situs bersejarah, melakukan aktivitas di alam ataupun sekedar melakukan relaksasi di luar ruangan.

Daerah Istimewa Yogyakarta merupakan salah satu provinsi Provinsi di Indonesia yang memiliki daya tarik wisata yang cukup besar. Hasil analisis dari Kurniawan (2010) menunjukkan bahwa D.I. Yogyakarta memiliki potensi wisata alam seperti pantai; wisata belanja di pasar dan *shopping mall*; wisata sejarah yang terdiri dari keraton kesultanan, museum dan candi; wisata budaya seperti tari-tarian dan teater ramayana, wisata kuliner yang tersebar di seluruh daerah serta wisata ziarah.. Dalam hasil analisisnya, Wahyuni (2020) turut mengidentifikasi berbagai macam objek wisata unggulan di D.I. Yogyakarta, pada Kabupaten Kulon Progo terdapat Desa Wisata, Wisata Pantai serta kerajinan tangan; pada Kabupaten Bantul, Desa Budaya, Wisata Pantai dan Cagar Budaya menjadi unggulan; Pada Kabupaten Gunungkidul Wisata Alam seperti pantai, dan goa menjadi unggulan; Pada Kabupaten Sleman, wisata candi dan museum; dan pada Kota Yogyakarta, wisata Museum, kuliner dan kampung wisata menjadi unggulan. Berbagai potensi wisata D.I. Yogyakarta tersebut menjadi aset serta daya tarik sehingga memberikan manfaat positif bagi D.I. Yogyakarta di seriap aspek, baik ekonomi, sosial, lingkungan maupun budaya.

Saat ini potensi wisata di D.I. Yogyakarta turut disampaikan dalam media sosial sehingga setiap orang dapat memperoleh informasi dengan mudah dan juga dapat menjadi sarana promosi agar banyak wisatawan yang akan mengunjungi Yogyakarta. Dalam penelitiannya, Wijayanti (2021) menyampaikan bahwa promosi wisata Yogyakarta menggunakan Instagram dari Dinas Pariwisata DIY secara signifikan dapat meningkatkan minat kunjungan wisatawan. Hal yang sama dikemukakan oleh Aji & Andadari (2021) dimana konten instagram Dinas Pariwisata Kota Yogyakarta @pariwisata.jogjakota mampu meningkatkan minat berkunjung wisatawan secara signifikan. Anggiyani (2021) turut menyampaikan bahwa akun Instagram @explore.bantul memiliki potensi yang sangat besar dalam membantu promosi daya tarik wisata yang berada di Kabupaten Bantul, terutama karena jumlah *followers* yang besar. Berbagai temuan tersebut menunjukkan bahwa informasi di media sosial Instagram telah mempu menjadi referensi masyarakat untuk berkunjung.

Tidak hanya konten-konten dari suatu akun saja tetapi juga komentar-komentar, opini maupun tinjauan dari para netizen turut menjadi informasi bagi calon wisatawan D.I. Yogyakarta. Jumlah komentar di konten Instagram dapat menunjukkan popularitas suatu objek wisata dan juga komentar di media sosial dapat berhubungan dengan bagaimana persepsi suatu brand apakah disukai konsumen atau tidak (Fithriya, 2020). (Wijayanti, 2021) turut menjelaskan bahwa komentar positif di media sosial dapat memengaruhi minat pembelian maupun berkunjung ke suatu destinasi dan komentar negatif berdampak pada berkurangnya minat kunjungan wisatawan. Adapun (Dwi & Kurniawati, 2016) turut menjelaskan bahwa berbagai wisatawan cukup sering berbagi komentar mengenai pengalamannya berwisata di akun Instagram yang menyediakan informasi wisata di suatu daerah. Berbagai komentar yang berada di kolom komentar suatu media sosial sering dinamakan sebagai *electronic Word of Mouth* (E-WOM) dan E-WOM memiliki dampak yang signifikan terkait bagaimana perasaan seseorang terhadap suatu destinasi pariwisata (Utama & Giantari, 2020). Komentar-komentar yang tersedia di media sosial mengenai pariwisata D.I. Yogyakarta dapat menjadi referensi bagi pemangku kepentingan terkait untuk mengevaluasi opini maupun pengalaman dari masyarakat.

3.2. Konsep Media Sosial

Menurut Robbins & Singer (2014) media sosial merujuk pada sebuah teknologi yang memfasilitasi diseminasi dan penyebaran informasi melalui internet. Media sosial dapat digunakan untuk menulis sebuah *long-form writing* atau blogging seperti Wordpress, Tumblr, Blogger; *short-form writing* atau *microblogging* seperti Facebook, dan Twitter; berbagi gambar seperti Instagram, atau Snapchat; serta berbagi video seperti Youtube. Media sosial sendiri mendukung konsep *User Generated Content* sehingga setiap penggunanya dapat memproduksi pesan atau konten. Seluruh konten yang dibuat oleh pengguna media sosial, baik berupa teks, gambar, maupun suara menjadi salah satu data yang dapat digunakan untuk dianalisis. Jumlah data yang besar dari media sosial dapat memberikan sebuah informasi bagi pihak-pihak yang membutuhkan (Renjith et al., 2022).

3.2.1 Instagram

Instagram merupakan sebuah layanan jejaring sosial/media sosial untuk berbagi konten yang berformat grafis atau foto dan video yang dimiliki oleh perusahaan Amerika, Meta Platforms. Melalui aplikasi ini, pengguna dapat mengunggah konten, melakukan *editing,* memberikan komentar dan lain sebagainya. Suatu unggah dapat disampaikan secara publik ataupun *private* sesuai dengan pengaturan yang ditentukan. Adapun media sosial ini dapat diakses melalui perangkat *mobile,* baik Android maupun iOS, sehingga memudahkan setiap penggunanya untuk mengakses layanan tersebut. Instagram diluncurkan pada bulan Oktober 2010 oleh Kevin Systrom dan Mike Krieger. Dalam dua bulan Instagram telah mendapatkan satu juta pengguna dan terus berkembang hingga 10 miliar pengguna di bulan Juni 2018. (Wikipedia, 2023)

Instagram adalah salah satu media sosial yang populer di Indonesia, seperti dikutip dari databoks tercatat sebanyak 104.8 juta pengguna Instagram di Indonesia pada bulan Oktober 2023 (Databoks, 2023). Indonesia menjadi negara keempat terbesar pengguna Instagram di bawah Brasil dan Amerika Serikat. Pengguna Instagram di Indonesia sendiri didominasi oleh kelompok usia 18-24 dan 25-34. Adapun Instagram adalah aplikasi terpopuler yang menempati urutan ke-4 di Indonesia setelah Facebook, Youtube dan Whatsapp. Sebagai salah satu media

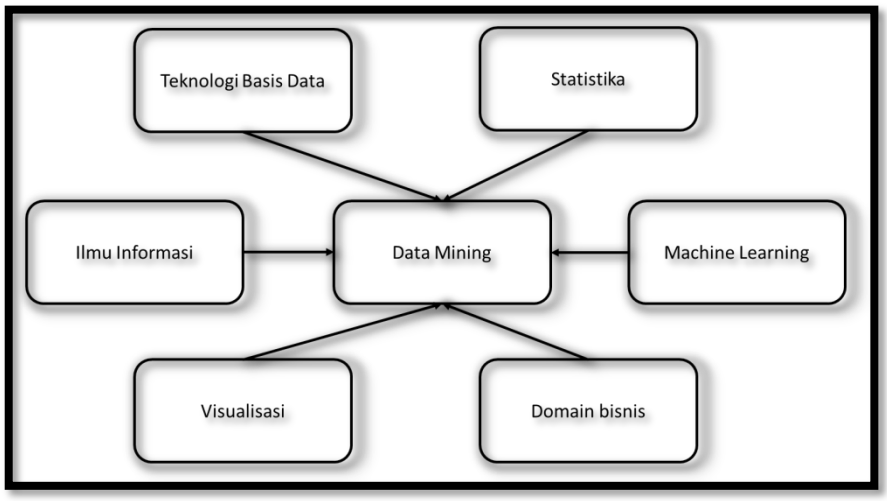
sosial, Instagram juga bisa menjadi sarana untuk bercerita, berbagi konten berbasis grafis dan mendukung interaktivitas terhadap sesama penggunanya.

3.3. *Web Scraping*

Web scraping merupakan salah satu teknik dalam pengumpulan data. Melalui teknik ini, berbagai data di suatu situs web dapat diambil secara otomatis tanpa harus disalin secara manual. Adapun tujuan dari *web scraping* ini adalah untuk mencari sebuah informasi tertentu dalam sebuah situs. Berbeda dengan teknik *web crawling* yang ditujukan untuk mengambil data dari seluruh aspek yang berada pada sebuah website, teknik *web scraping* hanya melakukan ekstraksi terhadap data tertentu di suatu situs yang dibutuhkan. Data yang diambil dalam suatu website dapat berupa teks maupun gambar. Hasil ekstraksi data dari teknik *web scraping* dapat digunakan kembali untuk dianalisis maupun dimodelkan lebih lanjut (Satriajati et al., 2020).

3.4. *Data Mining*

Data mining merujuk sebuah penambangan atau penemuan informasi baru terkait pola ataupun tren yang tersembunyi dan yang belum diketahui sebelumnya di dalam sebuah data. Proses data mining sendiri dikaitkan dengan analisis data dalam jumlah atau volume yang besar dan membutuhkan komputer dengan performa yang tinggi dalam analisisnya (Zaki, 2002). *Data mining* dapat mengganti aktivitas manual atau analisis tradisional yang memakan waktu dengan proses yang terotomatisasi dimana performa teknik tersebut dapat terus ditingkatkan dengan terus mencari ketidakteraturan dalam sebuah data. Teknik ini bertujuan untuk mengubah data mentah/*raw data* yang memiliki volume besar menjadi sebuah informasi yang dimengerti oleh manusia. Informasi yang diperoleh dari sebuah proses *data mining* digunakan untuk mendukung perumusan keputusan pada sebuah *domain* bisnis tertentu. Data mining sendiri dapat digunakan pada berbagai industri maupun disiplin ilmu baik dari keuangan, kesehatan, pemasaran maupun pada bidang komunikasi.



Gambar 3.1 Data Mining Sebagai Multi Disiplin

Sumber :Mehra et al. (2014)

Data Mining merupakan sebuah teknik pengolahan data yang multi disiplin. Mehra et al. (2014) menyampaikan setidaknya terdapat enam aspek dalam *data mining* yakni teknologi basis data, statistika, machine learning, ilmu informasi, visualisasi dan domain bisnis. Statistika memiliki peran penting untuk mendukung pengujian hipotesis maupun estimasi probabilitas. Machine learning sendiri merupakan bagian dari kecerdasan buatan yang berfokus pada pengembangan algoritma untuk mencari pola dalam data. Teknologi basis data merupakan bagian penting karena berperan untuk mengatur dan mengekstrak data dari sebuah basis data. Aspek visualisasi dibutuhkan untuk dapat menampilkan pola yang diperoleh melalui gambaran yang lebih mudah dimengerti oleh individu non teknis. Domain bisnis merupakan aspek mengenai pemahaman seseorang terhadap suatu ilmu tertentu. Adapun aspek ilmu informasi mendukung proses manajemen data yang besar dan juga cara mendapatkan data yang lebih efisien.

3.5. *Machine Learning*

Machine learning merujuk pada sebuah algoritma yang mampu mencari pola di dalam data dan mampu mengenalinya kembali dalam sampel yang baru (Delgadillo, 2021). Dasar dari *machine learning* antara lain adalah teknik statistika seperti korelasi, regresi atau inferensi bayesian. Contoh dari *machine learning* antara lain adalah model linear (*k-nearest neighbors, support vector machine*), *decision trees* (*random forest, xgboost*), model bayesian (*naïve bayes, bayesian*

network analysis), jaringan syaraf tiruan, dan lain sebagainya. Tidak seperti paradigma pengujian hipotesis pada umumnya, pendekatan machine learning cenderung tidak terikat pada sebuah teori dan sering digunakan tanpa sebuah hubungan yang diperkirakan sebelumnya. Dalam mengevaluasi performa dari sebuah model yang dibangun dari *machine learning*, suatu dataset akan dipecah menjadi data *training* dan data *testing*. Data *training* digunakan untuk membangun sebuah model dan data *testing* digunakan untuk mengevaluasi model yang telah dibuat sebelumnya. Perbandingan antara data *training* dan data *testing* dapat dibentuk antara 80:20, 70:30 maupun 60:40. *Machine learning* sendiri dibagi atas dua topik yakni *supervised learning* dan *unsupervised learning*.

3.5.1 *Supervised Learning*

Supervised learning bertujuan untuk memecahkan permasalahan prediksi. Salah satu topik *machine learning* ini digunakan untuk mengetahui sebuah hubungan atau pola tersembunyi antara atribut *input* dengan atribut *output*. Dalam jenis *machine learning* ini suatu algoritma akan dilatih pada sebuah dataset yang sudah diberikan label, dalam artian dimana sebuah data *input* akan disandingkan dengan label *output*. Hasil dari *supervised learning* adalah untuk memahami sebuah fungsi pemetaan dari *input* terhadap *output* sehingga suatu algoritma dapat membentuk prediksi ataupun klasifikasi yang akurat ketika dihadapkan kembali pada data yang terbaru. *Supervised learning* telah banyak digunakan dalam lingkup *natural language processing*.

3.5.2 *Unsupervised Learning*

Unsupervised learning merupakan salah satu jenis *machine learning* dimana penggunaan algoritma ditujukan untuk mengidentifikasi pola dalam suatu kumpulan data yang sebelumnya tidak diklasifikasikan atau diberikan label *output*. Dengan kata lain, suatu algoritma akan mengeksplorasi data tanpa diberikan label yang spesifik untuk memprediksi. Tujuan utama dari *unsupervised learning* adalah untuk mencari pola yang tersembunyi, hubungan atau struktur yang membentuk sebuah data. Jenis pembelajaran yang sering digunakan antara lain adalah *clustering* dan *dimension reduction*. Pada model ini tidak dapat dihitung akurasi seperti pada pemodelan *supervised learning* karena tidak ada atribut target yang menjadi acuan dalam pemodelan.

3.6. *Text mining*

Teknik *text mining* mengacu pada sebuah metode ekstraksi informasi yang komprehensif dan bernilai dari sebuah badan teks (Gao et al., 2020). *Text mining* merupakan bagian dari *data mining* dimana objek yang dianalisis adalah dokumen teks yang terstruktur maupun semi-terstruktur. Data tidak struktur yang dapat dianalisis dalam *text mining* merujuk pada suatu informasi tekstual yang tidak memiliki model data yang sebelumnya ditentukan seperti email, konten media sosial, artikel dan lain sebagainya. Adapun *text mining* ini juga dikenal sebagai *text analytics* atau *natural language processing/*NLP. NLP sendiri merupakan sebuah algoritma yang memungkinkan suatu mesin untuk mengenai pola bahasa manusia dan memprosesnya secara otomatis. Melalui *text mining* ini, suatu data yang tidak terstruktur dapat diolah menjadi sebuah data yang terstruktur sehingga dapat dimodelkan bagaimana pola yang terbentuk di dalamnya. Dalam *text mining* tidak hanya digunakan untuk mengeksplorasi istilah-istilah apa saja yang muncul di dalam kumpulan dokumen tetapi juga mengeksplorasi bagaimana hubungan tekstual antar dokumen yang dianalisis. *Text mining* ini dapat digunakan pada berbagai jenis industri seperti bisnis, kesehatan, keuangan, maupun media. Melalui teknik ini maka suatu organisasi dapat memperoleh informasi yang dapat digunakan untuk menenutkan sebuah keputusan. Adapun melalui *text mining* ini seseorang dapat memperoleh pola-pola kata yang tersembunyi dari sebuah dokumen yang berjumlah besar.

3.6.1 *Text Preprocessing*

Text processing merujuk pada sebuah analisis dan manipulasi dari suatu teks untuk menjadi informasi inti (Rani & Jain, 2023). Proses manipulasi teks pada *text processing* perlu menggunakan sebuah teknik komputasi. Teknik ini bertujuan untuk mengambil informasi yang bernilai, pola dan juga *insight* dalam sebuah dokumen yang berjenis teks. Adapun *text processing* merupakan salah satu aspek fundamental dalam NLP yang merupakan cabang dari *artificial intelligence* dimana berfokus pada interaksi antara komputer dengan bahasa manusia. Berikut adalah beberapa langkah umum yang terlibat dalam preprocessing teks:

1. *Cleaning data*

Cleaning data merupakan sebuah proses dalam mempersiapkan dan membersihkan data teks yang mentah sehingga kualitas, konsistensi, kecocokan untuk analisisnya meningkat. Suatu data teks biasanya sering mengandung *noise*, informasi yang tidak relevan dan penyimpangan sehingga dapat menghalangi pembuatan model yang efektif. Tahap ini terdiri atas eliminasi dari tanda baca seperti koma, titik, tanda seru, tanda tanya, dan karakter lainnya selain alfabet.

2. *Case folding*

Case folding adalah sebuah bagian dari *text processing* yang mengubah semua karakter menjadi sebuah bentuk yang terstandar, bentuk yang digunakan secara umum adalah *lowercase* atau mengubah seluruh karakter menjadi huruf kecil. Tujuan dari *case folding* ini adalah untuk memfasilitasi perbandingan dan analisis dari sebuah teks dengan tidak bergantung pada bentuk awal teks yang dimasukan. Teknik ini bermanfaat bagi pengolahan teks karena dapat mencari, menyortir dan membandingkan sebuah teks yang tidak relevan.

3. *Filtering/Stopword*

Filtering merupakan suatu tahap dalam *text processing* untuk menghapus sekumpulan kata yang tidak memiliki arti substantif yang signifikan pada sebuah analisis. Adapun sekumpulan kata yang dimaksud adalah *stopwords* atau kata-kata yang sering muncul dalam sebuah teks tetapi memiliki arti yang tidak dapat berkontribusi pada pemahaman sebuah teks. Contoh dari *stopwords* antara lain “dan”, “pada”, “dalam” dan lain sebagainya. Tujuan dari *filtering* ini adalah untuk hanya melibatkan kata-kata yang bermakna saja dalam sebuah analisis.

4. *Stemming*

Stemming merupakan sebuah teknik *text processing* untuk mengubah sebuah kata yang memiliki imbuhan atau partikel menjadi kata dasarnya. Tujuan dari proses ini adalah untuk menormalisasi huruf sehingga berbagai variasi dari huruf yang sama akan diperlakukan sebagai sebuah entitas yang sama. Contoh dari proses *stemming* dalam Bahasa Indonesia adalah mengubah kata “membeli” menjadi “beli”. Algoritma dari *stemming* harus mengikuti bahasa dari teks yang akan diproses.

5. *Tokenizing*

Tokenizing merupakan salah satu teknik *text processing* yang digunakan untuk memecah teks menjadi unit individual. Dalam NLP atau *text mining*, proses ini merupakan proses fundamental agar komputer dapat memahami data teks yang lebih terstruktur. Contoh dari proses ini adalah memecah kalimat "Yogyakarta memiliki banyak objek wisata yang indah" menjadi "Yogyakarta", "memiliki", "banyak", "objek", "wisata", "yang", "indah".

3.7. Analisis Sentimen

Analisis sentimen/*sentiment analysis* merupakan suatu teknik komputasi dan *natural language processing* untuk mengidentifikasi, mengekstrak atau mengkarakterisasi informasi subjektif seperti opini yang diekspresikan melalui sebuah teks (Beigi et al., 2016). Tujuan utama dari analisis ini adalah untuk mengklasifikasikan suatu opini yang disampaikan terhadap kategori positif, negatif maupun netral. Aplikasi dari analisis sentimen dapat digunakan secara luas diantaranya analisis dari masukan pelanggan, pengawasan media sosial, tinjauan produk dan analisis sentimen terhadap suatu produk. Melalui teknik ini maka dapat diperoleh informasi bernilai bagaimana perasaan seseorang terhadap suatu objek. Adapun terkait dengan sumber datanya, analisis sentimen dibagi menjadi dua jenis yakni analisis tingkat dokumen, analisis tingkat kalimat dan tingkat aspek (Behdenna et al., 2018).

1. Analisis sentimen tingkat dokumen

   Analisis pada tingkat dokumen ditujukan untuk menentukan keseluruhan opini pada semua dokumen. Analisis sentimen pada tingkat ini mengasumsikan bahwa setiap dokumen mengekspresikan opini pada suatu entitas tunggal.

2. Analisis sentimen tingkat kalimat

   Analisis tingkat kalimat bertujuan untuk menentukan apakah setiap kalimat mengekspresikan sebuah opini. Tingkat ini membedakan antara kalimat objektif yang menyatakan informasi faktual dan kalimat subjektif yang menyatakan pendapat. Dalam kasus ini terdapat dua perlakuan, pertama kenali apakah kalimat tersebut telah menyatakan kalimat atau tidak, kemudian menilai polaritas dari kalimat tersebut.

3. Analisis sentimen level aspek

Analisis sentimen ini menentukan sentimen terhadap aspek tertentu atau fitur yang diberikan seperti produk, layanan atau topik. Pada tingkat ini, opini ditandai dengan adanya polaritas dan sasaran opini. Dalam hal ini, terdapat dua perlakuan, pertama mengidentifikasi entitas dan aspek dari entitas yang dimaksud kemudian menilai pendapat pada setiap aspek.

Salah satu teknik yang digunakan dalam analisis sentimen adalah VADER (Valence Aware Dictionary and sentiment Reasoner). Teknik tersebut adalah teknik yang alat sentimen analisis yang berbasiskan lexicon dan digunakan untuk menganalisis sentimen baik dalam tingkat dokumen atau kalimat. Teknik ini dikembangkan untuk mengatasi sentimen yang disampaikan melalui teks media sosial atau sebuah kumpulan teks yang sering ditemukan istilah-istilah informal. Pendekatan ini akan menjadikan kalimat sebagai input, memberikan evaluasi persentase terhadap aspek positif, negatif, netral, dan juga compound (polaritas keseluruhan kalimat). Suatu kalimat masuk ke dalam sentimen tertentu dengan menjumlahkan *compound score* atau sebuah pengukuran yang merepresentasikan sentimen dimana dapat bernilai -1 atau memiliki sentimen negatif dan 1 untuk sentimen positif. (Elbagir & Yang, 2019)

3.8. *Term Frequency – Inverse Document Frequency (TF-IDF)*

TF-IDF merupakan sebuah angka statistik yang digunakan dalam analisis teks untuk mengevaluasi tingkat kepentingan suatu kata yang relatif terhadap kumpulan dokumen. TF-IDF bekerja dengan menentukan frekuensi relatif kata dalam dokumen tertentu dibandingkan dengan proporsi kebalikan dari kata tersebut di seluruh korpus dokumen (Ramos, 2003). Secara intuitif, perhitungan ini menentukan seberapa relevan suatu kategori tertentu dalam dokumen tertentu. Kata-kata yang umum dalam satu atau sekelompok kecil dokumen cenderung memiliki angka TF-IDF yang lebih tinggi dibandingkan kata-kata umum seperti artikel. TF-IDF sendiri secara umum diciptakan untuk keperluan pencarian dokumen dan pengambilan informasi.

Pada Term Frequency (TF), terdapat beberapa jenis formula yang dapat digunakan:

1. TF biner (binary TF) yakni memperhatikan keberadaan atau ketiadaan suatu kata dalam dokumen. TF ini merupakan bentuk TF yang lebih sederhana

dari TF murni, dimana jika dalam suatu dokumen terdapat istilah tertentu maka nilainya menjadi satu (1), jika tidak, nilainya nol (0).

2. TF murni (raw TF) memberi nilai TF berdasarkan frekuensi kemunculan kata dalam dokumen. Secara umum TF dimaksudkan untuk mengukur seberapa sering suatu istilah atau kata muncul di dalam sebuah dokumen. Contohnya, jika kata itu muncul lima kali, nilainya menjadi lima (5).
3. TF normalisasi menggunakan perbandingan antara frekuensi term dengan nilai tertinggi dalam seluruh frekuensi term di dokumen. Ide utama dari TF Normalisasi adalah untuk mencegah dokumen yang lebih panjang memiliki nilai TF yang lebih besar karena panjangnya dokumen.
4. TF logaritmik digunakan untuk menghindari dominasi dokumen dengan sedikit term dalam kueri tetapi memiliki frekuensi yang tinggi. Secara umum dokumen yang lebih panjang akan memiliki nilai TF yang besar sehingga dapat memunculkan sebuah representasi yang bias terhadap istilah yang penting.

Nilai TF diperoleh dengan menggunakan persamaan 3.1 (Ramos, 2003)

$$TF: \frac{1}{f_{t,d}} \qquad (3.1)$$

Di mana,

$f_{t,d}$: frekuensi kata (t) muncul di dokumen *d*

Inverse Document Frequency (IDF) adalah sebuah pengukuran yang digunakan dalam analisis teks dan pengumpulan informasi untuk mengevaluasi tingkat kepentingan suatu istilah dalam sebuah kumpulan dokumen tertentu. Tujuan dari IDF adalah mengidentifikasi istilah yang relatif jarang, lebih informatif maupun uni dalam seluruh korpus. Adapun IDF berfungsi mengurangi bobot suatu term jika kemunculannya banyak tersebar di seluruh dokumen. Perhitungan IDF diperlukan, karena jika hanya menggunakan TF saja dikhawatirkan akan muncul kata umum yang akan dominan, yang seharusnya kata tersebut dihilangkan. Metode IDF merupakan sebuah perhitungan dari bagaimana term didistribusikan secara luas pada koleksi dokumen yang bersangkutan. Untuk menghitung nilai IDF dari sebuah kata dalam kumpulan dokumen menggunakan persamaan 3.2.

$$IDF: \log \frac{N}{f_{t,d}} \qquad (3.2)$$

Di mana,

N : jumlah keseluruhan dokumen

$f_{t,d}$: jumlah dokumen D yang mengandung t

Maka rumus umum untuk TF-IDF adalah gabungan formula TF dan IDF yaitu dengan mengalikan TF dan IDF seperti persamaan 3.3

$$TF - IDF(t) = TF(t).IDF(t) \quad (3.3)$$

3.9. *Confusion matrix*

Untuk secara efektif mengevaluasi performa dari suatu model, maka diperlukan sebuah pengukuran performa untuk menentukan kualitas dari model *machine learning*. Salah satu alat pengukuran performa yang dapat digunakan adalah *confusion matrix*. Pengukuran tersebut merujuk pada sebuah tabel yang digunakan untuk menggambarkan rangkuman dari pengujian performa sebuah model klasifikasi. Pada garis diagonal tabel tersebut adalah letak dimana algoritma memberikan hasil yang benar. *Confusion martrix* berguna dalam permasalahan klasifikasi yang melibatkan kelas biner maupun *multi-class.* Dalam sebuah *confusion matrix* terdapat empat buah aspek yang digunakan sebagai pengukuran evaluasi yakni *true positive, true negative, false positive dan false negative:*

1. *True Positive* (TP): Merupakan jumlah sampel yang secara benar diklasifikasikan sebagai positif oleh model.
2. *True Negative* (TN): Merupakan jumlah sampel yang secara benar diklasifikasikan sebagai negatif oleh model.
3. *False Positive* (FP): Merupakan jumlah sampel yang keliru diklasifikasikan sebagai positif oleh model.
4. *False Negative* (FN): Merupakan jumlah sampel yang keliru diklasifikasikan sebagai negatif oleh model.

Tabel 3.1 Contoh *Matrix Confusion*

Predict values	***Actual value***	
	Positif	Negatif
Positif	TP	FP
Negatif	FN	TN

Recall, Precision, dan Accuracy merupakan beberapa metrik evaluasi yang dapat digunakan untuk mengukur kinerja model hasil klasifikasi. Adapun ketiga pengukuran tersebut dapat diperoleh dari aspek-aspek yang berada di dalam *confusion matrix*.

1. *Recall* mengukur seberapa baik model mampu mengidentifikasi semua *instance* positif yang sebenarnya. Pengukuran ini diperoleh dengan membandingkan antara jumlah positif yang diprediksi dengan benar (*True Positive*) dibagi dengan jumlah semua instance positif yang sebenarnya (*True Positive + False Negative*).
2. *Precision* mengukur seberapa baik model mampu mengidentifikasi *instance* yang diprediksi positif dengan benar. Pengukuran ini diperoleh dengan membandingkan jumlah positif yang diprediksi dengan benar (*True Positive*) dibagi dengan jumlah semua instance yang diprediksi positif (*True Positive + False Positive*).
3. *Accuracy* mengukur seberapa baik model mampu memprediksi secara keseluruhan dengan benar. Pengukuran ini diperoleh dengan cara membandingan antara jumlah prediksi benar (True Positives + True Negatives) dibagi dengan jumlah total instance.

3.10. *Naïve Bayes Classifier*

Naïve Bayes Classifier merupakan teknik klasifikasi berdasarkan teorema bayes dengan menggunakan metode probabilitas (Leong & M, 2020). Klasifikasi naïve bayes memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya. Naïve Bayes adalah algoritma yang mengasumsikan bahwa setiap fitur atau variabel tidak bergantung satu dengan yang lainnya. Dengan menggunakan kaidah *naïve bayes*, maka didapatkan formulasi sebagai berikut :

$$P(A|B) : \frac{P(B|A)\,P(A)}{P(B)} \qquad (3.4)$$

Keterangan :

A : Data

B : Kamus positif/negatif (hipotesis)

P(A) : Probabilitas A

P(B) : Probabilitas B

P(A|B) : Probabilitas A dengan syarat terjadi B

P(B|A) : Probabilitas B dengan syarat terjadi A

Merupakan teorema *Bayes* yang paling sederhana, jika ruang sampel **Ω** dapat dibagi menjadi banyak kejadian sehingga $A_{1,}$ $A_{1,\ \ldots}$ $A_{n,}$ dan jika *B* adalah *P (B)>0,* yang merupakan penyatuan semua A_i lalu untuk masing-masing A_i dengan rumus sebagai berikut :

$$P(Ai|B) : \frac{P(B|Ai)\ P(Ai)}{\sum_{j:1}^{n} P(B|Aj)\ P(Aj)} \qquad (3.5)$$

Teorema *Naïve Bayes* data digunakan untuk menurunkan probabilitas posterior dari suatu hipotesis berdasarkan data yang diamati.

Naïve Bayes Classifier memiliki berbagai kelebihan jika dibandingkan model lainnya, Epri (2022) menjelaskan bahwa model tersebut memiliki tingkat komputasi yang sederhana, kecepatan pengolahan data yang tinggi namun dapat memberikan performa yang baik jika dibandingkan dengan model klasifikasi lainnya. Adapun *Naïve Bayes Classifier* dapat digunakan secara efektif jika diaplikasikan pada basis data yang besar dan juga data yang beragam (Ardiansyah et al., 2023). Metode ini mengasumsikan independensi variabel, sehingga hanya varians dari variabel dalam suatu kelas yang diperlukan untuk menentukan klasifikasi, bukan seluruh matriks kovarians.

3.11. *Word Cloud*

Word cloud (atau juga disebut wordle) merujuk pada sebuah kumpulan kata yang memiliki bentuk dan ukuran yang berbeda (Gupta et al., 2019). Ukuran dari kata tersebut menggambarkan frekuensi atau tingkat kepentingan dari sebuah kata. Semakin besar suatu kata maka akan semakin sering juga kata tersebut diulang. Tujuan dari visualisasi *word cloud* ini adalah untuk memberikan *highlight*, merangkum kata-kata dari sebuah dokumen, dan visualisasi pada kata-kata yang penting dalam suatu korpus sehingga lebih mudah untuk mengidentifikasi tema maupun topik yang berada di dalam suatu kumpulan kalimat. Parameter visualisasi dari *word cloud* sendiri dapat ditentukan oleh penggunanya, mulai dari penentuan warna, skema, jenis huruf/*font*, serta bentuk *layout*. Kata-kata yang disampaikan

dalam visualisasi *word cloud* adalah kata-kata yang sudah melewati tahap *text processing* sehingga diharapkan dapat menggambarkan berbagai kata yang akan dianalisis dalam suatu model. Gambar di bawah ini adalah salah satu contoh visualisasi dari *word cloud.*

Gambar 3.2. Contoh *Word Cloud*

Pada **Gambar 3.2** kata yang digambarkan paling sering muncul yaitu “kemudi”, “maxim”, “aplikasi”, “batal”, “susah”. Secara visual kata kemudi menjadi katayang paling penting dan paling banyak keluar karena memiliki ukuran yang paling besar, kata-kata yang memiliki ukuran lebih kecil dari kata “kemudi” maka akan memiliki frekuensi lebih kecil.

BAB IV
METODOLOGI PENELITIAN

4.1. Populasi Penelitian

Dalam penelitian ini, populasi mengacu pada data komentar di media sosial Instagram yang berkaitan dengan wisata di Provinsi Daerah Istimewa Yogyakarta. Sampel yang diambil pada penelitian ini adalah data komentar Instagram dari lima akun. Yakni, @explorejogja, @jogjaviral, @wisatajogja_, @pariwisata.jogjakarta, @wonderfuljogja selama satu tahun, Agustus 2022 - Agustus 2023 yang dikumpulkan (dilakukan *scrapping* data) pada bulan September 2023. Secara keseluruhan, data yang diperoleh berjumlah 26406 baris.

4.2. Jenis dan Sumber Data

Penelitian ini menggunakan data yang diambil dari media sosial Instagram dengan kata kunci wisata di Jogja. Data diambil dari tanggal 1 Agustus 2023 – 30 Agustus 2023. Adapun lokasi dari penelitian ini adalah Fakultas MIPA UII.

4.3. Variabel penelitian

Terdapat 1 variabel yang digunakan dalam penelitian ini yakni variabel teks. Penjelasan mengenai variabel yang digunakan disajikan dalam tabel 4.1. di bawah ini.

Tabel 4.1. Variabel Penelitian

Nama Variabel	Definisi Variabel
Komentar	Penilaian atau opini dari warganet media sosial Instagram mengenai wisata di Provinsi Daerah Istimewa Yogjakarta

4.4. Metode Analisis

Penelitian ini menggunakan metode analisis sebagai berikut:

1. *Web Scraping*

   Teknik ini digunakan untuk mengambil data yang berada dalam suatu *website*. Dalam penelitian ini, *web scraping* akan digunakan untuk mengambil data di Instagram.

2. *Text mining*

   Teknik ini digunakan untuk mengolah dan menganalisis data berjenis teks yang tidak terstruktur. Dalam teknik ini terdapat salah satu tahapan yang

disebut dengan *text processing* dimana tahap tersebut akan membersihkan data teks dari *noise* sehingga dapat dianalisis dan menghasilkan model yang baik.

3. Analisis sentimen *lexicon-based*

   Teknik analisis sentimen adalah suatu teknik pemodelan data yang bertujuan untuk melakukan klasifikasi sentimen terhadap suatu teks atau kalimat. Penentuan kelas sentimen pada data yang digunakan untuk membangun model ditentukan melalui perhitungan skor sentimen melalui kamus sentimen. Kelas sentimen yang digunakan dalam penelitian ini adalah sentimen positif dan negatif.

4. TF-IDF

   Term Frequency – Inverse Document Frequency (TF-TDF) merupakan sebuah tabel yang digunakan untuk memberikan bobot pada setiap kata sehingga dapat ditentukan keterkaitan suatu kata dengan dokumen tertentu.

5. *Naïve Bayes Classifier*

   Algoritma *Naïve Bayes Classifier* merupakan salah satu teknik klasifikasi dalam *machine learning* yang digunakan untuk menentukan kelas suatu objek.

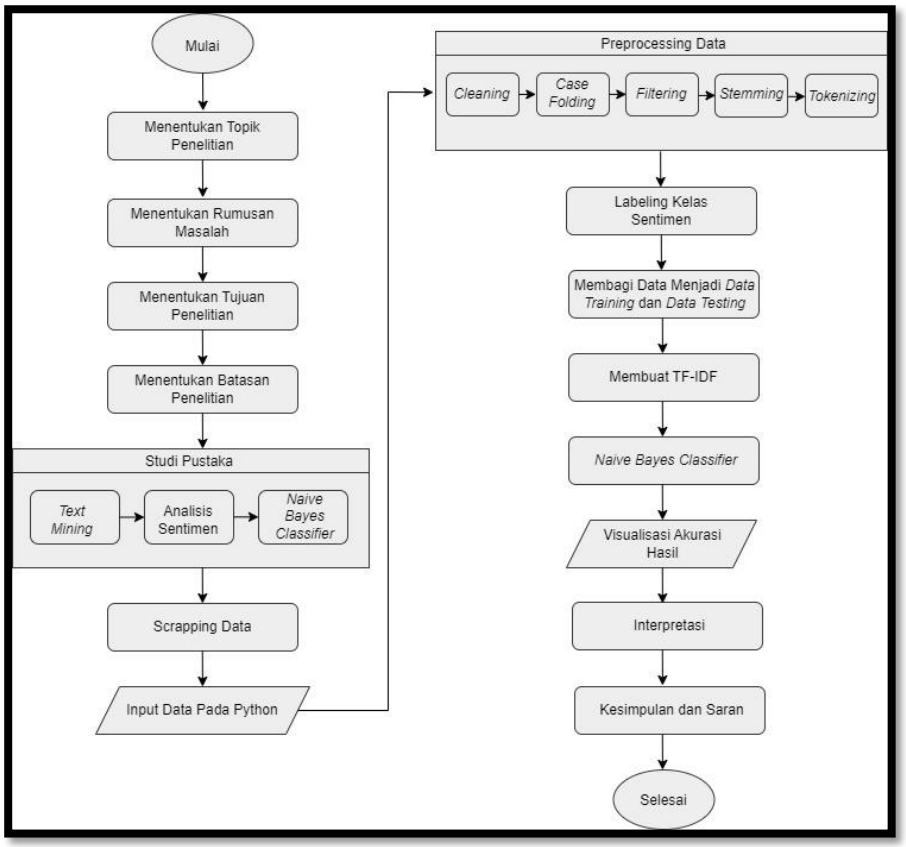
6. *Word cloud*

   Teknik *word cloud* digunakan untuk memvisualisasikan berbagai kata yang sering muncul dalam beberapa dokumen.

Data penelitian ini diambil dengan melakukan teknik *web scraping* dari media sosial Instagram. Melalui teknik tersebut maka berbagai elemen yang berada di website Instagram dapat diambil secara otomatis sehingga dapat digunakan untuk analisis data. Pada penelitian ini, elemen yang diambil dari website Instagram adalah komentar-komentar para netizen di akun-akun yang akan menjadi fokus analisis. Adapun aplikasi yang digunakan dalam pengolahan dan analisis data adalah *Google Colaboratory*. Aplikasi berbasis *cloud* tersebut memungkinkan penggunanya untuk menuliskan serta mengeksekusi bahasa pemrograman *python* melalui *web browser*.

4.5. Alat dan cara organisir data

Berbagai tahapan yang akan dilakukan dalam penelitian ini disampaikan pada diagram alur (*flowchart*) pada gambar 4.1. di bawah ini.



Gambar 4.1. Diagram Alir Tahapan Penelitian

Berikut adalah penjelasan dari *flowchart* di atas:

1. Tahap pertama adalah menentukan topik yang akan dianalisis dalam penelitian ini. Adapun pada penelitian ini topik yang diangkat adalah Analisis Sentimen Terhadap Opini Warganet Instagram Tentang Wisata D.I. Yogyakarta.
2. Tahap selanjutnya, peneliti menentukan rumusan masalah yang diangkat, tujuan dilakukannya penelitian serta merumuskan batasan-batasan dalam penelitian.
3. Tahapan selanjutnya, peneliti melakukan *web scraping* terkait data mengenai wisata D.I. Yogyakarta di Instagram.
4. Setelah data diperoleh dari Instagram, peneliti melanjutkan proses *preprocessing data* atau disebut dengan *text cleaning* sehingga data yang

diperoleh bersih dari *noise*. Adapun tahap ini terdiri dari, *case folding, filtering, tokenizing,* dan *stemming*.

5. Data dibagi menjadi data *training* dan data *testing.*
6. Setelah data dibagi menjadi 2, peneliti melakukan pembobotan menggunakan teknik TF-IDF sehingga terbentuk tabel yang menyampaikan keterkaitan kata dengan dokumen.
7. Setelah itu peneliti membangun model dengan *naïve bayes classifier* dengan data *training* dan mengevaluasi performa model tersebut pada data *testing*.
8. Memvisualisasikan kata pada kumpulan dokumen bersentimen negatif dan positif menggunakan *word cloud*.
9. Menyampaikan kesimpulan dari penelitian yang telah dilakukan.

BAB V
HASIL DAN PEMBAHASAN

5.1. Pengumpulan Data

Dalam penelitian ini, proses pengumpulan data dilakukan dengan menggunakan teknik *scraping* terhadap data yang terdapat di media sosial Instagram. Melalui teknik scraping tersebut, maka komentar-komentar yang terdapat di dalam suatu akun Instagram dapat secara otomatis terambil. Adapun *query* yang digunakan untuk mengambil data di Instagram disajikan pada tabel 5.1. berikut.

Tabel 5.1. Query Pengambilan Data Instagram

Atribut	Nilai
Kata kunci	wisata di DI Yogyakarta
Waktu mulai	01-08-2022
Waktu selesai	30-08-2023
Akun yang diambil	@explorejogja, @jogjaviral, @wisatajogja_, @pariwisata.jogjakarta, @wonderfuljogja

Melalui query tersebut maka diperoleh data dari Instagram sebanyak 26406 baris dan 12 kolom. Instagram dalam hal ini membatasi penggunanya untuk mengambil datanya sehingga hanya 26406 data saja yang dapat diambil dalam kurun waktu satu tahun. Adapun kolom yang akan diolah lebih lanjut adalah kolom yang bernama *text* dimana kolom tersebut berisikan komentar dari pengguna Instagram yang memenuhi kriteria pada tabel 5.1. Pada langkah selanjutya data kolom *text* tersebut akan diolah sehingga menghasilkan model yang dapat memprediksi sentiment.

5.2. *Text Processing*

Tahap *text processing* akan dilakukan untuk membersihkan data dari *noise* sehingga data lebih terstruktur dan pada akhirnya dapat memberikan model yang baik dalam memprediksi. Adapun tahap ini terdiri dari *text cleaning, case folding, filtering, tokenizing,* serta *stemming*.

5.2.1 *Cleaning Data*

Pembersihan data/*cleaning data* adalah tahap untuk menghapus komponen-komponen teks yang tidak diperlukan dalam pengolahan data. Adapun komponen yang secara umum terdapat dalam sebuah komentar adalah tanda pagar (#), tautan atau URL, tanda *tag* (@), maupun *emoticon*. Seluruh karakter selain huruf alfabet akan dihapus pada tahap ini. Contoh proses *cleaning data* disajikan pada tabel 5.2. dibawah ini.

Tabel 5.2. Contoh proses *cleaning data*

Sebelum	Sesudah
Yukk muter2rin jogja pake motor	Yukk muterrin jogja pake motor
Emang macet nggak bisa diatas 40 km jam lah	Emang macet nggak bisa diatas km jam lah
ojo aneh虏	Ojo aneh

5.2.2 *Case Folding*

Tahap selanjutnya setelah melakukan pembersihan data adalah *case folding*. Tujuan dari tahap ini adalah untuk menyeragamkan huruf-huruf pada kolom komentar yang akan dianalisis lebih lanjut. Adapun pada penelitian ini, seluruh huruf yang ada pada kolom komentar akan diubah seluruhnya menjadi huruf kecil (*lowercase*). Melalui proses ini pula maka akan diperoleh kata-kata yang unik antara satu dengan yang lainnya atau tidak bersifat *case sensitive*. Tabel 5.3. dibawah ini menjelaskan contoh proses *case folding*.

Tabel 5.3. Contoh proses *case folding*

Sebelum	Sesudah
Hujan membawa kenangan	hujan membawa kenangan
Harus ke semua tempat nih	harus ke semua tempat nih
Alhamdulillah bersih min	alhamdulillah bersih min

5.2.3 *Filtering*

Melalui tahap *filtering*, akan dilakukan penghapusan terhadap kata-kata yang secara umum tidak memiliki arti secara substansi maupun konteks bahasa. Kata-kata umum atau *stopwords* tersebut tersimpan dalam sebuah kamus yang terdapat dalam *library* Sastrawi. Adapun tahap *filtering* akan menghapus *slang words* atau kata-kata yang tidak baku yang sering digunakan dalam percakapan sehari-hari.

Pada penelitian ini, kumpulan dari *slang words* disatukan dalam sebuah *dataframe*. Tabel 5.4. menyajikan contoh hasil proses *filtering*.

Tabel 5.4. Contoh proses *filtering*

Sebelum	Sesudah
trauma servis ning ahass an motor ra soyo bener malah ragat akeh	trauma servis di ahass an motor ra soyo bener malah ragat banyak
hello jogja	halo jogja
jebul mungil	ternyata mungil

5.2.4 *Normalizing*

Normalisasi adalah istilah lain daripada *filtering*. Yaitu pembakuan suatu tata bahasa, baik itu secara penulisan ataupun jenis bahasa nya yang terklasifikasikan *slang* diubah ke dalam *stopwards* (bentuk baku). Contoh bentuk penulisan, "skrng" diubah menjadi "sekarang", "jd" diubah menjadi "jadi", "yg" diubah menjadi "yang" dan sebagainya. Sementara yang dibakukan dari segi jenis bahasanya adalah, bahasa non formal ataupun bahasa daerah. Sebagai contoh, "make" menjadi "pakai", "ngarti" menjadi "mengerti", "sing" menjadi "yang", dan seterusnya. Seperti yang ditampulkan pada gambar berikut.

```
[ ] slangword = pd.read_excel('/content/Normalisasi.xlsx', sheet_name='Normalisasi')
    slangword.head()
```

	slang	normal
0	skrg	sekarang
1	jd	jadi
2	make	pakai
3	ngarti	mengerti
4	dr	dari

Gambar 5.1. Normalisasi

5.2.5 *Stemming*

Tahap *stemming* bertujuan untuk menyederhanakan kata-kata yang memiliki imbuhan maupun partikel yang melekat. Imbuhan maupun partikel tersebut perlu dihapus sehingga setiap kata yang akan dianalisis adalah kata dasarnya. Proses *stemming* ini dilakukan dengan fungsi StemmerFactory pada *library* Sastrawi. Tabel 5.5. adalah contoh dari proses *stemming*.

Tabel 5.5. Contoh proses *stemming*

Sebelum	Sesudah
termurah yakin nih yuki auce termurah yuuki auce	murah nih yuk auce murah yuuki auce
asli topping melimpah tekstur roti juga lembut empuk cocok banget nemenin kopi penikmat gembong tidak pernah bohong	asli topping limpah tekstur roti lembut empuk cocok banget nemenin kopi nikmat gembong tidak bohong
niat mau irit tenaga cuci piring hasilnya malah nambah sampah mana jogja darurat sampah lagi	niat irit tenaga cuci piring hasil nambah sampah jogja darurat sampah

5.2.6 *Tokenizing*

Tahap terakhir pada *text processing* adalah *tokenizing*. Pada tahap ini setiap kalimat yang berada di setiap baris akan dipecah per kata. Tabel 5.6. dibawah ini menjelaskan contoh dari proses *tokenizing*.

Tabel 5.6. Contoh proses *tokenizing*

Sebelum	Sesudah
enak banget suka keluarga sukak pizza bianche alias cheese sauce recommend banget layan ramah	['enak', 'banget', 'suka', 'keluarga', 'sukak', 'pizza', 'bianche', 'alias', 'cheese', 'sauce', 'recommend', 'banget', 'layan', 'ramah']
yukk muterrin jogja pakai motor	['yukk', 'muterrin', 'jogja', 'pakai', 'motor']
mall jogja bagus yaa	['mall', 'jogja', 'bagus', 'yaa']

5.3. Pelabelan Kelas Sentimen

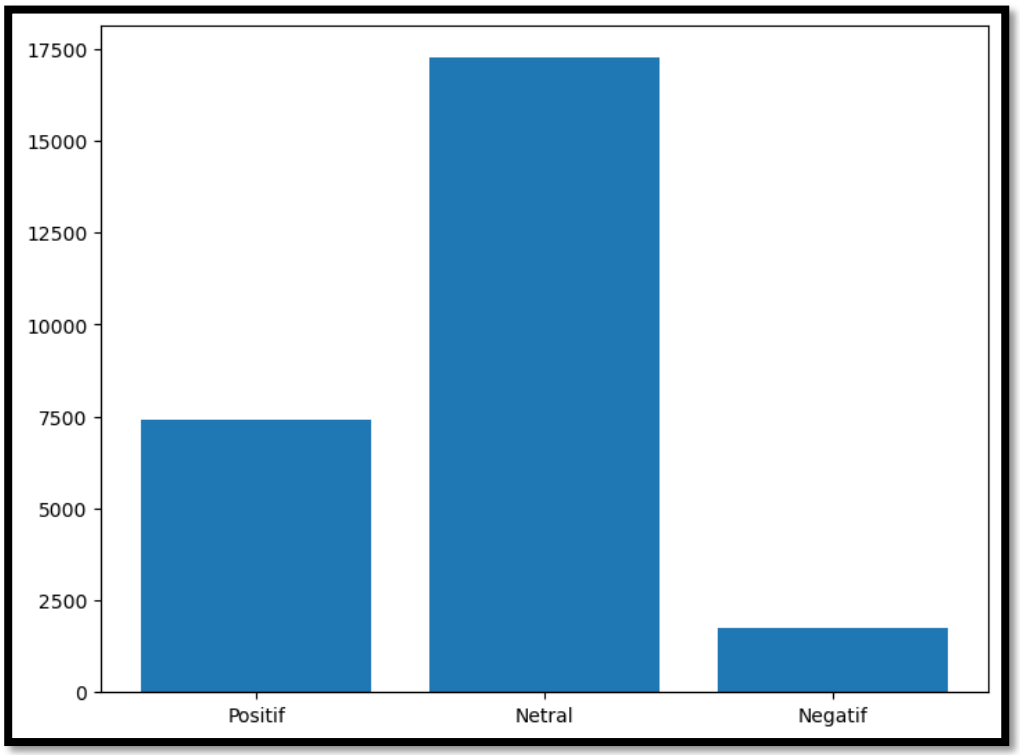
Setelah seluruh data dibersihkan dalam tahap *text processing* maka selanjutnya akan dilakukan pelabelan kelas sentimen pada setiap baris kolom komentar. Penelitian ini mengklasifikasikan label sentimen menjadi tiga kelas yakni sentimen positif, netral dan negatif. Adapun dalam penelitian ini pelabelan akan menggunakan teknik *lexicon-based* dimana penentuan label sentimen dihitung berdasarkan skor sentimen sesuai dengan referensi tabel *lexicon* yang digunakan.

Kamus *lexicon* yang akan digunakan adalah "Kamus Sentimen Bahasa Indonesia" yang dapat diunduh pada GitHub *https://github.com/masdevid/ID-OpinionWords*. Di dalam kamus *lexicon* tersebut setiap kata negatif memiliki skor -1 dan kata positif memiliki skor 1. Tabel 5.7. dibawah ini menyajikan contoh dari hasil pelabelan kelas sentimen.

Tabel 5.7. Contoh proses pelabelan

Data	Skor	Keterangan
['cahaya', 'ayo', 'gas']	3	positif
['pergi', 'jogja', 'luka', 'tiada'']	-1	negatif
['nikmat', 'pandang', 'rapi']	2	positif

Melalui hasil pelabelan sentimen di atas, maka diperoleh kalimat kelas sentimen positif sebanyak 7413, kalimat sentimen negatif sebanyak 1734 dan kalimat sentimen netral sebanyak 17259. Sebaran kelas sentimen disajikan dalam gambar 5.1. di bawah ini.



Gambar 5.2. Sebaran Kelas Sentimen

5.4. Data *Training* dan Data *Testing*

Dalam proses *machine learning* data akan dibagi menjadi 2 bagian yakni *data training* serta *data testing*. *Data training* merupakan sebuah bagian data yang digunakan untuk membangun atau melatih suatu model, sedangkan *data testing* merupakan bagian data lainnya yang digunakan untuk mengetahui serta menghitung performa dari model yang sudah dibagun pada proses latihan pada *data training*. Dalam membangun model dari data komentar, penelitian ini hanya menggunakan sentimen negatif dan positif saja sehingga kalimat atau baris yang memiliki sentimen netral akan dihilangkan dari analisis. Adapun dalam penelitian ini data *training* dan *testing* dibagi dengan persentase masing-masing 80% dan 20% dari keseluruhan data. Melalui pembagian ini diperoleh data *training* sebanyak 7317 baris dan data *testing* sebanyak 1830 baris.

5.5. *Term Frequency – Inverse Document Frequency* (TF-IDF)

Metode *Term Frequency – Inverse Document Frequency* atau TF-IDF merupakan salah satu metode yang paling umum digunakan untuk pencarian informasi dengan memberikan bobot pada kata. Data yang digunakan dalam tahap pembuatan TF-IDF adalah data *training* yang memiliki proporsi 80% dari keseluruhan data. Dalam membuat TF-IDF terdapat beberapa langkah yakni membuat *document term matrix* (DTM), lalu membuat tabel *term Frequency* (TF), dan selanjutnya membuat *Inverse Document Frequency* (IDF). Tabel DTM, langkah pertama dalam membuat TF-IDF, adalah suatu tabel yang menggambarkan seberapa sering suatu kata tertentu muncul dalam keseluruhan dokumen yang dianalisis. Tabel 5.8. menyajikan contoh dari tabel DTM yang dibangun pada penelitian ini.

Tabel 5.8. Tabel *Document Term Matrix*

Dok.	jogja	keren	bagus	Sampah	hancur
1	0	0	0	0	0
2	0	0	0	1	0
3	0	0	0	0	1
4	0	0	0	0	0
5	1	1	1	0	0
…	…	…	…	…	…

Dok.	jogja	keren	bagus	Sampah	hancur
7313	0	1	0	0	0
7314	0	0	0	0	0
7315	0	1	0	0	0
7316	0	0	1	0	0
7317	0	0	0	0	0
Jumlah	776	525	474	76	3

Berdasarkan tabel 5.8. diatas mengenai *document term matrix*, diketahui bahwa terdapat 7317 dokumen yang digunakan untuk perhitungan TF-IDF. Tabel tersebut hanya tangkapan dari 5 dokumen teratas dan terbawah sebagai contoh untuk perhitungan tabel DTM. Tabel tersebut menunjukkan frekuensi dari munculnya sebuah kata pada dokumen, dalam hal ini kalimat, tertentu. Dapat diketahui bahwa kata "jogja" berjumlah 776, "keren" berjumlah 525 dan seterusnya. Untuk menjelaskan hubungan antara kata dengan dokumen, maka tabel 5.8. menjelaskan berapa banyak kata tertentu yang keluar pada sebuah dokumen tertentu. Misal kata jogja muncul 1 kali pada dokumen 5. Adapun kata "sampah" muncul 1 kali pada dokumen 2. Setelah mendapatkan DTM maka selanjutnya menghitung tabel *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF). Tabel TF bertujuan untuk mengetahui proporsi suatu kata yang muncul pada dokumen tertentu terhadap jumlah munculnya suatu kata.

Tabel 5.9. Tabel *term frequency*

Dok	jogja	keren	bagus	Sampah	hancur
1	0	0	0	0	0
2	0	0	0	$\frac{1}{76} = 0.013$	0
3	0	0	0	0	$\frac{1}{3} = 0.33$
4	0	0	0	0	0
5	$\frac{1}{776} = 0.0012$	$\frac{1}{525} = 0.0019$	$\frac{1}{474} = 0.0021$	0	0
…	…	…	…	…	…
7313	0	$\frac{1}{525} = 0.0019$	0	0	0

Dok	jogja	keren	bagus	Sampah	hancur
7314	0	0	0	0	0
7315	0	$\frac{1}{525} = 0.0019$	0	0	0
7316	0	0	$\frac{1}{474} = 0.0021$	0	0
7317	0	0	0	0	0

Teknik TF digunakan untuk memberikan bobot pada setiap kata yang muncul berdasarkan seberapa sering kata tersebut muncul dalam suatu dokumen. Berdasarkan hasil tabel TF pada tabel 5.9. dapat disampaikan sebagai contoh bahwa 1 kata "jogja" memiliki bobot 0.0012. Adapun satu kata "keren" memiliki bobot 0.0019.

Tabel 5.10. Tabel *Inverse Document Frequency*

Dok	jogja	keren	bagus	Sampah	hancur
1	0	0	0	0	0
2	0	0	0	$\log\left(\frac{7317}{76}\right) = 1.983$	0
3	0	0	0	0	$\log\left(\frac{7317}{3}\right) = 3.386$
4	0	0	0	0	0
5	$\log\left(\frac{7317}{776}\right) = 0.974$	$\log\left(\frac{7317}{525}\right) = 1.143$	$\log\left(\frac{7317}{474}\right) = 1.188$	0	0
…	…	…	…	…	…
7313	0	$\log\left(\frac{7317}{525}\right) = 1.143$	0	0	0
7314	0	0	0	0	0
7315	0	$\log\left(\frac{7317}{525}\right) = 1.143$	0	0	0

Dok	jogja	keren	bagus	Sampah	hancur
7316	0	0	$\log\left(\frac{7317}{474}\right) = 1.188$	0	0
7317	0	0	0	0	0

Tabel IDF mengukur seberapa unik suatu kata dalam seluruh koleksi dokumen. Kata-kata yang muncul di banyak dokumen mendapatkan nilai IDF yang rendah sementara kata-kata yang hanya muncul dalam sedikit dokumen memiliki nilai IDF yang tinggi. Seperti yang disampaikan pada tabel 5.10, kata "jogja" memiliki nilai IDF yang rendah jika dibandingkan kata"hancur". Berdasarkan nilai IDF tersebut maka selanjutnya dibuatlah tabel TF IDF dengan mengalikan nilai TF dengan IDF pada masing-masing kotak. Semakin tinggi nilai TF-IDF maka akan semakin besar hubungan antara kata dengan dokumennya.

Tabel 5.11. Tabel TF-IDF

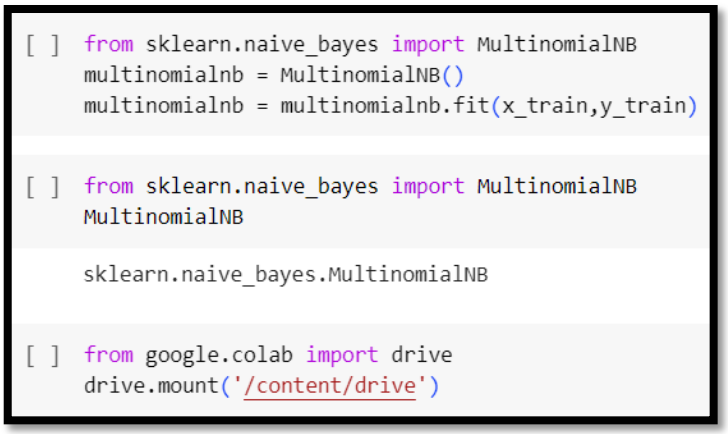
Dokumen	jogja	keren	bagus	Sampah	hancur
1	0	0	0	0	0
2	0	0	0	0.0257	0
3	0	0	0	0	1.117
4	0	0	0	0	0
5	0.0011	0.0021	0.0024	0	0
…	…	…	…	…	…
7313	0	0.0021	0	0	0
7314	0	0	0	0	0
7315	0	0.0021	0	0	0
7316	0	0	0.0024	0	0
7317	0	0	0	0	0

Tabel TF-IDF dapat digunakan untuk mengukur seberapa penting suatu kata dalam konteks kalimat tertentu dan dapat juga digunakan dalam berbagai macam aplikasi seperti pengambilan informasi, indeksasi dokumen, serta klasifikasi suatu teks. Skor dari TF-IDF ini menggambarkan pentingnya suatu istilah pada suatu dokumen relatif terhadap dokumen yang lainnya. Berdasarkan tabel TF-IDF, pembobotan paling besar diberikan pada kata "hancur" yakni 1.117 karena kata

tersebut paling jarang keluar dalam keseluruhan dokumen. Adapun kata “jogja” merupakan kata yang memiliki bobot paling kecil yakni 0.0011 karena paling sering keluar dari seluruh dokumen. Semakin tinggi skor TF-IDF maka semakin penting istilah tersebut dan semakin kecil skor TF-IDF maka istilah tersebut tidak begitu berkontribusi pada keseluruhan dokumen.

5.6. Membangun Model

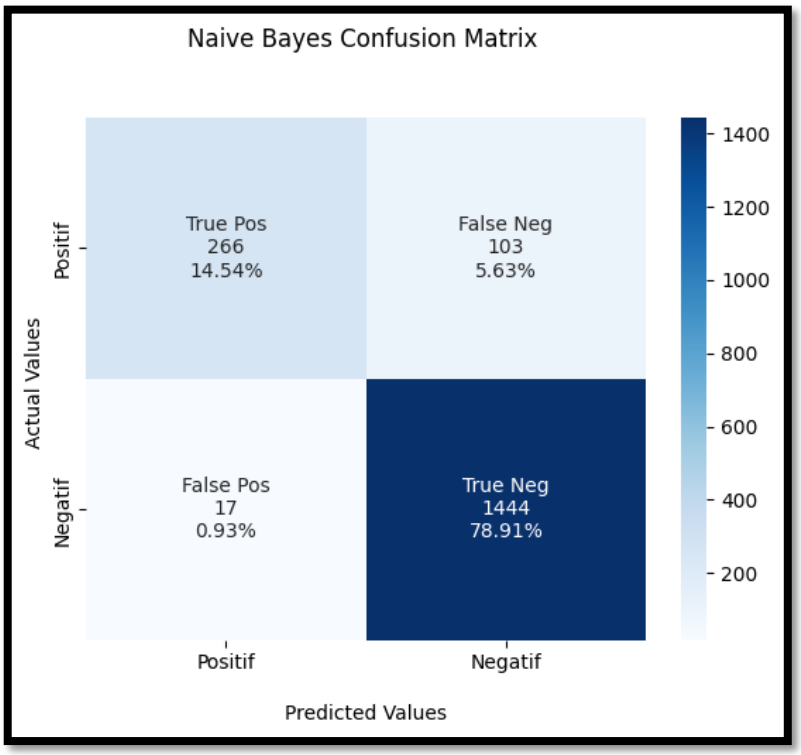
Melalui tabel TF-IDF yang telah dibentuk, maka model *machine learning* dapat dibangun. Pada penelitian ini algoritma *Naïve Bayes* Multinomial digunakan untuk memodelkan TF-IDF dengan kelas sentimen pada masing-masing dokumen atau kalimat. Model dibangun dengan menggunakan data *training* yang telah dibentuk menjadi tabel TF-IDF.



Gambar 5.3. Model *Naïve Bayes* Multinomial

5.7. Evaluasi Model

Melalui model yang telah dibangun dan dilatih dengan menggunakan *data training* dan algoritma Naïve Bayes, selanjutnya model tersebut akan digunakan untuk mengklasifikasi *data testing*. Terdapat beberapa metrik yang akan diukur dalam proses evaluasi model ini yakni akurasi, *recall*, dan *precision*. Dalam mengukur metrik tersebut, *confusion matrix* akan dibangun untuk menghitung hasil klasifikasi model. Berdasarkan 1830 baris data *testing* (20% dari keseluruhan data) maka diperoleh *confusion matrix* seperti yang disajikan pada gambar 5.12. di bawah ini. Pembuatan *confusion matrix* dilakukan dengan menggunakan jenis grafik *heatmap* pada *library seaborn*.



Gambar 5.4. *Confusion Matrix*

Melalui *confusion matrix* pada gambar 5.2 di atas maka dapat disampaikan interpretasi sebagai berikut:

- *True Positive* (TP): Kalimat yang memiliki sentimen positif dan diprediksi memiliki sentimen positif sebanyak 266 data (14,54%).
- *True Negative* (TN): Kalimat yang memiliki sentimen negatif dan diprediksi memiliki sentimen negatif sebanyak 1444 data (78.91%).
- *False Positive* (FP): Kalimat yang memiliki sentimen negatif tetapi diprediksi memiliki sentimen positif sebanyak 17 data (0.93%)
- *False Negative* (FN): Kalimat yang memiliki sentimen positif tetapi diprediksi memiliki sentimen negatif sebanyak 103 data (5.63%)

Berdasarkan hasil perhitungan *confusion matrix* tersebut maka selanjutnya dapat dihitung metrik evaluasi model. Melalui algoritma *naïve bayes classifier* maka nilai dari akurasi, *recall* dan *precision* yang didapatkan sebagai berikut:

$$Akurasi = \frac{TP + TN}{TP + FP + FN + TN} * 100\% = \frac{266 + 1444}{266 + 17 + 103 + 1444} = 93.4\%$$

$$Recall = \frac{TP}{TP + FN} * 100\% = \frac{266}{266 + 103} = 72.08\%$$

$$Precision = \frac{TP}{TP + FP} * 100\% = \frac{266}{266 + 17} = 93.99\%$$

5.8. Visualisasi Komentar Positif dan Negatif

5.8.1 Visualisasi Komentar Positif

Gambar 5.5. *Word cloud* Komentar Positif

Pada *word cloud* yang disampaikan pada gambar 5.3, semakin besar ukuran kata maka akan semakin sering juga kata tersebut muncul pada kumpulan komentar positif. Gambar 5.4. di bawah ini disampaikan untuk memperjelas jumlah kata yang muncul pada *word cloud* tersebut.

	Common_words	count
1	kesini	851
2	jogja	781
3	kak	691
4	keren	665
5	banget	652
6	bagus	571
7	nih	565
8	ya	537
9	nya	476
10	enak	401
11	min	358
12	rapi	301
13	gas	287
14	yuk	280
15	indah	280

Gambar 5.6. 15 kata paling sering muncul pada komentar positif

Berdasarkan hasil visualisasi *word cloud* dan juga frekuensi top 15 kata di komentar positif, maka dapat disampaikan bahwa komentar mengenai wisata di jogja secara umum mengajak para warganet untuk datang ke jogja karena kata "kesini" mendapatkan nilai frekuensi paling tinggi. Dan kata "jogja" juga

menempati urutan tertinggi kedua, yakni tepat setelah kata ajakan “kesini”, yang bermakna memang topik yang sedang dibicarakan adalah wisata Provinsi DIY, tidak peduli tentang apa yang dibicarakan atau bagaimana sentimennya, karena baik sentimen positif ataupun negatif akan berdampak pada penempatan kata “jogja” yang berada pada urutan teratas, karena memang itulah ide atau topik bahasan yang sedang dibicarakan. Sehingga bagaimanapun beropini atau berkomentar tentang wisata DIY, pasti akan banyak melibatkan kata “jogja” di dalamnya. Adapun istilah-istilah mengenai “keren”, “bagus”, “banget”, “indah”, “yuk” dan “rapi” terkategorikan dalam top 15 kata-kata pada komentar positif, karena jelas, kata-kata tersebut memberikan gambaran bagaimana warganet memberikan kesan maupun ulasan positif bagi wisata DIY. Sementara kata-kata yang secara makna dasar terkesan netral namun tetap dikategorikan sebagai kata positif, bahkan masuk ke dalam top 15 kata-kata dalam komentar positif seperti “kak”, “nih”, “ya”, “nya”, “min” adalah kata kata – kata tersebut umum digunakan saat *users* instagram berkomentar dengan intensi yang positif. Seperti kata “kak” yang digunakan saat mengajak kerabat, teman dengan cara me*mention username* akun kerabat atau temannya. Kata “nih” saat merekomendasikan wisata DIY yang dimaksud kepada kerabat atau temannya. “ya” saat membalas rekomendasi dari kerabat atau temannya atas saran wisata yang dimaksud, yang bermakna atau berintensi setuju, sepakat, atau mengiyakan atas intensi positif yang disimpulkan sebagai kata positif juga. Juga kata “nya” yang dapat bermakna postif ataupun negatif. Namun turut terkategorikan ke dalam komentar positif karena kata “nya” seringkali muncul saat membahas topik wisata terkait yang dianggap menarik atensi, dan digunakan kata ganti “nya” dalam merepresentasikan topik wisata terkait.

5.8.2 Visualisasi Komentar Negatif

Gambar 5.7. *Word cloud* Komentar Negatif

Pada *word cloud* yang disampaikan pada gambar 5.5, semakin besar ukuran kata maka akan semakin sering juga kata tersebut muncul pada kumpulan komentar negatif. Gambar 5.4. di bawah ini disampaikan untuk memperjelas jumlah kata yang muncul pada *word cloud* tersebut.

	Common_words	count
1	jogja	200
2	ya	167
3	nya	164
4	jalan	127
5	min	115
6	kalo	95
7	tidak	94
8	lambat	93
9	salah	89
10	sampah	86
11	pas	86
12	takut	84
13	lupa	84
14	belah	76
15	udah	73

Gambar 5.8. 15 kata paling sering muncul pada komentar negatif

Berdasarkan hasil visualisasi *word cloud* dan juga frekuensi top 15 kata di komentar negatif, maka dapat disampaikan bahwa terdapat beberapa ulasan yang menggambarkan hal negatif mengenai wisata di jogja. Kata "jogja" menempati kata teratas dalam sentimen negatif bukan berarti bermakna negatif secara harfiah. Namun seperti yang telah disampaikan pada paragraf sebelumnya, kata "jogja" bermakna bahwa topik yang sedang dibicarakan mengenai wisata Provinsi DIY, tidak peduli tentang apa yang dibicarakan atau bagaimana sentimennya, karena baik sentimen positif ataupun negatif akan berdampak pada penempatan kata "jogja" yang berada pada urutan teratas, karena memang itulah ide atau topik bahasan yang sedang dibicarakan. Sehingga bagaimanapun beropini atau berkomentar tentang wisata DIY, positif ataupun negatif, pasti akan banyak melibatkan kata "jogja" di dalamnya. Adapun istilah "nya" terkategorikan ke dalam sentimen negatif karena frekuensi atau keseringannya muncul saat para warganet atau instagram *users* memberikan ulasan atau komentar negatif tentang wisata di Provinsi DIY dan digunakan kata ganti "nya" dalam merepresentasikan topik wisata terkait. Kendati

kata “nya” yang sebenarnya dapat bermakna postif ataupun negatif. Sementara kata kata lain seperti “pantai”, “takut”, “hilang”, “belah”, “lupa” masuk di dalam frekuensi teratas dalam *wordcloud* komentar negatif karena kata – kata tersebut memberikan gambaran mengenai fenomena atau masalah yang terjadi pada wisata di Provinsi Daerah Istimewa Yogyakarta. Saat*scrapping* data yang dilakukan di bulan September 2023 dari komentar warganet di instagram selama satu tahun, Agustus 2022 sampai Agustus 2023 mengenai wisata di Provinsi Daerah Istimewa Yogyakarta.

Dimana ditemukan fenomena – fenomena di pantai Parangtritis, Baron, dan pantai lainnya di sekitar Gunung Kidul, DIY pada kurun waktu Agustus 2022-Agustus 2023 yang dianggap menakutkan yang membuat warganet beropini demikian melalui representasi kata “pantai” dan “takut”. Terdapat juga beberapa femonema lain yang sering menarik atensi warganet tentang sisi negatif wisata DIY adalah fenomena hilangnya wisatawan yang berenan di Pantai Parangtritis, terbelahnya air di Pantai Baron, Gunung Kidul, DIY pada Februari 2024 yang direpresentasikan dalam kata “hilang”, dan “belah”. Yang mana mampu membuat “takut” warganet, dan dianggap sebagai hal yang negatif. Sehingga sering dikomentari hal tersebut dengan komentar yang berintensikan negatif. Kendati fenomena terpisah warna air laut di lepas Pantai Baron yang terkesan seperti terbelah menjadi dua warna yang kontras, bukan merupakan hal yang negatif. Seperti yang dijelaskan oleh Peneliti ahli utama Kelompok Riset Petrologi dan Mineralogi Pusat Riset Sumber Daya Geologi Organisasi Riset Kebumian dan Maritim BRIN, Haryadi Permana bahwa fenomena tersebut biasa terjadi di muara sungai yang menjadi tempat pertemuan air tawar dan air asin. "Secara garis besar, air warna cokelat, tawar, mengandung banyak sedimen terlarut," kata Haryadi, Senin (20/2/2023). Haryadi juga menerangkan bahwa terpisahnya air laut menjadi dua warna bisa dipicu oleh hujan. Ia mengatakan, air sungai yang tawar akan membawa material halus hasil erosi sungai dan bertemu di muara dengaan air yang relatif asin. "Fenomena ini bisa dijumpai di muara sungai besar, terutama setelah hujan," sambungnya. Kendati belum banyak orang yang mengetahui fenomena terpisahnya air laut menjadi dua tidaklah berbahaya. Haryadi menampaikan, air yang terpisah menjadi dua warna akan tercampur secara alami.

BAB VI
PENUTUP

6.1. Kesimpulan

Berdasarkan hasil analisis data yang telah dilakukan, diperoleh kesimpulan penelitian sebagai berikut:

1. Melalui 26406 data yang telah dikumpulkan dari 5 akun Instagram dalam rentang waktu 1 Agustus 2022 – 30 Agustus 2023, analisis sentimen dengan teknik VADER berhasil mengelompokan 7413 komentar dalam sentimen positif, 1734 komentar dalam sentimen negatif dan 17259 dalam sentimen netral. Berdasarkan hasil tersebut maka dapat disimpulkan bahwa komentar mengenai pariwisata DIY lebih banyak berada pada sentimen positif dibandingkan sentimen negatif.
2. Melalui proporsi pembagian data *training* dan data *testing* sebesar 80:20 dari komentar yang memiliki sentimen positif dan negatif, model *naïve bayes classifier* yang dibangun menghasilkan performa akurasi sebesar **93.4%**, *recall* sebesar **72.08%**, dan *precision* sebesar **93.99%**. Nilai-nilai tersebut relatif cukup tinggi jika dibandingkan dengan penelitian-penelitian sebelumnya.
3. Berdasarkan analisis pada komentar-komentar di kelompok sentimen negatif dan positif diperoleh 15 kata yang paling banyak muncul. Kata-kata tersebut dapat menjadi gambaran mengenai keunggulan maupun permasalahan dari sektor pariwisata DIY.

6.2. Saran

Berdasarkan penelitian yang telah dilakukan, terdapat beberapa saran yang dapat menjadi pertimbangan bagi penelitian-penelitian selanjutnya, antara lain:

1. Penelitian selanjutnya dapat melakukan proses *crawling* data terkait pariwisata pada media sosial atau sumber lainnya seperti TikTok maupun *website* pemesanan hotel sehingga dapat diperoleh pemahaman lebih mendalam mengenai sentimen pariwisata jogja.

2. Penelitian selanjutnya dapat menggunakan model klasifikasi lainnya sebagai teknik *machine learning* untuk menganalisis komentar-komentar mengenai pariwisata di Yogyakarta, diantaranya *Support Vector Machine*, *K-Nearest Neighbors*, maupun *Artificial Neural Networks*.
3. *Stakeholder* pariwisata DIY dapat berfokus pada kata-kata yang muncul di kelompok sentimen negatif sehingga dapat menjadi bahan peningkatan kualitas pariwisata sesuai dengan keluhan masyarakat.

DAFTAR PUSTAKA

Ahmad, U. S. (2022). Implementasi Pariwisata terhadap Perekonomian Indonesia. *Al-DYAS*, *1*(1), 81–96. https://doi.org/10.58578/aldyas.v1i1.1319

Aji, C., & Andadari, R. K. (2021). Media Sosial Instagram Dan Website Terhadap Minat Kunjung Wisatawan. *Jurnal Penelitian Dan Pengembangan Sains Dan Humaniora*, *5*(1), 54–63.

Anggiyani, H. E. (2021). Sinergi Badan Promosi Pariwisata dan Akun Instagram Explore Bantul dalam Promosi Pariwisata Kabupaten Bantul. *Jurnal Pariwisata Indonesia*, *17*(2). https://doi.org/10.53691/jpi.v17i2.253

Ardiansyah, B., Daulay, I., Firdaus, M., Hutagaol, R., & Rahmaddeni, rahmadeni. (2023). *Analisis Sentimen Opini Publik Terhadap Penerimaan Bantuan Subsidi Upah (BSU) Menggunakan Algoritma Naive Bayes*. https://journal.irpi.or.id/index.php/sentimas

Azmi, M., Huda, A. K., & Setiyanto, A. (2020). PEMANFAATAN DATA INSTAGRAM UNTUK MENGETAHUI REPUTASI TEMPAT WISATA DI LOMBOK. *Jurnal TEKNIMEDIA*, *1*(1), 39–46.

Behdenna, S., Barigou, F., & Belalem, G. (2018). Document Level Sentiment Analysis: A survey. *EAI Endorsed Transactions on Context-Aware Systems and Applications*, *4*(13), 154339. https://doi.org/10.4108/eai.14-3-2018.154339

Beigi, G., Hu, X., Maciejewski, R., & Liu, H. (2016). An overview of sentiment analysis in social media and its applications in disaster relief. In *Studies in Computational Intelligence* (Vol. 639, pp. 313–340). Springer Verlag. https://doi.org/10.1007/978-3-319-30319-2_13

Databoks. (2023). *Indonesia Jadi Negara dengan Pengguna Instagram Terbanyak ke-4 di Dunia*. https://databoks.katadata.co.id/datapublish/2023/11/28/indonesia-jadi-negara-dengan-pengguna-instagram-terbanyak-ke-4-di-dunia#:~:text=Daftar%20sekarang%2C%20GRATIS!&text=Menurut%20laporan%20We%20Are%20Social,juta%20pengguna%20Instagram%20di%20Indonesia.

dataindonesia.id. (2023). *Jumlah Objek Daya Tarik Wisata di Indonesia*. https://dataindonesia.id/pariwisata/detail/indonesia-miliki-2563-objek-daya-tarik-wisata-pada-2021

Delgadillo, J. (2021). Machine learning: A primer for psychotherapy researchers. In *Psychotherapy Research* (Vol. 31, Issue 1, pp. 1–4). Routledge. https://doi.org/10.1080/10503307.2020.1859638

Dwi, W., & Kurniawati, N. (2016). Pemanfaatan Instagram oleh Komunitas Wisata Grobogan. *Komuniti*, *VIII*(2).

Elbagir, S., & Yang, J. (2019). Twitter Sentiment Analysis Using Natural Language Toolkit and VADER Sentiment. *Proceedings of the International MultiConference of Engineers and Computer Scientists 2019*, 575.

Epri, W. (2022). Text Classification With Naïve Bayes. *Teknologipintar.Org*, *2*(4).

Kurniawan, F. (2010). *POTENSI WISATA KULINER DALAM PENGEMBANGAN PARIWISATA DI YOGYAKARTA*.

Finandra, S., & Hamami, F. (2021). PENERAPAN ANALISIS SENTIMEN MELALUI DATA INSTAGRAM UNTUK MENGETAHUI REPUTASI WISATA KULINER DI KOTA BANDUNG MENGGUNAKANMETODE KLASIFIKASI NAÏVE BAYES. *E-Proceeding of Engineering*. https://www.instagram.com

Fithriya, D. N. L. (2020). CAPAIAN UNGGAHAN KONTEN AKUN INSTAGRAM GEMBIRA LOKA ZOO (GLZOO) YOGYAKARTA TERHADAP ONLINE ENGAGEMENT PADA MASA PANDEMI COVID-19. *Sosiologi Reflektif*, *15*(1).

Gao, X., Tan, R., & Li, G. (2020). Research on Text Mining of Material Science Based on Natural Language Processing. *IOP Conference Series: Materials Science and Engineering*, *768*(7). https://doi.org/10.1088/1757-899X/768/7/072094

Gupta, A., Singh, A., Pandita, I., & Parashar, H. (2019). Sentiment analysis of twitter posts using machine learning algorithms. *Proceedings of the 2019 6th International Conference on Computing for Sustainable Global Development, INDIACom 2019*, 980–983. https://doi.org/10.21541/apjes.939338

indonesia.go.id. (2023). *Rapor Biru Pariwisata Nasional*. https://indonesia.go.id/kategori/editorial/7771/rapor-biru-pariwisata-nasional?lang=1#:~:text=Pada%20triwulan%20III%2D2023%20nilai,pariwisata%20sebesar%203%2C76%20persen.

JogjaDataku. (2023). *Data Kinerja Dinas Pariwisata*. https://bappeda.jogjaprov.go.id/dataku/data_dasar/index/603-data-kinerja-dinas-pariwisata

Kusuma, R. K. C., Sulistyo, B. W., & Rachim, A. M. (2020). Desain Skywalk Jalan Malioboro-Stasiun Tugu Yogyakarta. *Seminar Teknologi Perencanaan, Perancangan, Lingkungan, Dan Infrastruktur II*.

Leong, J. Y., & M, B. P. (2020). SYMPTOM-BASED DISEASE PREDICTION SYSTEM USING MACHINE LEARNING. *Journal of Theoretical and Applied Information Technology*, *15*, 19. www.jatit.org

mediakeuangan. (2023). *Kian Melesat di 2023, Pariwisata Indonesia Bersiap Menuju Level Prapandemi*. https://mediakeuangan.kemenkeu.go.id/article/show/kian-melesat-di-2023-pariwisata-indonesia-bersiap-menuju-level-prapandemi

Mehra, L., Kumar Gupta, M., & Bhatt Guruji, M. (2014). An Effectual and Secure Approach for the Detection and Efficient Searching of Network Intrusion Detection System (NIDS). *International Journal of Computer Applications*, *108*(15), 37–41. https://doi.org/10.5120/18990-0442

Murnawan, M. (2017). PEMANFAATAN ANALISIS SENTIMEN UNTUK PEMERINGKATAN POPULARITAS TUJUAN WISATA. *Jurnal Penelitian Pos Dan Informatika*, *7*(2), 109. https://doi.org/10.17933/jppi.2017.070203

Purba, H., & Irwansyah, I. (2022). User Generated Content dan Pemanfaatan Media Sosial Dalam Perkembangan Industri Pariwisata: Literature Review. *Jurnal Professional*, *9*(2), 229–238.

Purwaningsih, O., Buchory, M. S., & Triwahana, T. (2020). Pemberdayaan Kelompok Masyarakat "Gardu Action" dalam Pengelolaan Sampah untuk Mewujudkan Kawasan Wisata Pantai Parangkusumo yang Bersih. *Jurnal Pengabdian Kepada Masyarakat*, *11*(4), 427–431. http://journal.upgris.ac.id/index.php/e-dimas

Putu, N. L. P. M., Ahmad Zuli Amrullah, & Ismarmiaty. (2021). Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *5*(1), 123–131. https://doi.org/10.29207/resti.v5i1.2587

Ramos, J. (2003). *Using TF-IDF to Determine Word Relevance in Document Queries*.

Rani, S., & Jain, A. (2023). Optimizing healthcare system by amalgamation of text processing and deep learning: a systematic review. *Multimedia Tools and Applications*. https://doi.org/10.1007/s11042-023-15539-y

Renjith, S., Sreekumar, A., & Jathavedan, M. (2022). An empirical research and comparative analysis of clustering performance for processing categorical and numerical data extracts from social media. *Acta Scientiarum - Technology*, *44*. https://doi.org/10.4025/ACTASCITECHNOL.V44I1.58653

Riswandha, Y., Kegiatan, P., Terhadap, W., Penggunaan, P., Di, L., Tawangmangu, K., Karanganyar, K., Riswandha, Y., & Wahyono, H. (2017). *TEKNIK PWK (Perencanaan Wilayah Kota) Corres ponding Author*. *6*(2), 131–141. http://ejournal3.undip.ac.id/index.php/pwk

Riyanto, G. P., & Pertiwi, W. K. (2023). *Apa Saja yang Terjadi di Internet Setiap 1 Menit?* . https://tekno.kompas.com/read/2023/04/25/07000077/apa-saja-yang-terjadi-di-internet-setiap-1-menit

Robbins, S. P., & Singer, J. B. (2014). From the editor-the medium is the message: Integrating social media and social work education. In *Journal of Social Work Education* (Vol. 50, Issue 3, pp. 387–390). Routledge. https://doi.org/10.1080/10437797.2014.916957

Saputra, D. (2020). Tatakelola Kolaborasi Pengembangan Kampung Wisata Berbasis Masyarakat. In *Jurnal Ilmu Pemerintahan* (Vol. 13, Issue 2).

Satriajati, S., Bagus Panuntun, S., & Pramana, S. (2020). Seminar Nasional Official Statistics 2020: Statistics in the New Normal: A Challenge of Big Data and Official Statistics IMPLEMENTASI WEB SCRAPING DALAM PENGUMPULAN BERITA KRIMINAL PADA MASA PANDEMI COVID-19 Studi Kasus: Situs Berita detik.com. *Seminar Nasional Official Statistics 2020: Statistics in the New Normal: A Challenge of Big Data and Official Statistics*.

Utama, I. P. H. B., & Giantari, I. G. A. K. (2020). PERAN CITRA DESTINASI MEMEDIASI PENGARUH E-WOM TERHADAP NIAT BERKUNJUNG KEMBALI WISATAWAN (Studi Pada Obyek Wisata Taman Edelweis Bali). *E-Jurnal Manajemen Universitas Udayana*, *9*(4), 1230. https://doi.org/10.24843/ejmunud.2020.v09.i04.p01

Wahyuni, S. (2020). ANALISIS POLA DAYA TARIK WISATA BERDASARKAN POTENSI SUMBERDAYA (SUPPLY) SEBAGAI ASET DAN DAYA TARIK DI DAERAH ISTIMEWA YOGYAKARTA. *Kepariwisataan : Jurnal Ilmiah*, *14*(1).

Wijayanti, A. (2021). Efektivitas Instagram dalam Meningkatkan Minat Kunjungan Wisatawan di Daerah Istimewa Yogyakarta. *Indonesian Journal of Tourism and Leisure*, *2*(1), 26–39. https://doi.org/10.36256/ijtl.v2i1.138

Wikipedia. (2023). *Instagram*. https://id.wikipedia.org/wiki/Instagram#cite_note-208

Zaki, M. J. (2002). Parallel and distributed data mining: An introduction. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *1759*, 1–23. https://doi.org/10.1007/3-540-46502-2_1

LAMPIRAN

Lampiran 1 Sintaks

Lampiran 2 Datasets