

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Penelitian ini dilakukan tidak terlepas dari hasil penelitian-penelitian terdahulu yang pernah dilakukan sebagai bahan perbandingan dan kajian. Adapun hasil-hasil penelitian yang dijadikan perbandingan tidak terlepas dari topik penelitian yaitu mengenai analisis sentiment. Tabel tersebut dapat memberikan gambaran menyeluruh tentang perkembangan penelitian, perbandingan antara metodologi yang digunakan, serta temuan yang telah dihasilkan oleh peneliti sebelumnya.

Table 2. 1 Tabel Penelitian Terdahulu

Penelitian Terdahulu 1	
Judul	Analisis Sentimen Pada Review Objek Wisata Dunia Fantasi menggunakan Algoritma K-Nearest Neighbor (K-NN)
Nama Penulis	Retno Sari
Jurnal	Evolusi: Jurnal Sains dan Manajemen [25]
Tahun	2020
Permasalahan	Banyaknya review mengenai tempat wisata yang telah dikunjungi, memudahkan pengunjung lain yang ingin berkunjung ke tempat tersebut mengetahui keadaan di tempat tersebut dan menjadikan masukan untuk pengunjung lain untuk memutuskan ke tempat tersebut atau tidak.
Algoritma	K-Nearest Neighbor (K-NN)
Temuan	Pada penelitian ini preprocessing yang digunakan yaitu tokenize, filter stopword dan bi-gram. Data yang digunakan terdiri dari 50 review positif dan 50 review negatif. Eksperimen telah dilakukan dengan mengubah nilai k dan didapat akurasi yang terbesar dengan nilai k=7 dengan akurasi 77.01, precision 92.38, recall 61.56 dan nilai AUC 0.894
Pembahasan	Pada penelitian ini review taman rekreasi dunia fantasi menggunakan K-Nearest Neighbor dengan teks berbahasa Indonesia menghasilkan akurasi sebesar 77.01% dengan nilai K 7 menggunakan K- fold cross validation
Relevansi	Penelitian ini bertujuan untuk mencari tahu ulasan masyarakat tentang objek wisata Dunia Fantasi, relevansi dengan penelitian ini adalah menggunakan Data Preparation yang terdiri dari casefolding, cleaning, tokenizing, filtering stopword, stemming, dan juga hasil yang didapatkan berupa akurasi.
Penelitian Terdahulu 2	
Judul	Analisis Sentimen Wisatawan Melalui Data Ulasan Candi Borobudur di Tripadvisor Menggunakan Algoritma Naïve Bayes Classifier
Nama Penulis	Yerik Afrianto Singgalen
Jurnal	Building of Informatics, Technology and Science (BITS) [26]
Tahun	2022
Permasalahan	Terdapat kesenjangan dalam studi tentang analisis sentimen wisatawan yang terbatas pada implementasi metode klasifikasi tanpa menghubungkan hasil klasifikasi untuk menghasilkan sejumlah rekomendasi berdasarkan kata populer dalam kelas positif maupun negatif untuk mengoptimalkan manajemen destinasi wisata.

Algoritma	Naïve Bayes
Temuan	Metode klasifikasi yang digunakan dalam analisis sentimen ialah Naive Bayes Classifier (NBC) terhadap 3850 ulasan pengunjung di Candi Borobudur. Data ulasan diambil dari halaman website Tripadvisor yang telah disaring berdasarkan bahasa, waktu mengulas, dan karakteristik perjalanan untuk dianalisis secara komprehensif preferensi wisatawan mancanegara. Tahapan penelitian ini terbagi menjadi tiga bagian yakni tahapan persiapan data, pemrosesan data, analisis sentimen dan evaluasi performa algoritma.
Pembahasan	Adapun, SMOTE Upsampling digunakan untuk menyeimbangkan data. Penelitian ini menunjukkan bahwa hasil implementasi metode klasifikasi Naïve Bayes Classifier (NBC) diperoleh nilai accuracy sebesar 96,36% dan nilai precision sebesar 93,23% serta nilai recall sebesar 100% dengan nilai Area Under Curve (AUC) sebesar 0,714. Selain itu hasil perankingan lima kata yang populer dari data ulasan menunjukkan adanya sorotan kondisi fisik candi, pemandangan dan aktivitas kunjungan wisata di Candi Borobudur. Terdapat empat kata paling populer dalam ulasan pengunjung ialah temple, visit, borobudur, sunrisedanplace
Relevansi	Penelitiannya menggunakan metode Naive Bayes Classifier untuk mendapatkan informasi ulasan di Objek Wisata Gunung Bromo dan memantau perkembangan di tiap tempat wisata yang ada, dengan adanya sistem ini diharapkan mempermudah pengelola Taman Nasional Bromo Tengger Semeru dalam mendapatkan informasi. Dan dapat cepat menentukan langkah apa yang harus diambil kedepannya.
Penelitian Terdahulu 3	
Judul	Analisis Sentimen E-Wallet Pada Google Play Menggunakan Algoritma Naive Bayes Berbasis Particle Swarm Optimization
Nama Penulis	Suwanda Aditya Saputra
Jurnal	Jurnal Resti [27]
Tahun	2021
Permasalahan	Pelanggan atau klien yang merasa tidak puas dengan layanan atau produk yang ditawarkan oleh sebuah perusahaan biasanya akan menuliskan keluhannya di media sosial atau ulasan di Google play. Akan tetapi, memantau dan mengorganisasi opini dari masyarakat juga bukanlah hal yang mudah. Oleh sebab itulah, diperlukan sebuah metode atau teknik khusus yang mampu mengkategorikan review-review tersebut secara otomatis, apakah termasuk positif atau negatif.
Algoritma	Naive Bayes
Temuan	Di dalam penelitian ini, akan dibahas tahapan yang dilalui untuk melakukan proses analisis sentimen terhadap komentar tentang aplikasi OVO pada Google Play. Dimulai dari tahap preprocessing sampai tahap analisis sentimen dengan Naive Bayes Classifier dan Feature selection Particle Swarm Optimization (PSO) serta bagaimana mengukur kualitas hasil analisis menggunakan tanpa feature selection dan dengan feature selection.
Pembahasan	Algoritma yang digunakan dalam penelitian ini adalah Naive Bayes Classifier (NB), dengan optimasi penggunaan Feature Selection (FS) Particle Swarm Optimization. Hasil dari cross validation NB tanpa FS adalah 82,30% untuk accuracy dan 0,780 untuk AUC. Sedangkan untuk NB dengan FS adalah 83,60% untuk accuracy dan 0,801 untuk AUC. Peningkatan sangat signifikan dengan penggunaan <i>Feature Selection (FS) Particle Swarm Optimization</i>
Relevansi	Relevansi dengan penelitian ini adalah pada bagian tahapan penelitian terbagi menjadi tiga bagian yakni tahapan persiapan data, pemrosesan data, analisis sentimen dan evaluasi performa algoritma
Penelitian Terdahulu 4	
Judul	Analysing TripAdvisor reviews of tourist attractions in Phuket, Thailand

Nama Penulis	Viriya Taecharungroj
Jurnal	Tourism Management 75 [28]
Tahun	2019
Permasalahan	Pengelola Wisata di Phuket Thailand tidak mengetahui ulasan Masyarakat tentang tempat wisata tersebut sehingga tidak bisa memperbaiki apa yang kurang dari tempat wisata tersebut.
Algoritma	Naive bayes
Temuan	Penelitian ini mempelajari ulasan TripAdvisor tempat wisata, termasuk pantai, pulau, kuil, jalan pejalan kaki, dan pasar di Phuket, Thailand. Secara total, 65.079 ulasan online dianalisis menggunakan dua teknik pembelajaran mesin: alokasi Dirichlet laten (LDA) dan pemodelan Naïve Bayes.
Pembahasan	Pemodelan LDA membantu peneliti menentukan dimensi setiap jenis atraksi. Empat dimensi ditentukan untuk pantai dan pulau, tiga dimensi untuk jalan pejalan kaki dan kuil, dan dua dimensi untuk pasar. Penelitian ini juga mengembangkan dua alat praktis – analisis arti-penting-valensi dimensional (DSVA) dan analisis arti-penting leksikal-valensi (LSVA) – dan menggunakannya untuk menyarankan tindakan bagi Otoritas Pariwisata Thailand (TAT).
Relevansi	Relevansi dengan penelitian ini adalah menggunakan metode Naive Bayes Classifier yang dapat diterapkan pada proses analisis sentimen pada pariwisata di Thailand menggunakan data latih dari ulasan-ulasan pada website TripAdvisor. Data latih juga harus dilabeli kelas sentimennya oleh pakar untuk proses klasifikasi. Dalam penyempurnaan penelitian ini maka penelitian selanjutnya disarankan untuk memperhatikan singkatan, gabungan dua kata atau lebih (bigram, trigram, n-gram)
Penelitian Terdahulu 5	
Judul	From digitization to the age of acceleration: On information technology and tourism
Nama Penulis	Zheng Xiang
Jurnal	Tourist Management Perspective [29]
Tahun	2020
Permasalahan	Sama halnya dengan review penilaian tempat tujuan wisata, para wisatawan yang susah pernah mengunjungi tempat tersebut biasanya melakukan review dengan membuat sebuah status atau pernyataan terhadap tempat yang mereka kunjungi, penilaian tersebut bisa saja positif, negatif ataupun bahkan netral atau biasa saja. Penilaian wisatawan terhadap tempat kunjungan wisata ini menjadi salah satu hal yang penting karena dapat dijadikan sebagai bagian dari pendukung keputusan baik itu wisatawan maupun pengelola tempat tersebut
Algoritma	Naïve bayes, Decision Tree
Temuan	Pada penelitian ini model sentimen analysis diusulkan sebagai solusi untuk mengatasi permasalahan. Tujuan dari penelitian ini adalah mencari model sistem untuk memberikan sebuah informasi pendukung keputusan bagi para wisatawan dan pengelola tempat wisata untuk dijadikan sumber informasi terhadap tempat wisata yang ada. Metode penelitian yang digunakan adalah dengan melakukan eksperimen untuk mendapatkan model yang diharapkan. Pada penelitian ini metode Naïve Bayes dan Decision Tree diterapkan untuk klasifikasi teks sehingga mendapatkan model terbaik yang dihasilkan untuk diimplementasikan pada sistem sehingga model yang didapatkan dapat dijadikan sebuah sistem pendukung keputusan untuk pengembangan sistem cerdas pada pihak terkait dalam upaya peningkatan nilai jual potensi daerah khususnya bidang pariwisata.
Pembahasan	Hasil penelitian menunjukkan bahwa model yang didapatkan setelah dilakukan eksperimen didapatkan tingkat akurasi naive bayes menghasilkan 77,50% lebih baik dibandingkan dengan menggunakan Decision Tree yang menghasilkan tingkat akurasi 60,83%.

Relevansi	Relevansi dengan penelitian ini adalah Analisis sentimen digunakan untuk menganalisis emosi dari semua komentar dan mengklasifikasikan ke dalam semua sentimen. Metode naive bayes digunakan untuk membagi atau memisahkan teks yang mengandung emosi positif atau negatif. Dalam metode naive bayes hasil klasifikasi akan tergantung pada probabilitas nilai frekuensi kemunculan semua dokumen karena komentar yang diberikan tergolong singkat maka perlu dilakukan optimasi dengan menggunakan query expansion recognition of synonym.
Penelitian Terdahulu 6	
Judul	Sentiment Analysis in Trip Advisor
Nama Penulis	Ana Valdivia
Jurnal	IEEE Computer Society [30]
Tahun	2020
Permasalahan	Dari hasil pengamatan yang telah dilakukan, banyak pengunjung objek wisata yang sudah memberikan ulasannya di Google Maps. Hal ini menjadi informasi yang berharga untuk meningkatkan kualitas pariwisata. Dengan ulasan tersebut kita bisa melakukan analisa untuk menemukan sentimen apa yang diberikan oleh pengunjung.
Algoritma	Naïve Bayes
Temuan	Metodologi yang digunakan dalam penelitian ini adalah kerangka kerja IS Research Alan Hevner. Dalam melakukan riset sentimen analysis objek wisata. Metode yang digunakan untuk klasifikasi adalah Naïve Bayes. Sebelum melakukan klasifikasi, dilakukan tahap pre-processing yang terdiri dari casefolding, tokenizing, filtering, stemming, dan tahap pembobotan kata menggunakan TF-IDF.
Pembahasan	Berdasarkan penelitian yang dilakukan, disimpulkan bahwa sistem dapat mengklasifikasikan kelas sentimen ulasan objek wisata yang terdapat pada Google Maps menggunakan metode Naive Bayes dengan nilai akurasi yang bervariasi dari setiap tempat wisata. Nilai akurasi tertinggi adalah 0,76 sedangkan terendah adalah 0,38. Hasil sentimen analisis yang dilakukan masuk dalam kategori yang positif. Hal ini berdasarkan performa metode Naive Bayes yang menunjukkan bahwa nilai rata-rata f1-score kelas positif adalah 0,73 lebih tinggi dibanding kelas netral 0,53 dan negatif 0.14
Relevansi	Relevansi dengan penelitian ini adalah Sebelum melakukan klasifikasi, dilakukan tahap pre-processing yang terdiri dari casefolding, tokenizing, filtering, stemming, dan tahap pembobotan kata menggunakan TF-IDF
Penelitian Terdahulu 7	
Judul	Analysis Of UMN Student Graduation Timeliness Using Supervised Learning Method
Nama Penulis	Christian Pangestu Kuncoro
Jurnal	IJNMT (International Journal of New Media Technology) [31]
Tahun	2021
Permasalahan	Pendidikan merupakan salah satu hal terpenting dalam kehidupan manusia, dan dalam dunia pendidikan. Namun masih banyak mahasiswa yang lulus tidak tepat waktu.
Algoritma	Naïve Bayes
Temuan	Metode penelitian yang digunakan adalah penelitian kuantitatif, karena mengukur suatu permasalahan data melalui angka-angka, dan juga mengukur sebagai kata-kata deskriptif yang lulus tepat waktu, atau tidak. Data dapat diubah dalam bentuk statistik dan diperhitungkan dalam pengambilan suatu solusi dan metode yang digunakan bukan dari kuisisioner, survei, jajak pendapat, atau wawancara yang bersifat pertanyaan, jumlah partisipan dalam metode kuantitatif cenderung lebih banyak dibandingkan kualitatif.
Pembahasan	Jumlah, dan persentase lulusan program studi di UMN tahun 2018-2020 berdasarkan program studi, dengan persentase terbesar ILKOM sebanyak 1025 mahasiswa dengan persentase 28,28%, dan untuk diploma hanya 17

	orang dengan persentase 0,47%. kemudian rata-rata IPK lulusan mahasiswa UMN tahun 2018 sampai dengan tahun 2020 berdasarkan program studi, pemenangnya adalah teknik fisika dengan rata-rata IPK 3,61, jumlah dan persentase program studi lulusan UMN tahun 2018-2020 berdasarkan program studi, dengan yang terbesar persentase ILKOM sebanyak 1025 mahasiswa dengan persentase 28,28%, dan untuk diploma hanya 17 orang
Relevansi	Relevansi dengan penelitian ini adalah menggunakan framework CRISP-DM untuk melakukan analisis terhadap mahasiswa UMN yang lulus tidak tepat waktu
Penelitian terdahulu 8	
Judul	Sentiment Analysis about Indonesian Lawyers Club Television Program Using K-Nearest Neighbor, Naïve Bayes Classifier, and Decision Tree
Nama Penulis	Nico Nathanael Wilim, Raymond Sunardi Oetama
Jurnal	IJNMT (International Journal of New Media Technology) [32]
Tahun	2021
Permasalahan	Pemenang Penghargaan dihitung dari rating. Nilai rating memang penting bagi keberlangsungan sebuah acara televisi, namun tidak menunjukkan tingkat kualitas acara tersebut. Penonton seringkali memberikan pendapatnya terhadap acara televisi melalui media sosial, salah satunya adalah Twitter. merupakan ungkapan keyakinan yang dianut bersama di antara anggota suatu kelompok atau masyarakat, mengenai suatu isu kontroversial yang menyangkut kepentingan umum. Pendapat tidak selalu logis, tidak berbentuk, selalu ambivalen, kontradiktif, dan mudah diubah. Opini masyarakat terhadap ILC dapat diperoleh dari tweet yang diambil dari Twitter, hal ini dikarenakan cukup banyak juga pengguna Twitter yang menonton ILC
Algoritma	Naive Bayes, Decision Tree, KNN
Temuan	Seperti terlihat pada Tabel VII, pada tahun 2018 Naïve Bayes merupakan algoritma terbaik, sedangkan pada tahun 2019 K-NN merupakan algoritma terbaik. Artinya tidak ada algoritma yang selalu berada di atas. Semua algoritma digunakan berdasarkan isi data dan tingkat keakuratan data tertentu selalu berubah. Pada tahun 2018, kinerja Mata Najwa dan ILC berfluktuasi, akibatnya K-NN semakin kesulitan mencari tetangga terdekat. Dalam situasi seperti ini, kemampuan probabilitas bersyarat Naïve Bayes berkinerja lebih baik. Sementara pada tahun 2019, opini publik terhadap Mata Najwa jelas berada di atas ILC. Jadi K-NN dapat dengan mudah memisahkannya menjadi tetangga terdekat. Akurasi terbaik terdapat pada ILC tahun 2019 dengan menggunakan K-NN karena opini positif masyarakatnya pada tahun 2018 berada di atas tahun 2019. Sehingga K-NN dapat lebih mudah memisahkannya ke tetangga terdekat.
Pembahasan	Opini publik di Twitter berhasil diterapkan untuk mengonfirmasi pemenang Panasonic Gobel Award. Berdasarkan analisis manual, ILC menjadi pemenang di tahun 2018. Jumlah sentimen positif ILC sedikit lebih baik dibandingkan Mata Najwa. Namun ILC kalah di tahun 2019 karena jumlah sentimen positifnya berkurang sedangkan Mata Najwa jelas berada di atas ILC. Selain itu, dari tahun 2018 hingga 2019, jumlah sentimen positif terhadap ILC menurun drastis, sedangkan Mata Najwa berfluktuasi. Tiga Algoritma diterapkan untuk memvalidasi hasil pelabelan manual dengan akurasi tertinggi sebesar 76,94% dengan menggunakan K-NN. Namun, tidak ada satu algoritma pun yang menunjukkan performa terbaik pada semua data. Pada tahun 2018, Naïve Bayes menjadi algoritma terbaik, sedangkan pada tahun 2019, K-NN menjadi algoritma terbaik.
Relevansi	Relevansi dengan penelitian ini adalah bertujuan untuk melakukan analisis sentimen dengan mengklasifikasikan sentimen tersebut ke dalam sentimen positif dan negatif.
Penelitian terdahulu 9	

Judul	Analisis Sentimen Objek Wisata Di Provinsi Sulawesi Selatan Berdasarkan Ulasan Pengunjung Menggunakan Metode Random Forest Classifier
Nama Penulis	Aisyah Nurul Izza
Jurnal	Jurnal Sistem Informasi, Teknologi Informasi, dan Edukasi Sistem Informasi (JUST-SI) [33]
Tahun	2022
Permasalahan	Permasalahan yang dihadapi oleh Pemerintah Sulawesi Selatan dalam sektor pariwisatanya diantara yaitu karena terus meningkatnya jumlah wisatawan dan objek wisata dapat menyulitkan pemerintah setempat untuk memantau perkembangan objek wisata terbukti dari salah satu berita dari medcom.id bahwa “tempat wisata di sulawesi selatan belum didukung akses memadai”. Sehingga dibutuhkan ulasan pengunjung objek wisata untuk pengembangan dan pemeliharaan objek wisata yang ada. Dengan ulasan tersebut kita dapat menemukan sentimen apa yang diberikan pengunjung terhadap objek wisata yang ada.
Algoritma	Random Forest Classifier
Temuan	Penelitian ini terdapat 3 parameter pengujian yang akan diuji yaitu: pengujian jumlah tree, pengujian kedalaman tree, pengujian jumlah data, dan pengujian algoritme random forest menggunakan k-fold cross validation. Pengujian ini menggunakan metode confusion matrix yang akan memberikan keluaran nilai accuracy, precision, recall dan f1-score. Pengujian ini bertujuan untuk mengetahui pengaruh jumlah kedalaman tree dengan performa terbaik dan bagaimana pengaruh perubahan jumlah kedalaman tree terhadap akurasi. Pengujian menggunakan jumlah kedalaman tree yang dipilih secara acak sebesar 30, 40, dan 50 dengan jumlah tree sebesar 300 kemudian menggunakan data latih sebanyak 90% dan data uji 10%. Data yang digunakan sebelumnya telah melewati proses text preprocessing. Tabel 4 dan Gambar 3 akan memperlihatkan hasil pengaruh kedalaman tree terhadap akurasi akhir. Pada Gambar 3 dapat dilihat bahwa rata-rata akurasi tertinggi pada pengujian terdapat pada jumlah kedalaman tree sebesar 50 dengan rata-rata akurasi sebesar 86% hal ini disebabkan karena semakin banyak kedalaman tree dalam pengklasifikasian maka semakin banyak proses splitting data. Splitting data merupakan proses yang dilakukan untuk mendapatkan aturan baru berisikan informasi mengenai data yang digunakan.
Pembahasan	Metode random forest dapat diimplementasikan pada proses analisis sentimen objek wisata di Provinsi Sulawesi Selatan menggunakan data dari ulasan-ulasan yang diberikan oleh pengunjung pada website TripAdvisor. Data yang digunakan harus dilabeli dengan kelas sentimen positif, netral atau negatif. Proses awal klasifikasi sentimen pada penelitian ini yaitu memberikan pelabelan data terhadap ulasan pengunjung, pelabelan ini meliputi positif, netral dan negatif. Setelah pelabelan, selanjutnya adalah proses text preprocessing yang meliputi tahapan case folding, tokenizing, Stopword Removal, dan stemming. Setelah proses text preprocessing selanjutnya adalah pembobotan kata menggunakan metode TF-IDF dan dilanjutkan untuk pengujian algoritme random forest menggunakan k-fold cross validation.
Relevansi	Relevansi dengan penelitian ini adalah menggunakan analisis sentimen untuk melakukan klasifikasi ulasan dan juga menggunakan k-fold cross validation dan confusion matrix.
Penelitian terdahulu 10	
Judul	Implementasi Algoritma K-Nearest Neighbors Pada Analisis Sentimen Masyarakat Terhadap Penerapan Pemberlakuan Pembatasan Kegiatan Masyarakat
Nama Penulis	Vitra bayu Anwari
Jurnal	Sistem Komputer dan Teknik Informatika [34]
Tahun	2022

Permasalahan	Pembatasan Sosial Berskala Besar (PSBB) yang berdampak pada berbagai sektor kehidupan di Indonesia. Namun seiring dengan berjalannya waktu kebijakan ini dinilai kurang efektif dalam menangani masalah virus Covid-19 yang terus meningkat dan kemudian diterapkan kebijakan lainnya yaitu Pemberlakuan Pembatasan Kegiatan Masyarakat (PPKM) yang diterapkan di wilayah zona merah saja, semua pihak berharap kebijakan ini adalah kebijakan yang lebih baik dari sebelumnya dalam menekan penyebaran virus Covid-19 dan sampai saat ini terbukti lebih efektif ketimbang PSBB. Dari berbagai kebijakan yang diterapkan pemerintah DKI Jakarta tentu banyak menimbulkan opini masyarakat yang bernilai positif maupun negatif yang disampaikan melalui pesan tweet pada media sosial twitter.
Algoritma	K-Nearest Neighbors
Temuan	Dari data tweet yang berisi opini masyarakat tersebut kemudian dikumpulkan untuk dilakukan proses analisis sentimen dengan menerapkan metode Cross-Industry Standard Process for Data Mining (CRISP-DM) untuk mempermudah dalam menjalankan tahapan proses text mining. Peneliti juga menerapkan algoritma k-nearest neighbors untuk pengolahan dan pengelompokan data berdasarkan jarak terdekat antara data testing dan data training. Kemudian Tahapan data preprocessing yaitu tahapan dimana peneliti mulai melakukan filter data tahap akhir dan merubah format data berdasarkan format yang dibutuhkan sebelum dilakukan proses perhitungan sebaran data yang sudah dipecah ke dalam data latih dan data percobaan dengan algoritma perbandingan KNN, naïve bayes, dan decision tree dengan nilai perbandingan 60:40, 70:30, 80:20. Berdasarkan hasil perhitungan yang dilakukan hasil yang didapatkan precision positif adalah 75.86%, precision negatif menghasilkan nilai sebesar 93.02%, hasil nilai recall sentimen negatif sebesar 85.11%, hasil nilai recall positif bernilai 88.00%, jadi kesimpulannya mengapa hasil tidak begitu akurat karena data training dan data testing yang dimiliki tidak seimbang, dan nilai accuracy yang dihasilkan adalah sebesar 86.11%, sehingga dapat disimpulkan berdasarkan hasil perhitungan dan perbandingan bahwa algoritma k-nearest neighbors dapat mengklasifikasikan dataset sentimen terkait penerapan PPKM di wilayah DKI Jakarta dengan baik.
Pembahasan	Berdasarkan hasil yang didapatkan dalam analisis sentimen yang dilakukan dengan menerapkan algoritma k-nearest neighbors memiliki nilai akurasi tertinggi yaitu sebesar 86.11% dengan nilai rasio perbandingan 80:20 dan nilai parameter k=7 yang sudah dibandingkan dengan hasil perhitungan algoritma naïve bayes dan decision tree terhadap dataset berupa data tweet hasil crawling menggunakan rapid miner yang memiliki label nilai sentimen yang dilakukan oleh pakar, kemudian hasil dari perhitungan evaluasi menggunakan metode confusion matrix memiliki hasil nilai precision sebesar 75.86%, recall sebesar 88.00%, dan nilai accuracy 86.11% terhadap total data sebanyak 365 data dan sebaran data positif sebanyak 131 dan data negatif sebanyak 234 data, Maka dapat disimpulkan berdasarkan hasil yang didapatkan pada penelitian ini dengan menerapkan algoritma k-nearest neighbors terkait penerapan PPKM di wilayah DKI Jakarta bahwa hasil yang didapatkan sentimen negatif lebih banyak ketimbang sentimen positif yang berarti penerapan PPKM di wilayah DKI Jakarta masih dianggap belum maksimal berdasarkan opini masyarakat yang disampaikan melalui aplikasi twitter.
Relevansi	Relevansi dengan penelitian ini adalah menggunakan metode CRISP-DM sebagai metode penelitian, menggunakan text mining, dan data mining.

Berdasarkan penelitian terdahulu yang dilakukan oleh peneliti sebelumnya yang berkaitan dengan topik analisa sentiment terdapat beberapa perbedaan antara penelitian ini dengan penelitian terdahulu yaitu penelitian-penelitian terdahulu fokus pada objek wisata yang berbeda-beda, seperti Dunia Fantasi, Candi Borobudur, Phuket (Thailand), dan lainnya [25]. Sementara itu, penelitian yang akan dilakukan fokus pada objek wisata Gunung Bromo di Indonesia. Meskipun banyak penelitian terdahulu menggunakan algoritma analisis sentimen seperti Naive Bayes, K-Nearest Neighbor (K-NN), dan Random Forest Classifier, penggunaan metode CRISP-DM yang akan digunakan dalam penelitian tersebut mungkin menjadi perbedaan signifikan [26]. Penelitian sebelumnya mungkin tidak secara eksplisit menggunakan kerangka kerja CRISP-DM [27].

Tujuan dari setiap penelitian terdahulu juga berbeda-beda, mulai dari mencari tahu ulasan masyarakat tentang objek wisata untuk pengembangan destinasi wisata, hingga mengevaluasi kebijakan pemerintah terkait pembatasan kegiatan masyarakat [28]. Dalam hal ini, penelitian yang akan dilakukan akan fokus pada analisis sentimen terhadap ulasan pengunjung Gunung Bromo dengan tujuan memberikan wawasan bagi pengelola destinasi wisata [29].

Penelitian terdahulu menggunakan berbagai teknik preprocessing seperti tokenization, filtering stopword, stemming, dan TF-IDF untuk mempersiapkan data ulasan [30]. Demikian juga, penelitian yang akan dilakukan akan menggunakan teknik preprocessing serupa untuk memastikan kualitas data yang baik sebelum dilakukan analisis sentiment [31].

Hasil dan evaluasi dari penelitian terdahulu menunjukkan akurasi, precision, recall, dan f-1 yang bervariasi tergantung pada metode dan algoritma yang digunakan [32]. Penelitian yang akan dilakukan juga akan mengevaluasi hasilnya berdasarkan metrik-metrik yang serupa [33]. Dengan mempertimbangkan perbedaan-perbedaan tersebut, penelitian yang akan dilakukan diharapkan dapat memberikan kontribusi yang berharga dalam pemahaman sentimen pengunjung terhadap Gunung Bromo, serta memberikan wawasan yang berguna bagi pengelola destinasi wisata [34].

2.2 Teori Tentang Topik

2.2.1 Gunung Bromo

Salah satu gunung berapi aktif di Indonesia, Gunung Bromo terletak di Jawa Timur dan dikelilingi oleh kabupaten Probolinggo, Lumajang, Malang, dan Pasuruan. Sejak aktifnya Gunung Bromo, kini menjadi salah satu objek wisata di Jawa Timur yang banyak disukai dan tidak selalu sepi [35]. Menjadi gunung berapi yang aktif, mudah untuk didaki, dan keunikan Kawah Bromo yang terletak di tengah kaldera Gunung Tengger yang sangat besar dan dikelilingi oleh pegunungan dan lautan pasir menjadi daya tarik sebagian besar orang untuk datang ke Gunung Bromo. Berdasarkan catatan sejarah, Kerajaan Majapahit pernah mengalami serangan dari berbagai arah di masa lalu, yang membuat penduduk asli kebingungan dan tidak dapat melakukan relokasi. Akhirnya kerajaan ini terbagi menjadi dua wilayah yaitu wilayah pertama terletak di Bali, sedangkan wilayah kedua berpusat di Gunung Bromo [36].

Satu-satunya kawasan lindung di Indonesia Taman Nasional Bromo Tengger Semeru (TNBTS), memiliki keunikan karena berbentuk lautan pasir dan luasnya mencapai 5.250 hektar. Kawasan Wisata Gunung Bromo (KWGB) merupakan salah satu destinasi wisata yang berada di bawah kepemilikan kawasan TNBTS, dengan fenomena alam yang sangat menarik [37]. Lokasi terbaik untuk melihat fajar atau yang sering diistilahkan dengan matahari terbit adalah di puncak pendakian Gunung Bromo. Untuk menyaksikan pemandangan matahari terbit yang menakjubkan dari Penanjakan Gunung Bromo, pengunjung harus menyewa kendaraan dari penduduk setempat dan berangkat dari akomodasi sekitar pukul tiga pagi untuk sampai di puncak Gunung Bromo [38].

Setelah menyaksikan matahari terbit, para wisatawan dapat mendaki Gunung Pananjakan dan kemudian melanjutkan perjalanan menuju Gunung Bromo. Selain itu, wisatawan juga bisa menaiki kuda seharga Rp 70.000/ orang untuk mendaki Gunung Bromo. Dengan ketinggian Gunung Bromo yang mencapai 2.392 m di atas permukaan laut, wisatawan juga bisa melihat kawah Gunung Bromo yang berasap dan melihat ke bawah terlihat lautan pasir dengan

pura di tengahnya. Tak hanya menampilkan keindahan matahari. Beberapa sudut pandang yang bisa dilihat di kawasan Gunung Bromo antara lain Strawberry Garden, Bukit Teletubbies, Pasir Berbisa, dan Kawah Gunung Bromo [39]. Sudut pandang Bukit Teletubbies merupakan padang savana yang tertutup Bukit Barisan. Jalur menuju Bukit Teletubbies atau disebut juga Savannah ini melewati jalan sempit dan jalan berbatu. Saat berkendara melewati gurun pasir, tidak ada satu pohon atau rumput, cuaca di Bukit Teletubbies sangat panas. Suhu di puncak Gunung Bromo 2° biasanya mencapai antara derajat Celcius hingga 20°C, sehingga bisa menghangatkan diri saat merasa kedinginan [40].

Bagi Wisatawan yang ingin bermalam dapat mencari akomodasi berupa *guesthouse* atau hotel yang terletak tidak jauh dari objek wisata Gunung Bromo. *Guesthouse* menawarkan berbagai jenis paket. Wisatawan dapat memilih paket berdasarkan biaya yang diberikan. Secara umum, tarif berkisar dari terjangkau hingga mahal berdasarkan permintaan pengunjung. Setiap lokasi wisata mempunyai kelebihan dan kekurangan, tidak terkecuali Gunung Bromo. Keuntungan mengunjungi Gunung Bromo antara lain bisa menikmati pemandangan yang menakjubkan, matahari terbit langsung dari dataran tinggi, dan pemandangan perbukitan yang memikat. Kerugiannya antara lain ketidakmampuan untuk menjaga area tersebut tetap bersih, pembuangan sampah sembarangan, kerusakan jalan, dan potensi ketidaknyamanan yang timbul saat bepergian dengan jeep. Gunung Bromo adalah salah satu destinasi wisata yang paling populer di Indonesia, menarik ribuan pengunjung setiap tahunnya untuk menikmati keindahan alamnya yang menakjubkan dan pengalaman petualangan yang tak terlupakan [41].

2.2.2 Analisis Sentimen

Analisis sentimen merupakan proses komputasi menggunakan teknik analisis teks dalam bentuk data tekstual dengan cara mengelola, memahami, dan menggolongkan emosional baik positif atau negatif [42]. Analisis sentimen banyak digunakan karena meningkatnya kebutuhan individu atau kelompok dalam mengetahui pendapat seseorang terhadap sesuatu. Manfaat adanya analisis sentimen yaitu sebagai evaluasi dan ide pada berbagai

bidang. Analisis sentimen dapat menganalisis suatu kejadian, pernyataan, dan komentar yang kontroversi. Tujuan utamanya adalah untuk memahami bagaimana orang merasa tentang suatu topik atau entitas tertentu, seperti produk, merek, layanan, acara, atau isu tertentu. Analisis sentimen dapat dilakukan dalam berbagai konteks, termasuk media sosial, ulasan produk, survei pelanggan, dan teks dari sumber-sumber lainnya [43].

Hasil dari analisis sentimen juga dapat menjadi sebuah gambaran bagi perusahaan, *public figure*, dan pemerintahan untuk menentukan langkah selanjutnya. Terdapat beberapa jenis analisis sentimen yaitu *emotion detection*, *aspectbased sentiment analysis*, dan *fine grand sentiment analysis*. *Fine sentiment analysis* adalah jenis analisis yang memiliki penilaian spesifik dan biasa digunakan pada bidang *e-commerce*. *Emoticon detection* adalah jenis analisis yang bertujuan untuk mengetahui emosi yang ada pada pesan misalnya emosi bahagia, sedih, marah, dan lain-lain. *Aspect-based sentiment analysis* merupakan jenis analisis untuk mengetahui aspek yang berpengaruh dan penilaian dari pelanggan [44].

2.2.3 Word Cloud

Wordcloud sering digunakan dalam analisis teks untuk memberikan pemahaman visual tentang konten teks dan menyoroti kata kunci atau tema-tema yang muncul dalam dokumen atau dataset teks. Ini adalah visualisasi data yang digunakan untuk menampilkan frekuensi kata-kata dalam teks secara grafis, dengan ukuran kata proporsional dengan frekuensinya. Langkah pertama adalah membersihkan dan mempersiapkan teks untuk analisis. Ini dapat mencakup langkah-langkah seperti menghapus tanda baca, mengonversi teks menjadi huruf kecil, menghapus kata-kata umum (*stopwords*), dan *stemming* atau *lemmatization* [45].

Teks dibagi menjadi bagian yang lebih kecil, seperti kata-kata atau frasa, yang disebut token. Setelah teks diberi token, perhitungan frekuensi kata dilakukan. Ini melibatkan menghitung berapa kali setiap kata muncul dalam teks. Setelah menghitung frekuensi kata, *wordcloud* digambarkan. Kata-kata dengan frekuensi tertinggi akan memiliki ukuran yang lebih

besar, sementara kata-kata dengan frekuensi terkecil akan memiliki ukuran yang lebih kecil. Biasanya, *wordcloud* dibuat dalam bentuk visual berwarna, yang memungkinkan penyesuaian ukuran dan warna kata-kata untuk menjadi lebih menarik dan lebih mudah dipahami [46].

2.2.4 TF-IDF

TF-IDF (*Term Frequency-Inverse Document Frequency*) adalah sebuah metode yang umum digunakan dalam pemrosesan teks dan information retrieval untuk mengevaluasi pentingnya sebuah kata dalam sebuah dokumen relatif terhadap sebuah kumpulan dokumen. Metode ini memberikan nilai bobot kepada kata-kata berdasarkan seberapa sering kata tersebut muncul dalam sebuah dokumen (TF) dan seberapa jarang kata tersebut muncul dalam seluruh koleksi dokumen [47].

2.3 Teori Framework dan Algoritma yang digunakan

2.3.1 Data Mining

Data Mining adalah proses penggalian informasi dan pola yang bermanfaat dari suatu data yang sangat besar. Proses *data mining* terdiri dari pengumpulan data, ekstraksi data, analisa data, dan statistik data. Teknik *data mining* digunakan untuk memeriksa basis data berukuran besar sebagai cara untuk menemukan pola yang baru dan berguna [48].

Ada dua jenis penggunaan *data mining* yaitu prediktif dan deskriptif. Ketika data mining bersifat deskriptif, yaitu mencari pola-pola yang dapat dipahami manusia yang menjelaskan fitur-fitur data. Di sisi lain, prediktif mengacu pada proses pembuatan model pengetahuan menggunakan data mining untuk membuat prediksi. Enam kelompok berikut dapat digunakan untuk mengatur pekerjaan data mining sesuai dengan fungsinya: [49]

1. *Clustering*

Data dengan label kelas yang tidak pasti dikelompokkan ke dalam sejumlah kelompok yang telah ditentukan berdasarkan metrik kesamaan. *Clustering* digunakan untuk menentukan seberapa mirip atau jauhnya suatu kumpulan data.

2. *Classification*

Dengan mengidentifikasi pola atau model yang dapat mengkarakterisasi dan membedakan suatu kelas data, klasifikasi adalah generalisasi dari struktur terkenal untuk diterapkan pada data baru.

3. *Regression*

Tujuan dari teknik regresi adalah untuk menemukan pola dalam data numerik daripada klasifikasi. Fungsi yang mendasarkan hasil pada nilai masukan adalah keluaran dari pendekatan regresi.

4. *Forecasting*

Salah satu teknik untuk memproyeksikan nilai yang akan dicapai dalam jangka waktu tertentu adalah peramalan. Data dan nilai kebisingan dari era sebelumnya digunakan sebagai dasar bahan prediksi ketika menerapkan prosedur peramalan.

5. *Sequencing*

Tujuan dari pendekatan analisis sekuensi adalah untuk menemukan pola dalam suatu rangkaian atau rangkaian kejadian.

6. *Descriptive*

Tujuan dari pendekatan deskriptif adalah untuk memperoleh pemahaman yang lebih mendalam terhadap data observasi. Mengetahui perilaku data itu sendiri adalah hasil akhirnya.

2.3.2 Text Mining

Text mining adalah proses analisis dan ekstraksi informasi yang berguna dari teks yang tidak terstruktur. Melibatkan penggunaan teknik-teknik komputasional dan algoritma untuk mengurai, memahami, dan mengekstrak pola-pola yang bermanfaat dari dokumen teks. Informasi berharga di dalam kumpulan makalah adalah tujuan dari *text mining*, yaitu penambangan data yang dilakukan melalui teks, karena penambangan teks mengubah informasi menjadi format yang dapat dimanfaatkan oleh mesin atau individu tanpa waktu untuk membaca keseluruhan teks, maka informasi tersebut harus tepat dan eksplisit [50].

Berbeda dengan *data mining*, yang mengambil input data terstruktur atau *database*, *text mining* menggunakan formulir input yang dikenal sebagai ekstraktor data terstruktur. Penambangan teks adalah proses dua langkah yang dimulai dengan memberikan struktur sumber data teks. Kemudian, dengan menggunakan teknik dan sumber daya yang sama seperti penambangan teks, mengekstraksi pengetahuan dan data terkait dari data teks terstruktur [51].

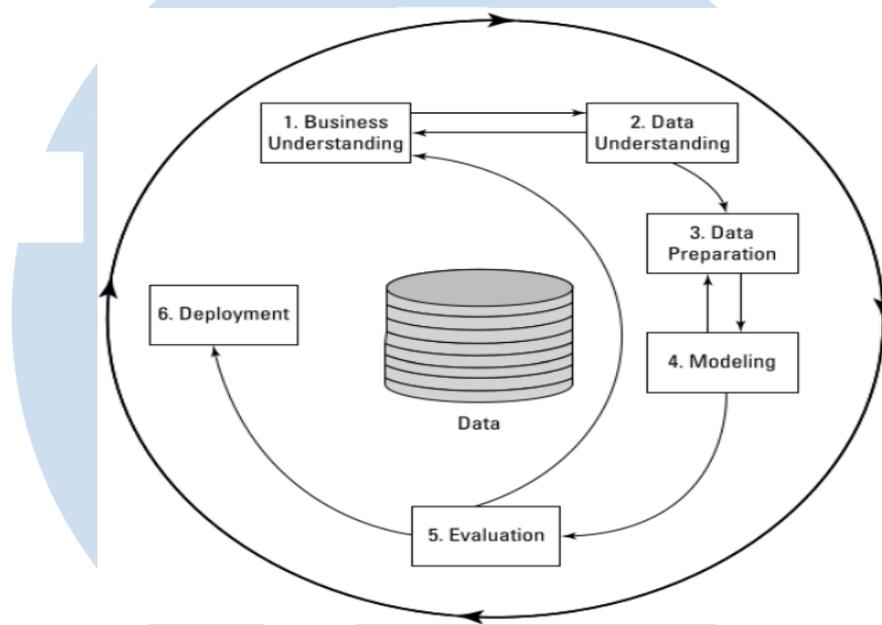
Analisis sentimen, pemrosesan bahasa alami, pemantauan media sosial, klasifikasi dokumen, ekstraksi informasi, penambangan opini, dan banyak lagi adalah beberapa aplikasi *teks mining*. Teknik-teknik ini dapat membantu organisasi merencanakan dan membuat keputusan dengan membantu teks yang tidak terstruktur.

Text mining memiliki berbagai aplikasi dalam berbagai domain, termasuk analisis sentimen, pemrosesan bahasa alami, pemantauan media sosial, klasifikasi dokumen, ekstraksi informasi, penambangan opini, dan banyak lagi. Dengan menggunakan teknik-teknik ini, organisasi dapat menggali informasi yang berharga dari teks yang tidak terstruktur untuk mendukung pengambilan keputusan dan perencanaan strategis [52].

2.3.3 CRISP-DM

Kerangka model yang digunakan dalam penambangan data disebut Proses Standar Lintas Industri untuk Penambangan Data, atau *CRISP-DM*. *CRISP-DM* menggambarkan fase dari tahapan-tahapan dalam sebuah proyek, pekerjaan yang terkait dalam tiap fase dan penjabaran terkait hubungan antar pekerjaan tersebut serta memberikan sebuah gambaran siklus hidup (*life-cycle*) dari *Data Mining*. Metode ini dikembangkan oleh sebuah kelompok kerja yang terdiri dari para profesional dan akademisi dalam bidang data mining dan analisis data. *CRISP-DM* menyediakan kerangka kerja yang terstruktur dan terdokumentasi dengan baik untuk memandu para praktisi data mining dari tahap perencanaan hingga implementasi Solusi .

Proses data mining berdasarkan CRISP-DM terdiri dari 6 tahapan yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*.



Gambar 2. 1 Framework CRISP-DM
Sumber:[53]

Berikut merupakan 6 tahapan dari framework *CRISP-DM*, yaitu: [53]

1. *Business Understanding* atau pemahaman domain yaitu pengetahuan domain atau keahlian bisnis. Memahami persyaratan dari sudut pandang bisnis dan sifat tugas data mining yang harus diselesaikan diperlukan untuk fase Pemahaman bisnis. Membuat rencana strategis dan jadwal penelitian adalah salah satu kegiatan yang melibatkan penentuan tujuan atau sasaran perusahaan, pemahaman lingkungan bisnis, dan penentuan tujuan penambangan data.
2. Proses pengumpulan data awal dan menganalisisnya untuk mengidentifikasi data yang akan digunakan disebut dengan pemahaman data. Untuk menghasilkan hipotesis awal, fase Pemahaman Data mencari masalah kualitas data dan menemukan subkumpulan data yang menarik.

3. Persiapan data, merupakan hal yang umum untuk menyebut langkah persiapan data sebagai proses yang padat karya. Memilih tabel dan kolom yang akan diubah menjadi database baru untuk konten penambahan data adalah salah satu tugas yang diselesaikan. Setelah peneliti mengembalikan data dengan label nilai sentimen, data tersebut harus menjalani pra-pemrosesan data yang meliputi pelipatan kasus, pembersihan, tokenisasi, pemfilteran stopwords, dan *stemming*. Penjelasan setiap langkah dalam proses penyiapan data disajikan di bawah ini. Berikut penjelasan tiap tahap dalam persiapan Data:

1. *Case Folding*

Salah satu teknik pra-pemrosesan teks yang disebut "*case folding*" berupaya menyamakan karakter dalam masukan. Tujuan prosedur ini adalah mengubah setiap huruf dalam dokumen menjadi huruf kecil.

2. *Cleaning*

Data teks yang berisi satu huruf, URL, dan karakter non-huruf (seperti tanda baca dan angka) dibersihkan selama langkah pembersihan. Hal ini dilakukan karena karakter tunggal, tanda baca, URL, dan karakter non-huruf dianggap sebagai elemen yang tidak berguna dan tidak penting.

3. *Tokenizing*

Pada tahap *Tokenizing* dilakukan proses *tokenizing* yaitu pemotongan teks menjadi potongan-potongan token atau fitur. *Tokenizing* pada penelitian ini menggunakan spasi, sehingga teks dipecah berdasarkan spasi. Tahap *Tokenizing* diperlukan agar mempermudah proses pada tahap selanjutnya.

4. *Filtering Stopword*

Proses pembuangan term atau kata dasar yang tidak memiliki arti atau tidak relevan. Jika *stopwords* tidak dihapus maka dapat memperbanyak fitur yang tidak bermakna dan dapat mengakibatkan penurunan performa model yang dihasilkan.

5. *Stemming*

Tahap *Stemming* digunakan untuk mengubah kata menjadi bentuk kata dasarnya. Hal ini dilakukan untuk karena kata yang seharusnya memiliki kata dasar sama namun memiliki imbuhan dianggap dua kata yang berbeda.

6. *TF-IDF*

Term Weighting (TF-IDF) merupakan proses untuk memberikan pembobotan kata dengan mencari nilai *Term Frequency (TF)*, kemudian mencari nilai *Document Frequency (DF)*, lalu mencari nilai *Invers Document Frequency (IDF)* setelah itu baru menghitung bobot.

4. *Modeling* adalah fase menentukan teknik *data mining* yang digunakan, menentukan tools data mining, teknik *data mining*, algoritma *data mining*, menentukan parameter dengan nilai yang optimal.

5. *Evaluation* adalah fase interpretasi terhadap hasil data mining yang ditunjukkan dalam proses pemodelan pada fase sebelumnya. Evaluasi dilakukan secara mendalam dengan tujuan menyesuaikan model yang didapat agar sesuai dengan sasaran yang ingin dicapai dalam fase pertama.

6. *Deployment* atau penyebaran adalah fase penyusunan laporan atau presentasi dari pengetahuan yang didapat dari evaluasi pada proses *data mining*.

2.3.4 Naive Bayes

Salah satu pendekatan pembelajaran mesin untuk masalah klasifikasi, algoritma *Naive Bayes*, banyak digunakan untuk kategorisasi teks menggunakan kumpulan data pelatihan berdimensi tinggi. Analisis sentimen, pemfilteran spam, dan klasifikasi adalah beberapa contohnya. Teknik-teknik ini terkenal karena efisiensinya dan juga kesederhanaannya. *Naive Bayes Classifier* menunjukkan kinerja yang cukup baik dalam klasifikasi teks. Metode *Naive Bayes* juga memperhitungkan informasi bobot, dengan demikian hasil klasifikasi akan lebih akurat daripada hanya menggunakan keputusan [54].

Teorema fundamental, terkadang disebut sebagai aturan *Bayes* atau hukum *Bayes*, adalah landasan dari algoritma *Naive Bayes*. Tujuannya adalah untuk menghitung probabilitas bersyarat, atau kemungkinan suatu kejadian bergantung pada pengetahuan sebelumnya yang kini dapat diakses.

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Keterangan:

X = Data dengan class yang belum diketahui

H = Hipotesis data merupakan suatu class spesifik

P(H|X) = Probabilitas hipotesis H berdasar kondisi X

P(H) = Probabilitas hipotesis H

P(X|H) = Probabilitas X berdasarkan kondisi pada hipotesis H

P(X) = Probabilitas X

Rumus 2.1 diatas merupakan rumus *Naive Bayes*. Dalam klasifikasi teks, X biasanya adalah vektor fitur yang mewakili dokumen teks (misalnya, dengan menggunakan representasi *Bag-of-Words* atau TF-IDF). y adalah label atau kelas yang ingin diprediksi. Karena asumsi dasar *Naive Bayes* bahwa setiap fitur X saling independen, meskipun ini seringkali tidak benar

dalam praktiknya. *Naive Bayes* seringkali menghasilkan hasil yang baik, terutama pada dataset yang besar dengan banyak fitur [55].

Asumsi naif pada *Naive Bayes* adalah bahwa setiap fitur (atau variabel independen) berkontribusi secara independen terhadap probabilitas kelas tertentu, meskipun kenyataannya fitur-fitur tersebut mungkin tidak benar-benar independen. Meskipun asumsi ini sering kali tidak terpenuhi di dunia nyata, *Naive Bayes* seringkali memberikan kinerja yang baik, terutama dalam data teks yang besar [56].

Dalam langkah ini, model *Naive Bayes* mempelajari distribusi fitur-fitur dari set data pelatihan. Setelah model dilatih, probabilitas kondisional dari setiap kelas berdasarkan fitur-fiturnya dihitung menggunakan *Teorema Bayes*. Ketika data test masuk, model *Naive Bayes* menggunakan probabilitas yang telah dihitung untuk memprediksi kelas yang paling mungkin untuk instance baru tersebut. *Naive Bayes* adalah algoritma yang kuat dan sederhana yang sering memberikan hasil yang memuaskan, terutama ketika asumsi independensi naif tidak mempengaruhi kualitas prediksi secara signifikan [57].

2.4 Tools yang digunakan

2.4.1 Microsoft Excel

Microsoft Excel merupakan tools yang banyak digunakan untuk mengelola data dalam bentuk spreadsheet, *Microsoft Excel* adalah program *spreadsheet* yang menawarkan fitur untuk menangani, memodifikasi, mengatur, dan menilai data. *Excel* dapat digunakan bersama dengan banyak program lain, seperti program *Microsoft Office* lainnya seperti *Word* dan *PowerPoint*, serta aplikasi bisnis, database, dan sistem ERP. Selain itu, program ini memiliki kemampuan untuk menampilkan data dalam berbagai format, seperti grafik atau diagram [58].

2.4.2 Trip Advisor

Sebagai komunitas perjalanan terbesar di dunia, *TripAdvisor* menawarkan jutaan wawasan, rekomendasi, dan saran wisatawan yang mungkin berguna bagi pengguna lain saat mereka merencanakan perjalanan mereka sendiri. *TripAdvisor* adalah platform penting dalam industri perjalanan yang membantu wisatawan mencari informasi, membuat keputusan, dan berbagi pengalaman tentang destinasi wisata di seluruh dunia. Dengan fitur-fitur yang kaya dan komunitas yang luas, *TripAdvisor* tetap menjadi sumber yang berguna bagi para pelancong. Ada sejumlah layanan di *TripAdvisor* yang mungkin bisa membantu wisatawan, seperti saran hotel, restoran, dan tempat wisata [59].

2.4.3 Jupyter Notebook

Jupyter Notebook adalah alat *online* untuk membuat program komputer interaktif yang menggunakan antarmuka konsol dan dapat dilihat dari mesin klien. Hasil pengembangan aplikasi akan membuat evaluasi instruksi pemrograman menjadi lebih mudah. Alat penilaian dirancang untuk siswa yang baru belajar pemrograman *Python* atau *C*. *Jupyter Notebook* adalah proyek yang memisahkan diri dari *IPython*, yang memiliki proyek aslinya sendiri yang disebut *IPython Notebook*. Karena mendukung bahasa pemrograman *Julian*, *Python*, dan *R*, ia dikenal dengan julukan *Jupyter* [60].

2.4.4 Python

Bahasa pemrograman seperti *Python* memprioritaskan kode yang dapat dibaca. Dinyatakan bahwa *Python* adalah bahasa yang menggabungkan berbagai fitur dengan sintaks pengkodean yang sangat mudah dibaca, serta fungsionalitas perpustakaan standar yang cukup besar dan luas. Salah satu cara untuk mendeskripsikan *Python* adalah sebagai bahasa pemrograman tujuan umum yang dirancang dengan mempertimbangkan kemudahan membaca kode sumber. *Python* memiliki perpustakaan lengkap yang memungkinkan pemrogram menulis aplikasi kompleks dengan kode sumber yang tampak sederhana. Pengembangan

web, otomatisasi, pengujian perangkat lunak, analisis data, pembelajaran mesin, dan pembuatan game adalah beberapa aplikasi paling populer untuk bahasa pemrograman *Python* [61].

2.4.5 Octoparse

Aplikasi scraping *online* bernama *Octoparse* dibuat untuk menyederhanakan proses pengambilan data dari website [30]. Antarmuka *Octoparse* yang intuitif dan fitur tunjuk dan klik memungkinkan pengguna dengan tingkat pengalaman apa pun untuk mengembangkan *web scraper* tanpa memerlukan pemahaman mendalam tentang pemrograman. Data web dapat dengan mudah diambil dan diekstraksi dengan *Octoparse*. Fitur sederhana namun efektif dari templat pengikisan *online Octoparse* mencoba memasukkan situs *web* atau kata kunci target dalam parameter, sehingga pengguna tidak perlu menulis kode atau menetapkan aturan pengikisan [62].

