

BAB II KAJIAN PUSTAKA

2.1 Analisis Sentimen

Menurut (Rahman Isnain dkk., 2021) Analisis Sentimen adalah proses menganalisis, memproses, merangkum, dan membuat inferensi terhadap teks subjektif. Saat ini, Analisis Sentimen dibagi menjadi dua pendekatan utama: klasifikasi berbasis pembelajaran mesin dan klasifikasi berbasis aturan. Metode pembelajaran mesin menggunakan kata-kata emosi sebagai fitur klasifikasi, sementara kamus emosi dapat digunakan untuk memudahkan pemilihan ciri-ciri sentimen dengan cepat dan efisien. Metode umum yang digunakan dalam kombinasi dengan analisis sentimen untuk menyelesaikan tugas klasifikasi adalah *Naïve Bayes* (NB), Support Vector Machine (SVM), dan Entropi Maksimum (EM).

2.2 Twitter

Twitter menurut (Giovani dkk., 2020) merupakan salah satu platform media sosial yang sangat populer di kalangan pengguna internet. Keberhasilan Twitter ini disebabkan oleh kesederhanaan dan kemudahan penggunaannya, serta memberikan kebebasan kepada pengguna untuk berbagi pendapat dan opini mereka dengan bebas. Aplikasi Twitter dibuat dan diluncurkan pada bulan Maret 2006 di San Francisco, California, Amerika yang lalu diakuisisi oleh Elon Musk pada tahun 2023. Dilansir dari *bankmycell.com*, tabel berikut adalah perkembangan pengguna twitter dari tahun ke tahun beserta prediksi di tahun mendatang.

Tabel 2. 1 Perkembangan Jumlah Pengguna Twitter

Tahun	Jumlah Pengguna Twitter
2024	335.70 juta
2023	353.90 juta
2022	368.40 juta
2021	362.40 juta
2020	347.60 juta
2019	312.70 juta

2.3 Data Mining

Penambangan data atau data mining menurut (Damuri dkk., 2021) merupakan metode yang berguna untuk menggali informasi berharga dari sejumlah data dengan menggunakan pengetahuan dalam bidang statistik, matematika, dan pengenalan pola. Tujuan dari data mining adalah untuk mengklasifikasikan, memprediksi, dan memperoleh informasi yang berarti. Melalui tahapan perencanaan dan analisis data, data mining memberikan informasi yang akurat untuk membuat prediksi berdasarkan tren masa lalu dan kondisi saat ini. (Yuli Mardi, 2019) menyebutkan, data mining melibatkan proses pencarian pola atau informasi yang menarik dalam dataset tertentu dengan menggunakan berbagai teknik, metode, atau algoritma. Pemilihan metode atau algoritma yang sesuai sangat tergantung pada tujuan dan keseluruhan proses *Knowledge Discovery in Database (KDD)*. Adapun varian dari data mining yang terfokus kepada pencarian informasi terhadap data yang berupa teks yaitu teks mining. Menurut (Runimerati dkk, 2023) teks mining merupakan proses eksplorasi dan analisis data teks yang tidak terstruktur dalam jumlah yang besar dengan menggunakan perangkat lunak yang dapat mengidentifikasi konsep, pola, topik, kata kunci, dan atribut lainnya dalam data.

2.4 Klasifikasi

Dalam penelitian (Yolanda Paramitha et al., 2023) dijelaskan bahwa klasifikasi adalah proses pengelompokan, yang berarti mengumpulkan benda/entitas yang sama dan memisahkan benda atau entitas yang tidak sama. Secara umum, batasan klasifikasi dapat dikatakan sebagai usaha untuk menata pengetahuan alam dalam urutan yang sistematis. Klasifikasi juga melibatkan penggunaan model untuk mengklasifikasikan *instance-instance* baru yang tidak diketahui kelasnya. Dalam *data mining*, terdapat berbagai teknik dan algoritma klasifikasi yang digunakan, seperti *decision tree*, *Naïve Bayes*, *K-NN*, dan algoritma lainnya.

2.5 TF-IDF (Term Frequency – Inverse Document Frequency)

Menurut (Umar & M. Adnan Nur, 2022) pengertian *TF-IDF* adalah metode yang bertujuan memberikan bobot pada kata-kata yang sering digunakan. *Term Frequency (TF)* adalah jumlah kata dalam vektor kata suatu kalimat dibagi dengan total kata dalam vektor kata tersebut. Sedangkan *Inverse Document Frequency (IDF)* bertujuan untuk mengurangi bobot kata-kata jika kata-kata tersebut ada di semua dokumen. Berikut adalah rumus *TF-IDF*:

$$tf_{t,d} = \frac{n_{t,d}}{N} \dots \dots \dots (2.1)$$

Keterangan :

$n_{t,d}$ = nilai istilah yang muncul

$tf_{t,d}$ = frekuensi kemunculan kata pada sebuah dokumen

N = semua term dalam dokumen

$$idf_d = \log \left(\frac{N}{df} \right) \dots \dots \dots (2.2)$$

Keterangan :

N = total dokumen

df = banyak dokumen yang mengandung term tersebut

idf_d = jumlah kemunculan suatu kata dalam dokumen

$$tfidf_{t,d} = tf_{t,d} \times idf_d \dots \dots \dots (3.3)$$

Keterangan :

$tfidf_{t,d}$ = Term Frequency-Inverse Document Frequency

$tf_{t,d}$ = nilai *TF*

idf_d = nilai *IDF*

2.6 Naïve Bayes

Merujuk pada penelitian (Ressan & Hassan, 2022) teori *Naïve Bayes* menentukan probabilitas terjadinya suatu peristiwa tertentu berdasarkan distribusi

probabilitas yang terkait dari peristiwa-peristiwa khusus lainnya. Model *Naïve Bayes* mengasumsikan bahwa keberadaan fitur tertentu dalam suatu kelas adalah independen dari keberadaan fitur-fitur lain dalam kelas yang sama. Mereka juga menjelaskan rumus *Naïve Bayes* adalah sebagai berikut :

$$P(c|X) = \frac{P(x|C)P(c)}{P(x)} \dots\dots\dots (4.4)$$

Keterangan :

$P(c|X)$: Peluang terjadinya c berdasarkan kondisi X (posteriori prob)

$P(c)$: Peluang terjadinya c (prior prob)

$P(X|C)$: Peluang terjadinya X berdasarkan kondisi pada hipotesis C

$P(X)$: Peluang terjadinya X

2.7 *Jupyter Notebook*

Jupyter Notebook merupakan sebuah aplikasi web yang bersifat open-source yang digunakan untuk membuat dan berbagi dokumen. Dokumen ini menggabungkan kode program, teks naratif, visualisasi, dan elemen-elemen interaktif lainnya. Dengan *Jupyter Notebook*, pengguna dapat menulis dan menjalankan kode program secara langsung dalam sel-sel notebook yang interaktif. Sel-sel ini dapat berisi kode dalam berbagai bahasa pemrograman seperti *Python*, *R*, *Julia*, dan lainnya. Selain itu, *Jupyter Notebook* juga mendukung penulisan teks naratif menggunakan format *Markdown* yang memungkinkan pengguna untuk menjelaskan langkah-langkah atau hasil analisis yang dilakukan. Dengan menggunakan *Jupyter Notebook*, pengguna dapat menggabungkan kode, hasil, dan penjelasan dalam satu dokumen yang mudah dibaca dan dapat diakses secara interaktif.

2.8 Penelitian Terdahulu

Sebelum melaksanakan penelitian ini, dilakukan riset terlebih dahulu tentang penelitian sebelumnya yang memiliki bidang dan topik yang serupa. Berikut adalah daftar penelitian sebelumnya yang dijadikan referensi.

Tabel 2. 2 Daftar Referensi

Penulis dan Tahun	Judul	Keterangan
(Febriyani & Februariyanti, 2020.)	Analisis Sentimen Terhadap Program Kampus Merdeka Menggunakan Algoritma Naive Bayes Classifier Di Twitter	Inti dari penelitian ini adalah untuk menganalisis sentimen opini publik terhadap program Kampus Merdeka di Twitter dan mengevaluasi tingkat akurasi, kinerja, dan program Kampus Merdeka itu sendiri. Hasil penelitian menunjukkan bahwa sistem mampu mengklasifikasikan sentimen positif sebanyak 272 opini dan sentimen negatif sebanyak 229 opini, dengan rata-rata akurasi 60%, presisi 64%, recall 58%, dan f1-score 58%.
(Ariyanti & Iswardani, 2020.)	Teks Mining untuk Klasifikasi Keluhan Masyarakat Menggunakan Algoritma Naive Bayes	Penelitian ini menggunakan teks mining dengan algoritma <i>Naive Bayes</i> untuk melakukan klasifikasi keluhan masyarakat. Hasil penelitian ini menunjukkan tingkat akurasi sebesar 95%.
(Giovani dkk., 2020)	ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI	Tujuan penelitian ini adalah untuk mencari algoritma yang paling optimal dalam mengklasifikasikan komentar terkait dengan aplikasi Ruang Guru. Hasil pengujian menunjukkan bahwa algoritma PSO berbasis SVM menghasilkan hasil yang paling baik dengan akurasi 78,55% dan AUC 0,853. Penelitian ini memberikan kontribusi dalam mendapatkan algoritma yang efektif untuk mengklasifikasikan sentimen terhadap aplikasi Ruang Guru.