

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Clarisa Hasya Yutika (2021) melakukan penelitian dengan mengimplementasikan metode *Naïve Bayes Classifier* dengan data yang diperoleh dari web *female daily* untuk melakukan analisis sentimen *Review Female Daily Naïve Bayes* dengan hasil pengujian diperoleh nilai *F1-Score* sebesar 62,81% .

Agri Yodi Prayoga, Asep Id Hadiana, Fajri Rakhmat Umbara (2021) melakukan penelitian dengan mengimplementasikan penggunaan algoritma *Bernoulli Naïve Bayes* dalam mendeteksi berita *hoax* pada berita *online* Bahasa Inggris. Hasil implementasi menunjukkan model prediksi yang dibangun dengan 8800 data berita mampu menghasilkan nilai akurasi sebesar 98,5%.

Indarti Mayasari(2022) melakukan penelitian dengan Klasifikasi topik *tweet* mengenai COVID-19 menggunakan metode *Multinomial Naïve Bayes* dilakukan pada *tweet* dalam bahasa Indonesia. Data *tweet* dikumpulkan dari 9 Juni 2021 sampai 9 Juli 2021 dengan kata kunci 'COVID'. Hasil pengujian diperoleh dengan akurasi sebesar 61,23%, rata-rata presisi sebesar 59%, rata-rata *recall* sebesar 45% dan rata-rata *F1-score* sebesar 47%.

Risa Wati, Siti Ernawati, dan Hilda Rachmi(2023) melakukan penelitian dengan mengimplementasikan klasifikasi analisis sentimen masyarakat terhadap isu kenaikan Biaya Perjalanan Ibadah Haji pada media sosial *Twitter*, Hasil

pengujian dengan memperoleh nilai akurasi sebesar 89% dan nilai ROC sebesar 0,91 masuk kedalam *Excellent Classification*.

Aang Aliyudin dan Rianto (2023) melakukan analisis sentimen yang menggunakan metode *Naïve Bayes* terhadap komentar tentang Piala Dunia u-20 yang batal diselenggarakan di Indonesia pada media *YouTube* menghasilkan akurasi prediksi yang paling sebesar 0.906 yang mana nilai tersebut cukup tinggi.

Pada penelitian kali ini akan menganalisis sentimen negatif, positif dan netral pada *YouTube* tentang pernyataan Ganjar Pranowo terhadap pembatalan Piala Dunia U-20 di Indonesia menggunakan metode *Naive Bayes Classifier* dengan data komentar pada *YouTube* pada tahun 2023.

Tabel 1.1. Tinjauan Pustaka

No	Nama Pengarang	Objek	Metode	Hasil
1	Clarisa Hasya Yutika (2021)	Sentimen Web <i>Review Female Daily</i>	<i>Naive Bayes Classifier</i>	hasil pengujian diperoleh nilai <i>F1-Score</i> sebesar 62,81%
2	Agri Yodi Prayoga, Asep Id Hadiana, Fajri Rakhmat Umbara (2021)	Sentimen untuk mendeteksi berita <i>hoax</i> pada berita <i>online</i> Bahasa Inggris	Algoritma <i>Bernoulli Naive Bayes</i>	Hasil klasifikasi nilai akurasi sebesar 98,5% nilai presisi 99,1%, <i>recall</i> 97,8%, dan <i>F1-score</i> 98,4%.
3	Indarti Mayasari (2022)	<i>Tweet</i> mengenai COVID-19	<i>Multinomial Naive Bayes</i>	Hasil klasifikasi dengan pembobotan <i>TF-IDF</i> diperoleh akurasi sebesar 61,23%, rata-rata presisi sebesar 59%, rata-rata <i>recall</i> sebesar 45% dan rata-rata <i>F1-score</i> sebesar 47%.
4	Risa Wati, Siti Ernawati, dan Hilda Rachmi (2023)	<i>Tweet</i> mengenai isu kenaikan Biaya Perjalanan Ibadah Haji	<i>Naive Bayes Classifier</i>	Hasil pengujian mendapatkan nilai akurasi sebesar 89%
5	Aang Aliyudin, Rianto (2023)	Komentar tentang Piala Dunia U-20 yang batal di Indonesia pada media <i>YouTube</i>	<i>Naive Bayes Classifier</i>	Sentimen positif, negatif dan netral menghasilkan nilai akurasi yaitu 0.096.
6	usulan (2023)	Komentar di <i>YouTube</i> tentang pembatalan Piala Dunia U-20 di Indonesia oleh Ganjar Pranowo	<i>Naive Bayes Classifier</i>	Sentimen positif, negatif dan netral di <i>YouTube</i> pernyataan Ganjar Pranowo terhadap pembatalan Piala Dunia U-20 di Indonesia.

2.2 Dasar Teori

2.2.1 Youtube

Youtube merupakan salah satu dari *social Network* yang berkategori *Multimedia sharing*. Pada umumnya, file-file *Multimedia* yang dibagikan berupa file gambar, audio, dan video melalui internet. Fitur-fitur yang disediakan *multimedia sharing* antara lain mengirim dan menanggapi komentar, pemberian *rate* (peringkat), berbagi (*share*) dalam bentuk alamat (URL). Proses yang dilakukan meliputi unggah (*upload*) dan unduh (*download*). Diantara *social network* yang ada saat ini seperti *Flicker* dan *Picasa*, *YouTube* tetap menjadi platform *sharing video* terbesar di Internet. Sudah banyak penelitian dilakukan tentang pola interaksi pengguna di *YouTube* ini. Interaksi ini berupa komentar yang biasanya digunakan untuk menunjukkan perasaannya, memberikan dukungan emosi, kesan, atau nasehat sebagai tanggapan atas video. Perilaku pengguna banyak digunakan untuk berbagai kepentingan seperti tujuan pemasaran dimana perusahaan dapat menggunakan informasi untuk menargetkan pesan dan produk mereka (Thomas, Santi, 2023).

2.2.2 Youtube API

Application Programming Interface (API) sebagai fungsi, kumpulan perintah dan juga protokol yang berguna untuk membantu programmer dalam membangun perangkat lunak dalam sistem operasi tertentu. Khususnya dalam

Youtube memungkinkan pengembang mengakses statistik video dan data saluran *YouTube* melalui dua jenis panggilan, *REST*, dan *XML-RPC*. Google sendiri mendeskripsikan Sumber Daya API *YouTube* sebagai *API* dan Alat yang memungkinkan Anda membawa pengalaman *YouTube* ke laman web, aplikasi, atau perangkat Anda. Pada dasarnya fungsi serta perintah pada *Youtube API* merupakan alat pemanggil *system calls* yang mana berhubungan langsung ke sistem operasinya(Saputra.2019).

2.2.3 Python

Python merupakan salah satu bahasa pemrograman yang banyak digunakan oleh perusahaan besar maupun para *engineer* untuk mengembangkan berbagai macam aplikasi berbasis *desktop*, *web* dan *mobile*. *Python* diciptakan oleh Guido Van Rossum di Belanda pada tahun 1990. Van Rossum mengembangkan *Python* sebagai hobi, kemudian *Python* menjadi bahasa pemrograman yang dipakai secara luas dalam industri dan pendidikan karena ringkas, sederhana, memiliki pustaka yang luas, dan sintak intuitif(Hanif.2022).

2.24 Analisis Sentimen

Analisis sentimen merupakan salah satu metode yang dapat mengukur dan menganalisis pada suatu kasus atau objek tertentu, dimana melalui sentimen analisis tersebut dapat diambil sebuah kesimpulan dan keputusan berdasarkan teks dalam bentuk kalimat ataupun dokumen.(Aang.2023).Salah satu algoritma yang

dapat digunakan untuk melakukan sentimen analisis adalah *Naïve Bayes* dimana pada algoritma ini sistem akan melakukan klasifikasi melalui probabilitas kemungkinan dari data yang diperoleh, algoritma lain yang dapat digunakan dalam teknik analisis sentimen adalah *Support Vector Machine (SVM)*, dimana pada algoritma tersebut proses pembobotan dilakukan melalui pembentukan pola garis untuk dilakukan pembobotan dan klasifikasi (Setiawan, H. 2021).

2.2.5 Naive Bayes Classifier

Metode *Naïve Bayes* merupakan metode yang memanfaatkan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes (Susanto & Indriyani, 2019) dengan pembelajaran statistik berbasis probabilitas dan memiliki asumsi premis bahwa semua atribut sepenuhnya independen pada nilai variabel kelas (Wang, Jiang, & Li, 2015). Teorema Bayes mengasumsikan bahwa semua variabel independen mempertimbangkan nilai variabel kelas (Saritas & Yasar, 2019) berdasarkan pengalaman di masa sebelumnya. Ciri utama dari algoritma ini adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi. Elemen utama dari pengklasifikasi *Naive Bayes* berkaitan dengan tiga aspek, yaitu, *prior*, *posterior* dan *class conditional probability* (Wibawa et al., 2019). Rumus untuk menghitung *Naïve Bayes* yang digunakan yaitu :

$$P(H | X) = \frac{P(X | H)P(H)}{P(X)} \dots\dots\dots(1)$$

$P(H | X)$: Probabilitas hipotesis berdasarkan kondisi (*posteriori probability*)

$P(H)$: Probabilitas hipotesis (*prior probability*)

$P(X | H)$: Probabilitas X berdasar kondisi pada hipotesis, jika kita ketahui bahwa H benar

$P(X)$: Probabilitas H

2.2.6 Akurasi, Presisi, Recall, dan F1-score

2.2.6.1 Akurasi

Akurasi adalah matrik evaluasi yang mengukur seberapa baik model membuat prediksi yang benar dari total prediksi yang dilakukan. Dalam konteks klasifikasi, akurasi memberikan gambaran mengenai seberapa sering model memprediksi kelas yang benar, baik itu kelas positif maupun negatif (Rina, 2023).

rumus yang digunakan yaitu :

$$\frac{\text{Total jumlah prediksi yang benar}}{\text{Total Jumlah Data}} \dots\dots\dots(2)$$

Keterangan:

- Total jumlah prediksi yang benar adalah jumlah data yang diprediksi dengan benar oleh model *Naive Bayes*.
- Total jumlah data adalah jumlah seluruh data yang dievaluasi oleh model.

2.2.7.2 Presisi

Presisi adalah matrik evaluasi yang mengukur seberapa baik model membuat prediksi yang benar untuk kelas positif dari total prediksi positif yang dilakukan. Dalam konteks klasifikasi, presisi memberikan gambaran mengenai seberapa sering model memprediksi kelas positif dengan benar, di antara semua prediksi positif yang dibuat oleh model(Rina.2023).rumus yang digunakan yaitu :

$$\frac{TP}{FP+TP} \dots\dots\dots(3)$$

Keterangan:

TP : Banyaknya data positif yang benar

FP : Banyaknya data positif yang salah

2.2.7.3 Recall

Recall adalah matrik evaluasi yang mengukur seberapa banyak data relevan yang berhasil diidentifikasi oleh sistem dan mengukur seberapa banyak kelas positif yang berhasil diidentifikasi oleh model(Rina.2023).rumus yang digunakan yaitu :

$$\frac{TP}{TP+FN} \dots\dots\dots(4)$$

Keterangan:

TP : Banyaknya data positif yang benar

FN : Banyaknya data negatif yang salah

2.2.7.4 F1 Score

Skor *F1* adalah matrik evaluasi pembelajaran mesin alternatif yang menilai keterampilan prediktif suatu model dengan mengelaborasi performa berdasarkan kelasnya, bukan performa keseluruhan yang dilakukan berdasarkan akurasi. Skor *F1* menggabungkan dua matrik yang saling bersaing, yaitu skor presisi dan perolehan model, yang menyebabkan penggunaannya secara luas dalam literatur terkini(Kundu,Rohit.2022).rumus yang digunakan yaitu :

$$\frac{2 * Presisi * Recall}{Presisi + Recall} \dots\dots\dots(5)$$