

BAB II LANDASAN TEORI

2.1. Tinjauan Pustaka

Beberapa penelitian terdahulu yang dijadikan alur referensi dalam penelitian ini adalah :

Tabel 2. 1 Tinjauan Pustaka

No	Nama (Tahun)	Judul	Metode	Hasil
1	Mahardhika and Zuliarso (2018)	Analisis sentimen terhadap pemerintahan Joko Widodo pada media sosial <i>twitter</i> menggunakan algoritma <i>Naive Bayes Classifier</i> .	Algoritma <i>Naive Bayes Classifier</i> .	Pada penelitian ini dapat diambil kesimpulan bahwa Metode <i>Naive Bayes Classifier</i> dalam melakukan klasifikasi <i>tweet</i> sentimen negatif dan positif dengan 300 data latih dan 100 data uji mendapat hasil akurasi sebesar 97% dan Dengan hasil akurasi yang cukup tinggi yaitu 97% maka metode <i>Naive Bayes Classifier</i> dapat digunakan untuk melakukan klasifikasi <i>tweet</i> dengan sentimen negatif dan positif secara otomatis
2	Pintoko and Muslim (2018)	Analisis Sentimen Jasa Transportasi <i>Online</i> pada <i>Twitter</i> Menggunakan Metode <i>Naive Bayes Classifier</i>	<i>Naive Bayes Classifier</i>	Berdasarkan hasil analisis yang telah dilakukan terkait sentimen terhadap jasa transportasi <i>online</i> dengan menggunakan metode <i>Naive Bayes</i> untuk

No	Nama (Tahun)	Judul	Metode	Hasil
				menentukan kelas prediksi, didapatkan hasil sentimen positif sebesar 88.60% dan sentimen negatif sebesar 11.40% dengan akurasi sebesar 86.80%. Hasil menunjukkan tingkat sentimen positif lebih besar dibandingkan dengan tingkat sentimen negatif.
3	Suryono, Utami and Luthfi (2018)	klasifikasi sentimen pada <i>twitter</i> dengan <i>Naïve Bayes Classifier</i> .	<i>Naïve Bayes Classifier</i> .	Berdasarkan hasil klasifikasi besar persentase untuk ketiga kelas, diletakkan pertama yaitu dengan kelas sentimen neutral dengan 52%, kedua sentimen positif dengan 28% dan ketiga sentimen negatif dengan 20%. Berdasarkan hasil persentase kelas sentimen, sentimen neutral merupakan sentimen yang paling banyak apabila dikaitkan dengan topik Presiden Joko Widodo dan Pemerintahannya.
4	Assuja and Saniati (2016)	Analisis Sentimen <i>Tweet</i> menggunakan	Backpropagation Neural Network	Hasil pengujian pada sentimen analisis menunjukkan jika

No	Nama (Tahun)	Judul	Metode	Hasil
		Backpropagation Neural Network.		tahap praproses diperlukan untuk meningkatkan akurasi hasil klasifikasi. Sedangkan bentuk praproses yang paling baik ditunjukkan oleh penggunaan transformasi kata gaul, <i>cleaning</i> (<i>retweet</i> dan <i>mention</i>) dan normalisasi (<i>case folding</i> dan perubahan makian).
5	Wilianto et al (2017)	Analisis Sentimen Terhadap Tempat Wisata Dari Komentar Pengunjung Dengan Menggunakan Metode Naïve Bayes Classifier Studi Kasus Jawa Barat.	Metode Naïve Bayes Classifier	Dari hasil penelitian yang telah dilakukan terlihat bahwa algoritma <i>naïve bayes classifier</i> dapat mengklasifikasikan suatu opini berupa komentar ke dalam dua kelas yaitu positif dan negatif dengan akurat. Tingkat keakurasian dari pengklasifikasian tersebut sangat dipengaruhi oleh proses training. Sehingga dapat disimpulkan dari pengklasifikasian yang dihasilkan dapat terlihat dengan jelas informasi sentimen

No	Nama (Tahun)	Judul	Metode	Hasil
				public terhadap tempat wisata.

2.2. Media Sosial

Media sosial merupakan perkembangan teknologi berbasis internet yang memungkinkan para penggunanya dapat dengan mudah membagikan, berkomentar, atau menuliskan sesuatu tentang hal-hal pribadi yang mereka lakukan pada blog, jejaring sosial, wiki, forum dan dunia virtual yang merupakan . bagian dari media sosial yang paling sering digunakan oleh banyak orang diseluruh belahan dunia.(Setyati, 2013).

2.3. Twitter

Twitter merupakan bagian dari layanan jejaring sosial dan microblog *online* yang memudahkan penggunanya untuk membagikan, menulis, mengirim dan membaca pesan berbasis teks. *Twitter* dapat membuat penggunanya terhubung dengan orang terdekat tanpa terhalang sekayang memiliki peran besar dalam mempengaruhi aspek dari segi sosial, komunikasi, psikologi, marketing, dan politik melalui twiter kita bisa menuliskan sebuah status atau pesan yang nantinya akan dibaca oleh pengguna lainnya, fitur ini disebut tweet atau kicauan. Keunikan dari Twitter adalah membuat tweet atau postingan yang ada di Twitter dengan ukuran maksimum 140 karakter (Pintoko & Muslim, 2018).

Twitter juga mampu membuat penggunanya dapat mengikuti trend yang terjadi secara realtime, cerita, informasi dan segala hal yang terjadi di seluruh penjuru

dunia hanya dari *tweet* sesama pengguna *twitter*. Berikut ini adalah beberapa istilah yang dikenal pada jejaring sosial *twitter* :

1. *Mention*

Mention atau yang dapat diartikan dengan menyebut. Adalah istilah yang digunakan saat pengguna *twitter* menyebut atau memanggil nama akun *twitter* pengguna lainnya. *Mention* dilakukan dengan menuliskan “@” dan diikuti dengan nama pengguna lainnya.

2. *Emoticon*

Emoticon atau yang dikenal dengan ekspresi wajah, adalah istilah yang digunakan pengguna *twitter* saat mengekspresikan mood yang mereka rasakan. *Emoticon* biasanya direpresentasikan dengan bentuk wajah atau kombinasi antara huruf, angka dan tanda baca.

3. *Hashtag*

Hashtag atau yang dapat diartikan dengan tagar ini biasanya disimbolkan dengan “#”. *Hashtag* biasanya diletakan pada awal kata pada topik yang sedang dibahas.

4. *Trending Topics*

Trending Topics adalah sebuah ungkapan khususnya pengguna *twitter* tentang suatu populer yang sedang terjadi. Suatu hal bisa dikatakan *trending topics* apalagi hal/topik tersebut sedang banyak dibicarakan di *twitter*.

Sebuah fakta yang cukup mengejutkan disampaikan oleh Giummole et al. (Giummole et al.,2013) bahwa *trending topics* di *Twitter* akan meningkatkan prediksi hasil pencarian di *Google*. Hal ini bisa terjadi

karena dalam setiap lima menit, *Twitter* akan mengeluarkan daftar topik yang sangat populer (*trending topics*) dengan cara memonitor dan menganalisa pembicaraan, sedangkan *Google* juga akan mengeluarkan daftar pencarian populer yang dicari oleh penggunanya pada setiap jam.

2.4. Review

Review Tokoh Publik Opini terhadap tokoh publik sangat banyak dan mudah ditemukan di dunia maya tetapi untuk menjadikan opini tersebut sebuah informasi yang sangat bermanfaat sangatlah sulit. Untuk mempermudah pengamatan terhadap tokoh publik maka digunakan sosial media *twitter* untuk pengumpulan opini yang dilontarkan oleh masyarakat sehingga dapat terkumpul bermacam-macam opini (Basari *et al*, 2013).

2.5. Analisis Sentimen

Analisis Sentimen atau opinion mining didefinisikan sebagai bidang ilmu yang mempelajari bagaimana mengekspresikan sentimen, opini atau pendapat dan emosi yang diekpresikan didalam teks. Terdapat beberapa topik pembahasan dalam sentimen analisis, salah satu yang paling sering diteliti yaitu klasifikasi sentimen. Topik ini berfokus pada kegiatan pengelompokan sentimen berdasarkan teks opini terhadap pembahasan masalah tertentu (Liu, 2012).

Langkah-langkah yang umumnya ditemukan pada klasifikasi teks analisa sentimen adalah:

1. Definisikan domain dataset pengumpulan dataset yang melingkupi suatu domain, misalnya dataset review film, dataset review produk, dan lain sebagainya.

2. *Pre-processing* Tahap pemrosesan awal yang umumnya dilakukan dengan proses *Tokenization*, *Stopwords removal*, dan *Stemming*.
3. *Transformation Proses* representasi angka yang dihitung dari data tekstual. Binary representation yang umumnya digunakan dan hanya menghitung kehadiran atau ketidakhadiran sebuah kata di dalam dokumen. Berapa kali sebuah kata muncul di dalam suatu dokumen juga digunakan sebagai skema pembobotan dari data tekstual. Proses yang umumnya digunakan yaitu TF-IDF, Binary transformation, dan *Frequency transformation*.
4. *Feature Selection* Pemilihan fitur (*feature selection*) bisa membuat pengklasifikasi lebih efisien/efektif dengan mengurangi jumlah data untuk dianalisa dengan mengidentifikasi fitur yang relevan yang selanjutnya akan diproses. Metode pemilihan fitur yang biasanya digunakan adalah Expert. Knowledge, Minimum Frequency, Information gain, Chi-Square, dan lain sebagainya
5. *Classification Proses* klasifikasi umumnya menggunakan pengklasifikasi seperti *Naïve Bayes*, *Support Vector Machine*, dan lain sebagainya. f. *Interpretation/Evaluation* Tahap evaluasi biasanya menghitung akurasi, recall, precision, dan F-1.

2.6. Text mining

Text mining adalah istilah yang mengacu pada teknik penambangan data untuk menganalisis dan memproses data tekstual yang tidak terstruktur dan semi terstruktur (Hardaya et al., 2017). *Text mining* memiliki proses yang sama dengan data mining tetapi memiliki inputan yang berbeda. Pada *text mining*, pertama

diperlukan pengambilan data kemudian data tersebut perlu di *pre-processing* sebelum proses klasifikasi.

Text mining menggunakan penerapan data mining untuk mengubah data tidak terstruktur menjadi data terstruktur melalui tahap-tahap sebagai berikut :

1. Text Preprocess yaitu pemecahan sekumpulan karakter ke dalam kata-kata.
2. Feature Generation / Text Transformation yaitu mengubah kata-kata ke dalam bentuk dasar sekaligus mengurangi jumlah kata-kata tersebut.
3. Feature Selection yaitu seleksi feature untuk mengurangi dimensi dari suatu kumpulan teks.
4. *Text mining* / Pattern Discovery yaitu dapat berupa unsupervised learning (clustering) atausupervised learning (classification).
5. Interpretation / Evaluation yaitu pengukuran efektifitas untuk mengevaluasi metode yang diterapkan menggunakan parameter precision.

2.7. Naïve Bayes Classifier

Salah satu tugas Data Mining adalah klasifikasi data, yaitu memgklasifikasikan (memetakan) data ke dalam satu atau beberapa kelas yaang sudah didefinisikan sebelumnya. *Naïve bayes classifier (NBC)* adalah salah satu metode dalam klasifikasi data. *Naïve Bayes Classifier* merupakan salah satu metode machine learning yang memanfaatkan perhitungan probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya (Amolik, 2016).

Dasar dari Naïve Bayes yang dipakai dalam pemrograman adalah rumus Bayes:

$$P(A|B) = (P(B|A) * P(A))/P(B) \dots\dots\dots (1)$$

Keterangan :

Peluang kejadian A sebagai B ditentukan dari peluang B saat A, peluang A, dan peluang B.

Pada pengaplikasiannya nanti rumus ini berubah menjadi :

$$P(C_i|D) = (P(D|C_i)*P(C_i)) / P(D) \dots\dots\dots (2)$$

Naïve Bayes Classifier atau bisa disebut sebagai *Multinomial Naïve Bayes* merupakan model penyederhanaan dari Metode klasifikasi yang berakar pada Teorema bayes. Teorema Bayes sendiri dikenal dengan memprediksi peluang dimasa depan berdasarkan pengalaman dimasa sebelumnya. Penyederhanaan Teorema bayes ke naïve bayes didasarkan pada asumsi naïve yaitu asumsi yang sangat kuat akan independensi dari masing-masing kondisi/kejadian.

Dibawah ini adalah persamaan model penyederhanaan dari Metode Bayes adalah:

$$V_{MAP} = \arg \max P(V_j | a_1, a_2, \dots, a_n) \dots\dots\dots (3)$$

Menurut persamaan (3), maka persamaan (1) dapat ditulis:

$$c_{NB} = \arg \max_{c_j \in C} P(c_j) \prod_i P(x_i | c_j)$$

2.8. Confusion Matrix

Confusion matrix merupakan salah satu metode yang digunakan untuk menilai akurasi dan mengukur kemampuan suatu metode klasifikasi. *Confusion matrix* menyimpan informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang sebenarnya. *Confusion matrix* merupakan suatu *tools* penting dalam metode visualisasi yang digunakan di dalam mesin pembelajaran yang biasanya berisi dua kategori atau lebih *Invalid source specified*. *Confusion matrix* prediksi dua kelas dapat dilihat pada gambar berikut:

Tabel 2. 2 Confusion Matrix

		Actual Values	
		Positive (1)	Negative (0)
Predicted Values	Positive (1)	TP	FP
	Negative (0)	FN	TN

Matriks tersebut memiliki empat nilai yang dijadikan acuan dalam perhitungan, dimana:

True Positive (TP) = ketika kelas yang diprediksi positif dan faktanya positif.

True Negative (TN) = ketika kelas yang diprediksi negatif dan faktanya negatif.

False Positive (FP) = ketika kelas yang diprediksi positif dan faktanya negatif.

False Negatif (FN) = ketika kelas yang diprediksi megatif dan faktanya positif.

Berdasarkan nilai TP, TN, FP dan FN dapat diperoleh nilai akurasi. Nilai akurasi menggambarkan seberapa akurat system dapat mengklasifikasi data secara benar. Nilai akurasi menggambarkan seberapa akurat system dapat mengklasifikasi data secara benar. Dari nilai akurasi, presisi dan *recall* diperoleh persamaan sebagai berikut:

$$\text{Akurasi} = \frac{TP+TN}{\text{Total}/(TP+TN+FP+FN)}$$

$$\text{Presisi} = \frac{TP}{(FP+TP)}$$

$$\text{Recall} = \frac{TP}{(FN+TP)}$$

$$\text{F1-Score} = \frac{\text{Presisi} \times \text{Recall}}{\text{Presisi} + \text{Recall}}$$

2.9. Python

Python merupakan bahasa pemrograman yang freeware atau perangkat bebas dalam arti sebenarnya, tidak ada batasan dalam penyalinan atau mendistribusikannya. Lengkap dengan source code, debugger dan profiler, antarmuka yang terkandung di dalamnya untuk pelayanan antarmuka, fungsi sistem, GUI (antarmuka pengguna grafis), dan basis datanya Python dapat digunakan dalam beberapa sistem operasi, seperti kebanyakan sistem UNIX, PCs (DOS, Windows, OS/2), Macintosh, dan lainnya. Pada kebanyakan sistem operasi linux, bahasa pemrograman ini menjadi standarisasi untuk disertakan dalam paket distribusinya (Muhardian, 2018).