

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Guna membedakan penelitian ini dengan penelitian sebelumnya maka penulis melakukan studi pustaka terhadap peneliti peneliti terdahulu.

Agnes Rossi Trisna Lestari, Rizal Setya Perdana, dan M. Ali Fauzi (2017), “Analisis Sentimen Tentang Opini Pilkada Dki 2017 Pada Dokumen Twitter Berbahasa Indonesia Menggunakan Naïve Bayes dan Pembobotan Emoji”. Pada penelitian tersebut memiliki hasil akurasi sistem pada pembobotan tesktual 68,52%, pada pembobotan non-tesktual 75,93%, dan pada penggabungan kedua pembobotan 74,81% sehingga disimpulkan bahwa penggabungan kedua pembobotan meningkatkan hasil akurasi sistem. Perbandingan konstanta α untuk pembobotan tekstual dan β untuk pembobotan non-tekstual adalah 0,4 dan 0,6 sampai dengan 0,1 dan 0,9 sehingga dapat disimpulkan bahwa pembobotan non-tekstual sangat mempengaruhi hasil dari klasifikasi sentimen.

Yuna Sophia Dewi Febriant (2017), “Analisis Dan Klasifikasi Sentimen Terhadap Twitter STMIK Akakom Yogyakarta Menggunakan Naïve Bayes Classifier”. Pada penelitian tersebut memiliki hasil yaitu sistem ini hanya dapat menampilkan status twitter secara *realtime* otomatis sebanyak < 100 status twitter. Hasil Akurasi pengujiam klasifikasi dengan metode Naive Bayes Classifier diperoleh sebesar 78%.

Ajeng Widya Ayuningtyas (2020), “Analisis Sentimen Terhadap STMIK Akakom Yogyakarta Menggunakan Metode Naïve Bayes Berdasarkan Komentar Facebook Dan Ulasan Di Google Maps”. Pada penelitian tersebut memiliki hasil yaitu metode Naïve Bayes dalam kasus ini dapat digunakan untuk klasifikasi teks dengan tingkat akurasi sebesar 89,73% dan tingkat kesalahan bernilai 10,27%.

Widya Sulistyani (2020), “Analisa Topik Data Media Sosial Twitter Menggunakan *Latent Semantic Analysis* (Studi Kasus : Gubernur Jawa Tengah)”. Pada penelitian tersebut memiliki hasil yaitu dari pengujian yang dilakukan terhadap 5210 tweet didapatkan informasi yang beragam. Dari 10 topik utama yang ditampilkan dapat diketahui bahwa data twitter dengan menggunakan *keyword* Ganjar Pranowo periode 7 juli-12 juli 2020 memberikan informasi tentang pemeliharaan wilayah di Jawa Tengah, gerakan masyarakat sehat yang digalakkan, kesadaran masyarakat akan *COVID-19* yang menjadikan masyarakat mematuhi protokol kesehatan, pembahasan tentang banjir tegal dan aktivitas vulkanik gunung merapi.

Hena Triana (2021), “Analisis Sentimen Pencitraan Perguruan Tinggi Di Yogyakarta Menggunakan Metode Naïve Bayes Classifier”. Pada penelitian tersebut memiliki hasil yaitu hasil klasifikasi dari data uji didapat presentase sentimen netral 82.1%, positif 14.8% dan negatif 3.1%. Dari data uji model mampu memprediksi data dengan menghasilkan akurasi sebesar 73%.

Tabel 2. 1 Perbandingan Penelitian

Peneliti	Objek	Metode	Hasil
<p>Agnes Rossi Trisna Lestari, Rizal Setya Perdana, dan M. Ali Fauzi (2017)</p>	<p>Pilkada Dki 2017</p>	<p>Naïve Bayes</p>	<p>1. Hasil akurasi sistem pada pembobotan tekstual 68,52%, pada pembobotan non-tekstual 75,93%, dan pada penggabungan kedua pembobotan 74,81% sehingga disimpulkan bahwa penggabungan kedua pembobotan meningkatkan hasil akurasi sistem.</p> <p>2. Perbandingan konstanta α untuk pembobotan tekstual dan β untuk pembobotan non-tekstual adalah 0,4 dan 0,6 sampai dengan 0,1 dan 0,9 sehingga dapat</p>

Lanjutan Tabel 2.1 Perbandingan Penelitian

			disimpulkan bahwa pembobotan non-tekstual sangat mempengaruhi hasil dari klasifikasi sentimen.
Yuna Sophia Dewi Febriant (2017)	Twitter STMIK Akakom Yogyakarta	Naïve Bayes	<ol style="list-style-type: none"> 1. Sistem ini hanya dapat menampilkan status twitter secara <i>realtime</i>/otomatis sebanyak < 100 status twitter. 2. Hasil Akurasi pengujian klasifikasi dengan metode Naive Bayes Classifier diperoleh sebesar 78%.
Ajeng Widya Ayuningtyas (2020)	Komentar Facebook Dan Ulasan Di Google Maps	Naïve Bayes	Dalam kasus ini, metode Naïve Bayes dapat digunakan untuk klasifikasi teks dengan

Lanjutan Tabel 2.1 Perbandingan Penelitian

	Terhadap STMIK Akakom Yogyakarta		tingkat akurasi sebesar 89,73% dan tingkat kesalahan bernilai 10,27%.
Widya Sulistyani (2020)	Gubernur Jawa Tengah	Latent Semantic Analysis	Dari pengujian yang dilakukan terhadap 5210 tweet didapatkan informasi yang beragam. Dari 10 topik utama yang ditampilkan dapat diketahui bahwa data twitter dengan menggunakan <i>keyword</i> Ganjar Pranowo periode 7 juli-12 juli 2020 memberikan informasi tentang pemeliharaan wilayah di Jawa Tengah, gerakan masyarakat sehat yang digalakkan, kesadaran masyarakat akan <i>COVID-19</i> yang

Lanjutan Tabel 2.1 Perbandingan Penelitian

			menjadikan masyarakat mematuhi protokol kesehatan, pembahasan tentang banjir tegal dan aktivitas vulkanik gunung merapi.
Hena Triana (2021)	Pencitraan Perguruan Tinggi Di Yogyakarta	Naïve Bayes	<ol style="list-style-type: none"> 1. Hasil klasifikasi dari data uji didapat presentase sentimen netral 82.1%, positif 14.8% dan negatif 3.1%. 2. Dari data uji model mampu memprediksi data dengan menghasilkan akurasi sebesar 73%.

2.2 Dasar Teori

Pada bagian ini akan diberikan penjelasan, definisi dan uraian yang diperoleh dari berbagai referensi yang dipublikasikan pada media cetak atau elektronik yang berkaitan dengan topik penelitian.

2.2.1 Twitter

Twitter sebagai salah satu situs *microblogging* yang memungkinkan pengguna untuk berbagi pesan teks pendek yang disebut tweet. Twitter merupakan media yang paling banyak digunakan terutama di Indonesia. Berdasarkan data dari PT Bakrie Telecom, twitter memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna seluruh dunia. Indonesia menempati urutan kelima setelah USA, Brazil, Jepang dan Inggris sebagai negara pengguna Twitter terbesar di dunia. Media sosial Twitter digunakan untuk mengutarakan berbagai pendapat atau opini akan sebuah produk, layanan atau hal lainnya. Twitter diciptakan oleh Jack Dorsey di tahun 2006 dan pertama meluncur di dunia maya saat Juli 2006 dengan alamat <http://www.Twitter.com> yang masih digunakan hingga saat ini. Pengguna dapat menulis pesan berdasarkan topik dengan menggunakan tanda *#(hashtag)*. Sedangkan untuk menyebutkan atau membalas pesan dari pengguna lain bisa menggunakan tanda *@*.

2.2.2 Text Mining

Text mining merupakan bagian dari data *mining* dimana proses yang dilakukan utamanya adalah melakukan ekstraksi pengetahuan dan informasi dari pola-pola yang terdapat dalam sekumpulan dokumen teks menggunakan alat analisis tertentu (R. Feldman, 2016). *Text mining* dapat diolah untuk berbagai macam keperluan diantaranya adalah untuk *summarization*, pencarian dokumen teks dan *sentiment analysis*.

Text mining didefinisikan sebagai data yang berupa teks yang biasanya sumber data didapatkan dari dokumen, dengan tujuan adalah mencari kata-kata yang dapat mewakili isi dari dokumen tersebut yang nantinya dapat dilakukan analisa hubungan antar dokumen.

Tahapan dalam *text mining* secara umum dibagi menjadi 6 (enam), yaitu *Knowledge Discovery Goal, Data Preparation, Data Preprocessing, Data Modelling, Evaluation, dan Knowledge and Result.*

2.2.3 Analisis Sentimen

Analisis Sentimen masih bagian dari penelitian *opinion mining* merupakan salah satu bidang dari *Natural Language Processing (NLP)* (Annisa, 2020) yaitu proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini.

Analisis Sentimen adalah kegiatan melakukan analisa terhadap pendapat, opini, sikap atau emosi seseorang mengenai suatu produk, topik atau permasalahan tertentu sehingga bisa diketahui hal tersebut masuk kedalam sentimen positif, negatif atau netral.

Tahapan yang umum dilakukan dalam analisis sentimen dengan *Machine Learning* (Septiar, 2019):

1. Data

Data yang dimaksud adalah data yang sudah dilabeli. Melakukan pelabelan terhadap data adalah usaha pertama yang besar untuk melakukan analisis sentimen.

2. *Preprocessing Data*

Preprocessing adalah suatu rangkaian langkah yang digunakan untuk menghasilkan dataset sesuai dengan kebutuhan. Pada kasus *Text Mining*, umumnya terdapat 5 rangkaian *preprocessing* yaitu *case folding*, *tokenizing*, *normalization*, *stopword removal*, dan *lemmalization*.

3. *Feature Extraction*

Komputer tidak dapat mengolah data selain data numerik, sehingga dibutuhkan langkah untuk mengekstrak data menjadi numerik dalam proses analisis sentimen. Secara umum terdapat 3 teknik ekstraksi fitur yaitu *Bag of Word (TF, IDF)*, *Word Embedding (Glove, Word2vec, FastText)* dan *Character Embedding*.

4. *Modelling*

Pada tahap ini dilakukan pemodelan menggunakan metode yang telah ditentukan untuk dapat melakukan analisis sentimen.

5. *Evaluation*

Model yang terbentuk dari data pelatihan dievaluasi agar diketahui seberapa efektif model tersebut bekerja dalam menyelesaikan permasalahan. Evaluasi model biasanya menggunakan data uji untuk mengetahui seberapa prediktif model.

2.2.4 Klasifikasi Naïve Bayes

Naïve Bayes merupakan teknik prediksi berbasis probabilistik sederhana yang berdasar pada penerapan teorema Bayes dengan asumsi independensi yang kuat (Suryadi, 2017). Metode pengklasifikasian dengan menggunakan metode

probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema Bayes. Naïve Bayes classifier mengasumsikan ada atau tidaknya suatu fitur tertentu pada sebuah kelas tidak mempengaruhi keberadaan fitur lainnya. Ciri utama dr Naïve Bayes Classifier ini adalah asumsi yg sangat kuat (naïf) akan independensi dari masing-masing kondisi / kejadian. Adapun persamaan dari Teorema Bayes dirumuskan sebagai berikut (Bustami, 2013).

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

X = Data dengan kelas yang belum diketahui

H = Hipotesis data merupakan kelas spesifik

$P(H|X)$ = Probabilitas hipotesis H berdasarkan kondisi X (*posterior probability*)

$P(H)$ = Probabilitas Hipotesis H (*prior probability*)

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

$P(X)$ = Probabilitas X