

BAB II

KERANGKA TEORITIS

A. Tinjauan Pustaka

Twitter merupakan salah satu media social yang berdiri pada tahun 2006 yang memungkinkan para pengguna untuk mengirim dan berbagi tautan dengan maksimal 140 karakter namun pada tahun 2017 bertambah hingga 280 karakter yang terkenal dengan istilah *tweet*. Di *twitter* pengguna yang belum memiliki akun hanya bisa membaca *tweet* berbalas melalui aplikasi seluler.

Sampai dengan saat ini *twitter* memiliki pertumbuhan yang sangat pesat hingga pada awal 2019 menempati urutan ke 6 untuk *global ranking*. Dengan popularitas tersebut *twitter* telah dimanfaatkan untuk banyak hal seperti berbagi opini, keluhan hingga ajang kampanye.

B. Landasan Teori

(1) Sentimen Analisis

a. Pengertian Sentimen Analisis

Sentiment analysis adalah jenis natural language yaitu pengolahan kata untuk melacak mood masyarakat tentang produk atau topik tertentu (Vinodhini & Chandrasekaran, 2012). Analisis sentimen atau disebut juga *opinion mining*, adalah bidang studi dalam menganalisis pendapat orang, evaluasi, penilaian, sikap dan emosi terhadap suatu entitas seperti produk, jasa, organisasi, individu, isu-isu, peristiwa dan topik. Fokus utama dari analisis sentimen adalah untuk menyatakan mana yang termasuk opini positif dan mana yang termasuk opini negatif. Salah satu contoh penggunaan analisis sentimen dalam dunia nyata adalah identifikasi kecenderungan pasar dan opini pasar terhadap suatu objek barang.

Selain mengekstraksi sentimen, hal yang biasa orang ingin tahu adalah, kapan terjadi perubahan sentimen dan apa yang menyebabkan sentimen tersebut berubah. Hal ini menjadi penting, karena dengan mengetahui apa yang menyebabkan sentimen berubah, pihak yang bersangkutan bisa mengambil keputusan dengan lebih baik. Misal saja ketika mengetahui suatu topik atau kejadian menyebabkan sentimen turun, maka pihak yang

bersangkutan akan menghindari kejadian serupa untuk meningkatkan sentimen. Kebutuhan-kebutuhan tersebut biasanya muncul ketika suatu pihak ingin mendapatkan sentimen publik yang baik atau melakukan promosi. Kebutuhan seperti ini biasa dimiliki oleh tokoh-tokoh publik, produk dan jasa layanan pada perusahaan untuk mengetahui tentang *review* produk, merek dan layanan untuk menentukan apakah dilihat sebagai hal positif atau negatif. Dalam dunia bisnis, hal ini memungkinkan untuk melacak :

- (1) Persepsi produk baru.
- (2) Persepsi Merek.
- (3) Manajemen reputasi.

Hal ini juga memungkinkan individu untuk mendapatkan sebuah pandangan tentang sesuatu (*review*) pada skala global.

b. Tujuan Sentimen Analis

Adapun tujuan sentimen analis sebagai berikut :

- (1) Untuk menganalisis sentiment postingan terhadap suatu produk atau layanan.
- (2) Untuk memperkirakan kelas yang tidak diketahui dari suatu objek .

c. Manfaat Sentimen Analis

Manfaat yang dapat diambil dari sentiment analisa:

- (1) Mendapatkan informasi yang sebelumnya tidak terstruktur dapat diubah menjadi data yang lebih terstruktur.
- (2) Mendapatkan *feedback* atau masukan dari pengguna suatu layanna hingga mengetahui keinginan layanan yang menunjang keberlangsungan pengguna.

d. Tahapan Sentimen Analisa

Tahapan sentiment analisa melauai tahapan preprocessing dimana pemrosesan teks merupakan proses menggali, mengolah, mengatur informasi dengan cara menganalisis hubungannya, aturan-aturan yang ada di data tekstual semi terstruktur atau tidak terstruktur. Untuk lebih efektif dalam proses pemrosesan dilakukan langkah transformasi data ke dalam suatu format yang memudahkan untuk kebutuhan pemakai. Proses ini disebut *preprocessing* dokumen. Setelah dalam bentuk yang lebih terstruktur dengan adanya proses diatas data dapat dijadikan sumber data yang dapat diolah lebih lanjut. Sama halnya *preprocessing* pada *Information Retrieval (IR)*, tahapannya terdiri dari

normalisasi fitur, tokenisasi, *case folding*, *stopword removal*. Namun pada *preprocessing* analisis sentimen, ada tambahan tahapan seperti *convert emoticon* dan *convert negation*.

1. Normalisasi Fitur

Ada beberapa komponen khas yang biasa ada di *tweet* yakni, *username*, URL, dan "RT" (tanda *retweet*). Karena *username*, URL, dan "RT" tidak memiliki pengaruh apapun terhadap nilai sentimen, maka ketiga komponen di atas akan dibuang. Komponen *username* diidentifikasi dengan kemunculan karakter '@', sedangkan komponen URL dikenali melalui ekspresi reguler (*http*).

2. Case Folding

Case folding adalah mengubah semua huruf dalam dokumen menjadi huruf kecil. Hanya huruf 'a' sampai dengan 'z' yang diterima. Karakter selain huruf dihilangkan dan dianggap delimiter. Tidak semua dokumen teks konsisten dalam penggunaan huruf kapital. Oleh karena itu, peran **Case Folding** dibutuhkan dalam mengkonversi keseluruhan teks dalam dokumen menjadi suatu bentuk standar (biasanya huruf kecil atau *lowercase*). Sebagai contoh, user yang ingin mendapatkan informasi "KOMPUTER" dan mengetik "KOMPOTER", "KomPUter", atau "komputer", tetap diberikan hasil retrieval yang sama yakni "komputer".

3. Convert Negation

Pada tahap ini, setiap *tweets* yang mengandung kata-kata yang bersifat negasi akan diubah nilai sentimennya. Kata-kata yang bersifat negasi seperti "bukan", "tidak", "enggak", "ga", "jangan", "nggak", "tak", dan "gak". *Convert negation* dilakukan jika terdapat kata negasi sebelum kata yang bernilai positif, maka kata tersebut akan diubah nilainya menjadi negatif dan begitupun sebaliknya.

4. Stopword Removal

Stopword removal adalah tahap mengambil kata-kata penting dengan menggunakan algoritma *stoplist* (membuang kata kurang penting) atau *wordlist* (menyimpan kata penting). *stopword* adalah kata umum yang biasanya muncul dalam jumlah besar dan dianggap tidak memiliki makna. Contoh *stopword* dalam bahasa Indonesia adalah "yang", "dan", "di", "dari", dll. Makna di balik penggunaan *stopword* yaitu dengan

menghapus kata-kata yang memiliki informasi rendah dari sebuah teks, kita dapat fokus pada kata-kata penting sebagai gantinya. Contoh penggunaan *filtering stopwords* dapat kita temukan pada konteks mesin pencarian. Jika permintaan pencarian anda adalah “apa itu pengertian manajemen?” tentunya anda ingin sistem pencarian fokus pada memunculkan dokumen dengan topik tentang “pengertian manajemen” di atas dokumen dengan topik “apa itu”. Hal ini dapat dilakukan dengan mencegah kata dari daftar *stopword* yang dianalisa.

(2) **Text Mining**

Text mining juga dikenal sebagai *data mining* teks [8] atau penemuan pengetahuan dari database tekstual [9]. Sesuai dengan buku *The Text Mining Handbook* [10], text mining dapat didefinisikan sebagai suatu proses menggali informasi dimana seorang *user* berinteraksi dengan sekumpulan dokumen menggunakan *tools* analisis yang merupakan komponen-komponen dalam data mining. Tujuan dari *text mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Jadi, sumber data yang digunakan *dalam text mining* adalah sekumpulan teks yang memiliki format yang tidak terstruktur atau minimal semi terstruktur.

Adapun tugas khusus dari *text mining* antara lain yaitu :pengkategorian teks dan pengelompokan teks. *Text mining* dapat memberikan solusi dari permasalahan seperti pemrosesan, pengorganisasian / pengelompokan dan menganalisa *unstructured data* dalam jumlah besar, dalam hal ini data yang akan digunakan adalah data yang diambil dari Twitter. Dalam memberikan solusi, *text mining* mengadopsi dan mengembangkan banyak teknik dari bidang lain, seperti *Data Mining, Information Retrieval, Statistik dan Matematik, Machine Learning, Linguistic, Natural Language Processing dan Visualization*. Kegiatan riset untuk *text mining* antara lain ekstraksi dan penyimpanan teks, *preprocessing* akan konten teks, pengumpulan data statistik serta *indexing* dan analisis sentiment.

(3) **Naïve Bayes Classifier**

Algoritma yang digunakan untuk mencari nilai probabilitas tertinggi untuk mengklasifikasikan data pada kategori yang paling tepat (Feldman & Sanger 2007). Dalam penelitian ini data uji adalah data dokumen tweet.

Dalam algoritma *naïve bayes classifier* setiap dokumen direpresentasikan dengan pasangan atribut “x1, x2, x3,...xn” dimana x1 adalah kata pertama, x2 adalah kata kedua dan seterusnya. Sedangkan V adalah himpunan kategori *Tweet*.

$$V_{MAP} = \frac{\arg \max_{V_j \in V} P(x_1, x_2, \dots, x_n | V_j) P(V_j)}{P(x_1, x_2, \dots, x_n)}$$

Keterangan:

V_{MAP} : semua kategori yang diujikan

V_j : Kategori komentar, dengan:

x₁ : Sentimen positif

x₂ : Sentimen Negatif

x₃ : Sentimen Netral

P(X₁|V_J) : probabilitas Xi pada kategori V_j

P(V_j) : probabilitas dari V

(4) *Bag Of Words*

BoW, seperti namanya, adalah sebuah tas (kresek) berisi semua kata pada teks yang akan diproses. Pertama teks (atau kalimat) dibagi menjadi beberapa bagian, proses ini diistilahkan tokenization. Pada proses ini yang diambil hanya kata, tanda baca tidak. Terdapat dua teks.

Tabel 2.1 Bag Of words

	ini	ibu	budi	pergi	ke	pasar	bersama
Ini ibu Budi	1	1	1	0	0	0	0
Ibu Budi pergi ke pasar bersama Budi.	0	1	2	1	1	1	1

Tiap kolom merepresentasikan **kata** dan baris adalah **jumlah kemunculan** kata tersebut pada teks. Tiap baris pada tabel di atas dapat juga diistilahkan sebagai vektor. Pada **teks pertama** kita mendapatkan sebuah vektor **[1, 1, 1, 0, 0, 0, 0]**, **[0, 1, 2, 1, 1, 1, 1]** pada baris kedua dan begitu seterusnya. Ada variasi lain dari BoW yaitu dengan mengambil **sebanyak**

n kata untuk tiap kolom (kolom diistilahkan pula sebagai *feature*), konsep ini disebut *n-gram* menggunakan satu kata.

(5) **Unified Modeling Language (UML)**

UML adalah bahasa pemodelan untuk sistem atau perangkat lunak berorientasi objek. Pemodelan sesungguhnya digunakan untuk penyederhanaan permasalahan permasalahan yang kompleks sehingga lebih mudah dipelajari dan dipahami. UML digunakan untuk melakukan analisis dan perancangan sistem/perangkat lunak. UML merupakan bahasa standar untuk merancang dan mendokumentasikan perangkat lunak yang berorientasi objek.

a. **Use Case Diagram**

Use case diagram digunakan untuk memodelkan bisnis proses berdasarkan perspektif pengguna sistem. *Use case diagram* terdiri atas diagram untuk *use case* dan *actor*. *Actor* merepresentasikan orang yang akan mengoperasikan atau orang yang berinteraksi dengan sistem aplikasi.

b. **Activity Diagram**

Activity diagram adalah representasi grafis dari alur kerja tahapan aktifitas. Diagram ini mendukung pilihan tindakan, iterasi dan *concurrency*. Pada pemodelan UML, *activity diagram* dapat digunakan untuk menjelaskan bisnis dan alur kerja operasional secara *step by-step* dari komponen suatu sistem. *Activity diagram* menunjukkan keseluruhan dari aliran kontrol.

c. **Sequence Diagram**

Sequence diagram menjelaskan secara detail urutan proses yang dilakukan dalam sistem untuk mencapai tujuan dari *use case*. Interaksi yang terjadi antar *class*, operasi apa saja yang terlibat, urutan antar operasi, dan informasi yang diperlukan oleh masing-masing operasi.

C. Tinjauan Studi

Penelitian rujukan merupakan acuan yang dibutuhkan peneliti untuk melakukan penelitian dalam kasus yang sama dengan metode yang berbeda sebagai bahan pertimbangan pada penelitian ini dan untuk mengetahui perbedaan penelitian sebelumnya dengan penelitian yang akan dilakukan. Berikut adalah penelitian yang telah dilakukan sebelumnya:

Tabel 2.2 Tinjauan Studi

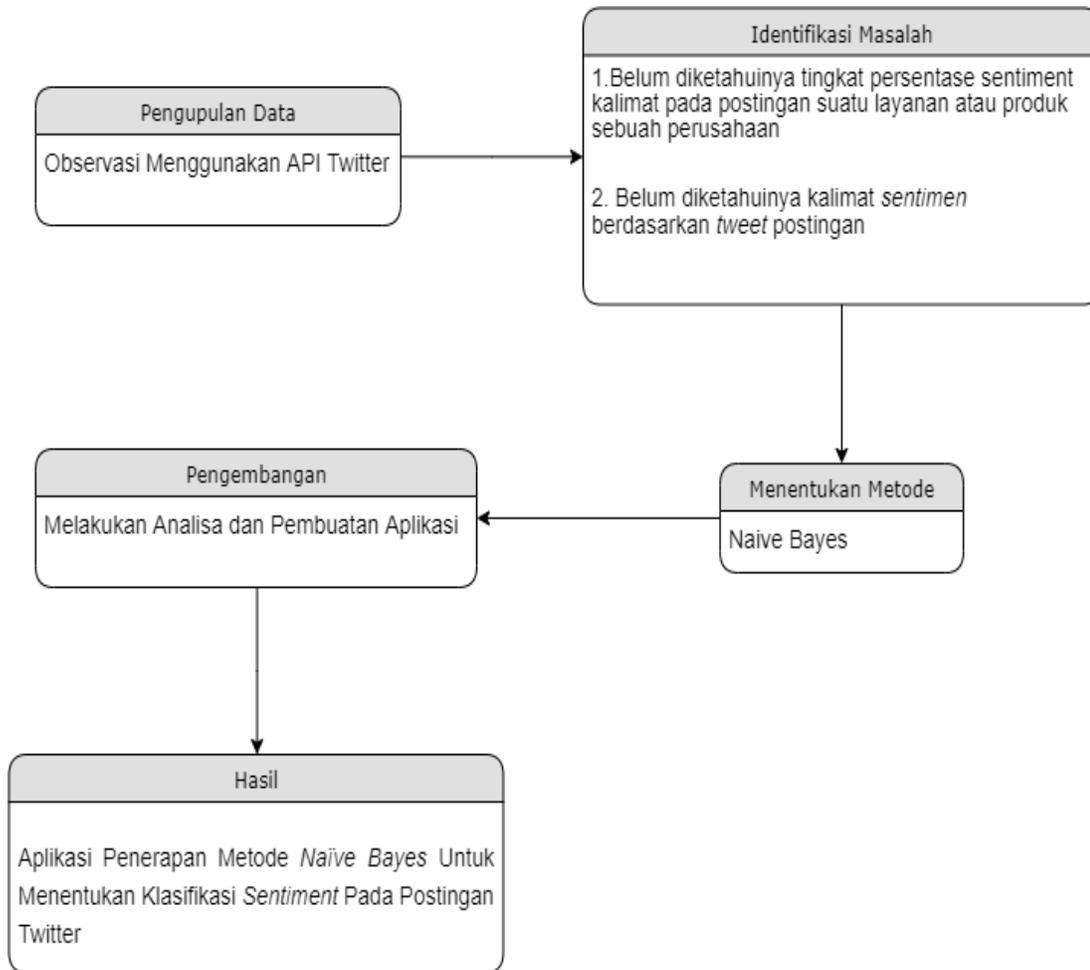
No	Peneliti	Judul	Permasalahan	Perbedaan
1.	Indra Budi, (2015)	Analisis Sentimen Masyarakat Terhadap Pemerintahan Jokowi Menggunakan Data Twitter	• Data tweet yang terkumpul dari hasil crawling terbagi menjadi tiga domain yakni politik, hukum dan ekonomi Keluar dari program tersebut menghasilkan tiga file yaitu data bersentimen positif, negatif dan netral Rata-rata persentase hasil penelitian 51.29% bersentimen netral	Pada tweet yang mengandung retweeted tidak dihapus. Hal tersebut dikarenakan bisa saja pengguna hanya meretweeted tweet yang memiliki sentimen. Hal ini dapat menambahkan jumlah data klasifikasi
2.	Ahmad Fathan Hidayatullah, Azhari SN. (2014)	Analisis Sentimen Dan Klasifikasi Kategori Terhadap Tokoh Publik Pada Twitter	Rancangan sistem dalam penelitian terbagi menjadi beberapa bagian diantaranya retrieval module, preprocessing, feature selection, dan classification module. Klasifikasi menggunakan tools rapidminer	Menggunakan tools rapidminer sedangkan yang saya lakukan menggunakan data twitter yg diambil menggunakan minig dari twitter API
3.	Ghulam Asrofi Buntoro, (2017)	Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter	Nilai akurasi tertinggi didapat saat menggunakan metode klasifikasi Naive Bayes Classifier (NBC) untuk klasifikasi data AHY, dengan nilai rata rata akurasi mencapai 95%, nilai presisi 95%, nilai recall 95% nilai TP rate 96,8% dan nilai TN rate 84,6%.	Hanya membandingkan akurasi dari NB classification dan SVM sedangkan yang saya ambil hanya menggunakan Deep belief network dan mengimplemetasikan polaritas kata tweet
4.	I Putu Eka N. Kencana	Analisis Sentimen Menggunakan Metode Naive Bayes Classifier Dengan Seleksi Fitur Chi Square	Terdapat kesalahan klasifikasi karena pada data uji terdapat fitur yang muncul pada bukan kategorinya. Untuk penelitian selanjutnya yaitu dapat dibandingkan hasil seleksi fitur dari chi square terhadap hasil seleksi fitur dari Mutual Information berdasarkan segi	Hanya Menggunakan pengujian 100 data saja sedangkan yang saya ajukan bias memuat lebih besar data

			waktu komputasi dan segi ketepatan klasifikasi	
5.	Hartanto	Text Mining Dan Sentime Analisis Twitter Pada Gerakan Lgbt	Metode "bag of words" dimana memakai bantuan algoritma khusus dan words list yang berperan sebagai "filter" dari teks Yang dianalisis (polarity) dan melakukan ekstraksi sentiment atau opini sebagai hasilnya. Kedua dengan metode statistic konsep hubungan dimana mengukur polaritas bukan dari word list namun dengan cara melakukan "comparing" terhadap satu set variable baru setelah itu dilakukan ekstraksi opini /sentiment.	Terbatasnya bag of word di atasi oleh word2vec yang memiliki kemampuan tinggimenebak persamaan kata sedangkan pengujian disini hanya pada tweet berbahasa Indonesia pada data olahnya
6.	Robet Habibi	Analisis Sentimen Pada Twitter Mahasiswa Menggunakan Metode Backpropagation	Mengenali sentimen pada twitter mahasiswa. Sentimen yang terkandung merupakan representasi emosi dan kecenderungan emosi mahasiswa. Hasil pengenalan kecenderungan Emosi mahasiswa digunakan sebagai acuan untuk memberikan perlakuan yang sesuai sehingga dapat meningkatkan keberhasilan dalam proses belajar mahasiswa.	Pada penelitian ini peneliti sebelum nya hanya menganalisa dan mengklasifikasikan keadaan psikologis menggunakan data set sedangkan penelitian yang saya lakukan lebih ke mengklasifikasikan polaritas dari kata
7.	ASWINDA PRIMA PUTRA	Analisis Sentimen Data Twitter Menggunakan Naïve Bayes Dengan Negation Handling Pada Data Twitter Bahasa Indonesia	Membandingkan akurasi Dari analisis sentiment dengan Negation Handling dan tanpa menggunakan Negation Handling	Menggunakan data yang lebih banyak dan juga mempunyai tweet negasi yang lebih banyak sehingga meningkatkan peluang dari kata negasi

8.	Ira Zulfa	Sentimen Analisis Tweet Berbahasa Indonesia Dengan Deep Belief Network	Sistem ini tidak dapat digunakan untuk menentukan sentimen tweet dari bahasa selain bahasa Indonesia.	Penelitian yang saya buat multilingual sentimen klasifikasi yaitu sistem yang dapat menentukan sentimen dari berbagai bahasa dengan data training yang sudah di listing.
9.	Nabila Safina	Analisis Sentimen Pada Twitter Terhadap Jasa Transportasi Online DiIndonesia Dengan Metode Support Vector Mechine	Membandingkan dua Competitor transportasi online hanya menggunakan mention	Menggunakan data hasil minning dari API twitter sendiri dengan mencari kata baik tagar ataupun mention yang diinginkan
10.	Faishol Nurhuda	Analisis Sentimen Masyarakat Terhadap CalonPresiden Indonesia 2014 Berdasarkan Opini Dari Twitter Menggunakan Metode Naive Bayes Classifier	Hanya membandingkan kedua pasangan saja tanpa tahu adanya parameter yang membatasi	Menggunakan data olah dan data training yang sudah di set lalu dicocokkan pada data yang di ambil

D. Kerangka Berpikir

Kerangka berfikir merupakan arah penalaran untuk bisa memberikan jawaban sementara atas rumusan masalah yang telah disebutkan. Metode Naïve bayes pada dasarnya merupakan salah satu metode referentif yang bisa dijadikan untuk memecahkan masalah penentuan pemenang Tender dimana masalah ini termasuk kedalam kategori penelitian kuantitatif. Adapun kerangka pemikiran ini dapat ditujukan seperti berikut :



Gambar 2.1 Kerangka Berfikir

E. Hipotesis Penelitian

Penerapan Metode *Naive Bayes* diduga dapat menyelesaikan masalah dalam menentukan persentase sentiment kalimat pada postingan kalimat *twitter*.