BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Tinjauan pustaka yang pertama adalah karya tulis dari Noviah Dwi Purtanti dan Edi Winarko pada tahun 2014 dengan judul Analisis Sentimen Twitter untuk Teks Berbahasa Indonesia dengan menggunakan metode Maximum Entropy dan Support Vector Machine. Data opini diperoleh dari jejaring sosial Twitter berdasarkan *query* dalam bahasa Indonesia. Analisis sentimen dalam penelitian ini merupakan proses klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen positif dan negatif. Implementasi klasifikasi diperoleh akurasi 86,81 % pada pengujian 7 *fold cross validation* untuk tipe kernel Sigmoid. Pelabelan kelas secara manual dengan POS *tagger* menghasilkan akurasi 81,67%.

Tinjauan pustaka yang ke dua adalah karya tulis dari Ahmad Fathan Hidayatullah dan Azhari SN pada tahun 2014 dengan judul Analisis Sentimen dan Klasifikasi Kategori terhadap Tokoh Publik pada Twitter. Pengambilan data pada karya tulis ini dengan menggunakan *cron job* dan diproses dengan menggunakan metode Support Vector Machine dengan bantuan software RapidMiner. Hasil dari penelitian pada karya tulis tersebut adalah tingkat akurasi metode Support Vector Machine sebagai metode yang lebih akurat dengan menggunakan data yang diambil dari Twitter dengan akurasi sebesar 79,66%.

Tinjauan pustaka yang ke tiga adalah karya tulis dari Rizky Maulana pada tahun 2016 dengan judul Penerapan Analisis Sentimen Pengguna Twitter menggunakan metode Support Vector Machine berbasis Cloud Computing. Data

dari penelitian ini yaitu Twitter Tokoh Publik dengan hasil analisis sentimen positif dan negatif pengguna twitter terhadap tokoh publik secara *real time*. Tweet yang sudah melalui tahapan *preprocessing* dan *streaming* serta perubahan format data yang sesuai dengan LibSVM menggunakan linier kernel dengan hasil akurasi sebesar 79,5%.

Tinjauan pustaka yang ke empat adalah Penelitian tentang "Analisis Sentimen pada Review Restoran dengan Teks Bahasa Indonesia Menggunakan Algoritma Naïve Bayes" dilakukan oleh Muthia (2017). Dari pengolahan data yang sudah dilakukan, Genetic Algorithm terbukti dapat meningkatkan akurasi pengklasifikasian Naïve Bayes. Data *review* restoran dapat diklasifikasi dengan baik ke dalam bentuk positif dan negatif. Akurasi Naïve Bayes sebelum menggunakan penggabungan metode pemilihan fitur mencapai 86,50%. Sedangkan setelah penggabungan metode pemilihan fitur, yaitu Genetic Algorithm, akurasinya meningkat mencapai 90,50%.

Tinjauan pustaka yang ke lima adalah karya tulis dari Nazuar Adnan pada tahun 2017 dengan judul "Analisis sentimen dan klasifikasi topik pada headline berita online". Dalam penelitian ini akan dibuat sistem untuk mengklasifikasi apakah berita tersebut termasuk berita positif, negatif, atau netral dan mengklasifikasi topik dari berita tersebut. Kasus yang akan di klasifikasi merupakan kasus yang umum yang muncul di berita. Seperti politik, kriminal, ekonomi, olahraga, nasional, dunia, teknologi dan lain-lain. Analisis sentimen dengan menggunakan Naive Bayes Classifier menghasilkan akurasi sebesar 74,2% untuk sentimen dan 32% untuk topik.

Penelitian ini melakukan analisis dan klasifikasi sentimen terhadap Twitter STMIK AKAKOM Yogyakarta. Metode yang digunakan dalam klasifikasi kategori adalah Naïve Bayes Classifier Pada Penelitian ini *Analysis Sentiment* digunakan untuk menentukan polaritas sentimen mana yang mengandung sentimen positif, sentimen netral dan sentimen negative secara otomatis atau *realtime*.

Tabel 2.1. Tinjauan Pustaka

Peneliti	Tahun	Data Penelitian	Metode	Hasil
Noviah Dwi Purtanti	2014	Twitter teks berbahasa Indonesia	Maximum Entropy dan Support Vector Machine	Klasifikasi dokumen tekstual ke dalam dua kelas, yaitu kelas sentimen positif dan negatif.
Ahmad Fathan Hidayatullah	2014	Twitter Tokoh Publik	Support Vector Machine	klasifikasi tweet berdasarkan sentimen dan kategori yang berasal dari fitur yang dimiliki oleh tokoh publik.
Rizky Maulana	2016	Twitter Tokoh Publik	Support Vector Mchine	Analisis sentimen positif dan negatif pengguna twitter terhadap tokoh publik secara real time.
Dinda Ayu Mutia	2017	*Review* Restoran dengan teks bahasa Indonesia	Naïve Bayes Classifier dan Genetic algorithm	Review restoran diklasifikasian dalam bentuk positif dan negatif.
Nazuar Adnan	2017	Topik pada headline berita online	Naive Bayes Classifier	Analisis Sentimen dan Klasifikasi Topik pada Headline Berita Online Untuk mengklasifikasi apakah berita tersebut termasuk berita positif, negatif, atau netral dan mengklasifikasi topik dari berita tersebut.

Tabel 2.2. Tinjauan Pustaka (Lanjutan)

Peneliti	Tahun	Data Penelitian	Metode	Hasil
Penelitian yang dilakukan	2017	Twitter STMIK AKAKOM Yogyakarta	Naïve Bayes Classifier	Analisis sentimen yang mengandung sentimen positif, netral dan negatif secara real time dan input manual.

2.2. Dasar Teori

2.2.1. Twitter

Twitter adalah salah satu layanan *microblogging* yang cukup terkenal dan memungkinkan para penggunanya untuk menulis atau membuat status yang sering dinamakan kicauan atau tweet. 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. Twitter API

Seperti dikatakan oleh Ade Suryansyah (2014). Bahwa API (*Application Programming Interface*) merupakan sekumpulan perintah, fungsi, dan protokol yang dapat digunakan dalam membangun perangkat lunak untuk sistem operasi tertentu, juga merupakan suatu dokumentasi yang terdiri dari antar muka, fungsi,

kelas, struktur untuk membangun *software.* Seorang *programmer* akan lebih mudah dalam membongkar suatu *software* untuk kemudian dapat dikembangkan atau diintegritasikan dengan perangkat lunak yang lain melalui API. Dengan demikian API dapat dikatakan sebagai penghubung antar aplikasi yang satu dengan aplikasi yang lainnya. Suatu rutin *standart* yang memungkinkan *developer* menggunakan *system function* dimana *operation system* berperan dalam mengelola hal ini. Sehingga API ini mempunyai keunggulan dalam hal interaksi antar aplikasi.

Sejalan dengan pendapat Suryansyah (2014). Keuntungan menggunakan API adalah sebagai berikut :

1. Portabilitas.

   Developer yang menggunakan API dapat menjalankan programnya dalam sistem operasi apapun asalkan sudah ter-install API tersebut.

2. Lebih mudah dimengerti.

   API menggunakan bahasa yang lebih terstruktur dan mudah di mengerti daripada bahasa system call. Hal ini sangat penting dalam hal editing dan pengembangan.

Menurut (Togias & Kemeas 2012) API digunakan untuk menggabungkan dua sumber yang berbeda untuk membuat suatu program aplikasi yang saling berintegrasi.

Sehingga dari kedua pendapat yang ada bisa disimpulkan bahwa API atau yang biasa disebut *Application Programming Interface* adalah suatu program atau

aplikasi yang disediakan oleh pihak *developer* tertentu agar pihak pengembang aplikasi lainnya dapat lebih mudah mengakses aplikasi tersebut, intinya API ini berfungsi sebagai jembatan antara aplikasi satu dengan aplikasi yang lain.

Twitter API yaitu sebuah aplikasi yang diciptakan oleh pihak Twitter agar mempermudah pihak *developer* lain untuk mengakses informasi web Twitter tersebut dengan ketentuan dan syarat yang berlaku seperti yang terdapat pada http://dev.twitter.com/oauth.

Ada beberapa jenis Twitter API :

1. Twitter *REST* API

   Terdiri dari Twitter REST dan Twitter Search. Twitter REST memberikan *core data* dan *core twitter objects*. Twitter *search* berfungsi untuk melakukan pencarian mengenai suatu *instance* objek Twitter maupun mencari *trend*.

2. Twitter *Streaming* API

   API ini biasa digunakan untuk penggalian data karena melalui API ini informasi bisa didapatkan secara realtime dengan volume yang sangat tinggi.

2.2.3. 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 diantarannya adalah untuk *summarization*, pencarian dokumen teks dan *sentiment analisys*.

Text mining bertujuan untuk mencari kata-kata yang dapat mewakili apa yang ada didalam dokumen sehingga dapat dilakukan analisa keterhubungan antar dokumen. Text mining mempunyai 5 tahapan yaitu Tokenizing, Filtering, Stemming, Tagging, dan Analyzing.

Tahapan Tokenizing adalah proses pemotongan string masukan berdasarkan tiap kata yang menyusunnya. Pada prinsipnya proses ini adalah memisahkan setiap kata yang menyusun suatu dokumen. Tahapan Filtering adalah suatu proses dimana diambil sebagian data dari data tertentu, dan membuang data pada frekuensi yang lain. Tahapan Stemming adalah proses pemetaan dan penguraian berbagai bentuk (variants) dari suatu kata menjadi bentuk kata dasarnya. Tahapan Tagging adalah kata yang belum lama dilahirkan. Dahulu sebelum ada tagging, dunia informasi yang ada di internet berserakan dan tidak tersusun berdasarkan kategorinya. Hal itu bagaikan, perpustakaan tanpa ada pengurusnya atau pustakawan. Tahapan Analyzing yaitu untuk mencari seberapa jauh keterhubungan antar kata-kata setiap dokumen.

2.2.4. Sentiment Analisys

Sentiment Analisys atau *opinion mining* merupakan proses memahami, mengekstrak dan mengolah data tekstual secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat pendapat atau kecendrungan opini terhadap sebuah

masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif (B. Liu. 2010).

Sentiment analysis 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.

2.2.5. Naïve Bayes Classifier

Bayesian classification didasarkan pada teorima bayes yang memiliki kemampuan hampir serupa dengan *decision tree* dan *neural network*. Teorema Bayes adalah teorema yang digunakan dalam statistika untuk menghitung peluang suatu hipotesis. Bayes merupakan teknik prediksi berbasis *probabilistic* sederhana yang berdasar pada penerapan teorema Bayes (atau aturan Bayes) dengan asumsi independensi (ketidaktergantungan) yang kuat atau naïf (Eko Prasetyo, 2012). Dengan kata lain, dalam Naïve Bayes, model yang digunakan adalah "model fitur independen"

Dalam Bayes (terutama Naïve Bayes), maksud independensi yang kuat pada fitur adalah bahwa sebuah fitur pada sebuah data tidak berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Eko Prasetyo, 2012).

Bayesian classification adalah suatu metode pengklasifikasian data dengan model statistik yang dapat digunakan untuk menghitung probabilitas keanggotaan suatu kelas. Metode Bayesian classification digunakan menganalisis dalam membantu tercapainya pengambilan keputusan terbaik suatu permasalahan dari

sejumlah alternatif. *Bayesian classification* merupakan salah satu metode yang sederhana yang dapat digunakan untuk data yang tidak konsisten dan data bias. Metode bayes juga merupakan metode yang baik dalam mesin pembelajaran berdasarkan data training dengan berdasarkan probabilitas bersyarat.

Kaitan antara Naïve Bayes dengan klasifikasi, korelasi hipotesis, dan bukti dengan klasifikasi adalah hipotesis dalam teorema Bayes merupakan label kelas yang menjadi target pemetaan dalam klasifikasi, sedangkan bukti merupakan fitur-fitur yang menjadi masukan dalam model klasifikasi (Eko Prasetyo, 2012).

Dalam penelitian ini yang menjadi data uji adalah data headline. Ada dua tahap pada klasifikaasi dokumen. Tahap pertama adalah pelatihan terhadap dokumen yang sudah diketahui kategorinya. Sedangkan tahap kedua adalah proses klasifikasi dokumen yang belum diketahui kategorinya.

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 headline. Pada saat klasifikasi algoritma akan mencari probabilitas tertinggi dari semua kategori dokumen yang diujikan (V_{MAP}), dimana persamaannya adalah sebagai berikut :

$$V_{MAP} = \underset{V\ j\ e\ V}{\arg\ max}\ \frac{P(x_1, x_{2,} x_{3,\ldots} x_n|Vj)P(Vj)}{P(x_1,x_2,x_{3,\ldots}x_n|Vj)} \quad \text{..... (2.1)}$$

Untuk P(x1, x2, x3,...xn) nilainya konstan untuk semua kategori (Vj) sehingga persamaan dapat ditulis sebagai berikut :

$$V_{MAP} = \underset{V\ j\ e\ V}{\arg max}\ P(x_1, x_{2,} x_{3,\ldots} x_n | Vj) P(Vj) \quad \text{..... (2.2)}$$

Persamaan diatas dapat disederhanakan menjadi sebagai berikut :

$$V_{MAP} = \underset{V\ j\ e\ V}{\arg max} \prod_{i=1}^{n} P(x_1|Vj) P(Vj) \quad \text{..... (2.3)}$$

Keterangan :

Vj = Kategori Headline

J = 1, 2, 3,…n. Dimana dalam penelitian ini

j1 = kategori headline sentimen positif

j2 = kategori headline sentimen negatif

j3 = kategori headline sentiment netral

P(xi|Vj) = Probabilitas xi pada kategori Vj

P(Vj) = Probabilitas dari Vj

Untuk P(Vj) dan P(xi|Vj) dihitung pada saat pelatihan dimana persamaannya adalah sebagai berikut :

$$P(Vj) = \frac{|docs\ j|}{|contoh|} \quad \text{..... (2.4)}$$

Jika P(Vj) sudah ditentukan maka hitung jumlah dokumen setiap kategeri j dan jumlah dokumen dari semua kategori dengan menggunakan rumus 2.4. diatas.

$$P(xi|Vj) = \frac{nk+1}{nk+|kosakata|} \quad \text{..... (2.5)}$$

Jika P(xi|Vj) sudah ditentukan maka hitung jumlah frekuensi kemunculan setiap kata ditambah 1 dan jumlah frekuensi kemunculan kata dari setiap kategori dengan menggunakan rumus 2.5. diatas.

Keterangan :

|docs j| = jumlah dokumen setiap kategori j

|contoh| = jumlah dokumen dari semua kategori

nk = jumlah frekuensi kemunculan setiap kata

n = jumlah frekuensi kemunculan kata dari setiap kategori

|kosakata| = jumlah semua kata dari semua kategori

Analisis Penerapan Algoritma Naïve Bayes Classifier :

Pada pengklasifikasian menggunakan Naïve Bayes dibagi kedalam 2 proses, yaitu proses training dan testing. Proses training digunakan untuk menghasilkan model analisis sentimen yang nantinya akan digunakan sebagai acuan untuk mengklasifikasikan sentimen dengan data testing atau data mentah yang baru.