

BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1. Tinjauan Pustaka

Tinjauan pustaka yang pertama penelitian yang ditulis oleh Falahah, Dyar Dwiki dan Adriadi Nur pada tahun 2015. Peneliti melakukan klasifikasi data dari twitter yang berhubungan dengan Pemerintah kota Bandung menjadi tiga kategori yaitu positif, negatif dan netral. Tingkat akurasi yang diperoleh dengan data testing sebanyak 300 *tweet* dan dilakukan tiga kali pengujian yaitu sebesar 73%, 68% dan 58%.

Tinjauan pustaka yang kedua penelitian yang ditulis oleh Nuke Y. A. Faradhillah, Renny P. Kusumawardani dan Irmasari Hafidz yang ditulis pada tahun 2016. Data yang digunakan dalam penelitian ini diperoleh dari crawling dua akun twitter yaitu @e100ss dan @SapawargaSby periode 1 September 2015 sampai dengan 13 Oktober 2015. Dari 2 metode algoritma yang digunakan yaitu *Naive Bayes* dan *Support Vector Machine* (SVM) didapatkan hasil akurasi lebih tinggi dengan metode *Support Vector Machine* (SVM).

Tinjauan pustaka yang ketiga penelitian yang ditulis oleh Ahmad Fathan Hidayatullah dan Azhari SN pada tahun 2014. Data yang digunakan diambil dari twitter menggunakan *Twitter Search API* v1.1 dengan kata kunci yaitu nama tokoh publik pada tahun 2014. Hasil akurasi yang diperoleh dengan menggunakan 2

metode algoritma pada *RapidMiner* yaitu untuk metode *Naive Bayes Classifier* sebesar 73,81% dengan fitur *Term frequency* dan 71,11% dengan fitur *TF-IDF*. Sedangkan untuk metode *Support Vector Machine* (SVM) sebesar 83,14% dengan fitur *Term frequency* dan 82,69% dengan fitur *TF-IDF*.

Tinjauan pustaka yang keempat ditulis oleh Naufal Riza Fatahillah pada tahun 2017 dengan judul Implementasi Algoritma *Naive Bayes Classifier* Pada Social Media (Twitter) Terhadap Ujaran Kebencian Berbahasa Indonesia. Pengambilan data pada penelitian ini menggunakan API twitter berdasar *hashtag* dengan batas waktu satu minggu yang selanjutnya diklasifikasikan menjadi negatif atau positif dengan menggunakan metode *Naive Bayes Classifier*.

Tinjauan pustaka yang kelima ditulis oleh Yuna Sophia Dewi Febriant pada tahun 2017 dengan judul Analisis Dan Klasifikasi Sentimen Terhadap Twitter STMIK Akakom Yogyakarta Menggunakan *Naive Bayes Classifier*. Data yang diteliti adalah *tweet* yang berhubungan dengan STMIK Akakom Yogyakarta dan diklasifikasikan dalam *tweet* positif, netral atau negatif dengan metode *Naive Bayes Classifier*.

Pada penelitian ini sistem akan melakukan analisis dan klasifikasi terhadap Twitter Pemerintah Provinsi Jawa Tengah. Kategori yang akan dipilih yaitu Ekonomi dan Layanan Sosial dengan data yang diambil yaitu periode satu tahun ke belakang. Sentimen yang akan dihasilkan yaitu positif, negatif, dan netral

dengan menggunakan metode *Naive Bayes Classifier*.

Tabel 2.1 Tinjauan Pustaka

No	Penulis	Tahun	Objek	Metode	Hasil
1.	Falahah	2015	Data teks twitter pemerintah kota bandung	Naive Bayes	Kategori <i>tweet</i> yaitu negatif, positif, netral dengan akurasi pengujian sebanyak tiga kali yaitu 73%, 68% dan 58%.
2.	Nuke Y. A. Faradhillah	2016	Data teks twitter pemerintah kota surabaya	Naive Bayes dan Support Vector Machine (SVM)	Nilai akurasi sebesar 67,05%. hasil kelas positif 50%, negatif 70,96%, dan netral 63,99%.
3.	Ahmad Fathan Hidayatullah	2014	Data teks twitter tokoh publik populer tahun 2014	Naive Bayes dan Support Vector Machine (SVM)	Klasifikasi dengan <i>Naive Bayes</i> menghasilkan akurasi 73,81% untuk fitur <i>term frequency</i> dan 71,11% untuk TF-IDF. Klasifikasi dengan <i>Support Vector Machine</i> menghasilkan akurasi 83,14% untuk fitur <i>term frequency</i> dan 82,69% untuk TF-IDF.
4.	Naufal Riza Fatahillah	2017	Data teks twitter berdasar hastag tertentu	Naive Bayes Classifier	Pengujian klasifikasi menggunakan <i>Naive Bayes Classifier</i> menghasilkan tingkat akurasi sebesar 93%.
5.	Yuna Sophia Dewi Febriant	2017	Twitter STMIK Akakom Yogyakarta	Naive Bayes Classifier	Klasifikasi <i>tweet</i> positif, negatif, netral dengan hasil akurasi pengujian menggunakan metode <i>Naive Bayes Classifier</i> sebesar 78%.

6.	Penelitian yang dilakukan	2018	Twitter Pemerintah Provinsi Jawa Tengah	Naive Bayes Classifier	Sentimen positif, negatif dan netral.
----	---------------------------	------	---	------------------------	---------------------------------------

2.2. Dasar Teori

2.2.1. Twitter

Twitter merupakan suatu layanan sosial media yang berkategori *microblogging*. Maksud dari *microblogging* ini yaitu sebuah layanan *blog* yang singkat dalam sebuah paragraph. Pada intinya twitter merupakan suatu tempat layanan informasi yang bisa kita bagikan kepada banyak orang diseluruh dunia. Twitter terdiri dari pesan singkat sebanyak 140 karakter yang diibaratkan sebagai kicauan burung dan dinamakan dengan *tweet*. Pengguna dapat menulis pesan dengan tanda #(tagar) untuk menentukan topik. Sedangkan tanda @ digunakan untuk menyebut atau membalas pesan dari pengguna lain.

2.2.2. Twitter API

Dalam laman resmi <https://help.twitter.com/> dikatakan API (*Application Programming Interface*) merupakan cara program komputer “berbicara” satu sama lain agar mereka dapat meminta dan menyajikan informasi. Hal ini dilakukan dengan mengizinkan aplikasi perangkat lunak memanggil apa yang disebut *endpoint*. Twitter mengizinkan akses ke bagian layanan melalui API untuk memungkinkan orang-orang membangun perangkat lunak yang terintegrasi dengan Twitter seperti solusi yang membantu sebuah perusahaan menjawab

umpan balik pelanggan Twitter.

Platform API Twitter menyediakan akses luas ke data Twitter publik yang telah dipilih pengguna untuk dibagikan ke dunia. Twitter juga mendukung API yang memungkinkan pengguna mengelola informasi Twitter mereka yang non-publik dan memberikan informasi ini ke pengembang yang telah diizinkan pengguna untuk melakukannya.

2.2.3. Text mining

Text Mining memiliki definisi menambang data yang berupa teks dimana sumber data biasanya di dapatkan dari dokumen, dan tujuannya adalah mencari kata-kata yang dapat mewakili isi dari dokumen sehingga dapat dilakukan analisis keterhubungan antar dokumen (Hendro Nindito, 2016).

Tujuan *Text Mining* adalah untuk mendapatkan informasi yang berguna dari sekumpulan dokumen. Sumber data yang digunakan pada *Text Mining* adalah kumpulan teks yang memiliki format tidak terstruktur atau semi terstruktur. Adapun tugas khusus *Text Mining* yaitu pengkategorian teks (*text categorization*) dan pengelompokan teks (*text clustering*).

2.2.4. Sentimen Analisis

Sentimen Analisis merupakan sebuah cabang penelitian di *domain Text Mining*. Proses yang dilakukan yaitu mengekstrak dan mengolah data *text* secara otomatis untuk mendapatkan informasi sentimen yang terkandung dalam satu

kalimat opini. Sentimen yang didapatkan dari sebuah masalah atau topik tertentu dengan hasil positif, negatif atau netral (Falahah, 2015).

2.2.5. Naive Bayes Classifier

Menurut Olson Delen (2008) *Naive Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar. Algoritma ini mengasumsikan bahwa atribut obyek adalah *independen*. *Probabilitas* yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari “*master*” tabel keputusan.

Kaitan antara *Naive 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).

Pada *Naive Bayes* proses untuk pelatihan (*training*) menggunakan persamaan (Budi Santoso, 2007) :

$$P(W_k|V_j) = \frac{n_k+1}{n+|vocabulary|} \dots\dots\dots(1)$$

Keterangan :

$P(W_k|V_j)$ = Peluang kategori j ketika terdapat kemunculan kata i

n_k+1 = Frekuensi kemunculan kata pada sebuah katagori

n = Jumlah seluruh kata pada dokumen dalam suatu kategori

$|vocabulary|$ = Jumlah semua kata pada semua kategori

Kemudian menghitung probabilitas kategori dengan menggunakan persamaan (Budi Santoso, 2007) :

$$P(V_j) = \frac{|doc\ j|}{|contoh|} \dots\dots\dots(2)$$

Keterangan :

$P(V_j)$ = Probabilitas dokumen kategori

$|doc\ j|$ = Jumlah dokumen setiap kategori j

$|contoh|$ = Jumlah dokumen dari semua kategori

Selanjutnya untuk proses klasifikasi, data yang diinputkan merupakan data yang belum memiliki kategori. Proses klasifikasi data menggunakan persamaan (Budi Santoso, 2007) :

$$V_{map} = \arg \max_{V_j \in V} P(V_j) \prod_{i=1}^n P(x_i|V_j) \dots\dots\dots(3)$$

Keterangan :

V = Kategori tweet $j = 1,2,3,\dots,n$ Dimana dalam penelitian ini

j_1 = sentimen negatif,

j_2 = sentimen positif,

j_3 = sentimen netral.

$P(x_i|V_j)$ = Probabilitas x_i pada kategori V_j

$P(V_j)$ = Probabilitas dari V_j

Python mempunyai *library Naïve Bayes Classifier* dengan nama *nltk.classify (Natural Language Toolkit)* yang di kembangkan oleh Steven Bird dan Edward Loper dalam kaitannya dengan mata kuliah komputasi *linguistic* di Universitas Pennsylvania pada tahun 2001.