BAB II

TINJAUAN PUSTAKA DAN DASAR TEORI

2.1 Tinjauan Pustaka

Dalam penelitian ini digunakan beberapa sumber pustaka. Pustaka yang relevan pada penelitian ini ditinjau dari sisi kasus penelitian dan metode yang digunakan. Kasus penelitian yang dilakukan adalah mengenai Analisis Sentimen Berita Kemiskinan. Metode yang digunakan dalam penelitian ini adalah metode *Naive Bayes Classifier.* Penelitian yang sudah dilakukan dapat dilihat pada tabel 2.1

1. Nazuar Adnan (2017) yang berjudul 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.
2. Yuna Sophia Dewi Febriant (2017) yang berjudul Analisis dan klasifikasi sentiment terhadap twitter STMIK AKAKOM Yogyakarta.
3. Rizky Maulana (2016) yang berjudul Analisis sentimen positif dan negatif pengguna twitter terhadap tokoh public secara online.
4. Ahmad Fathan Hidayatullah (2014) yang berjudul Klaifikasi tweet berdasarkan sentimen dan kategori yang berasal dari fitur yang dimiliki oleh tokoh publik.

Tabel 2.1 Tinjauan pustaka yang digunakan pada penelitian ini :

Peneliti	**Tahun**	**Data Penelitian**	**Metode**	**Hasil**
Nazuar Adnan	2017	Topik pada headline berita online	Naïve Bayes Classifier	Analisis Sentimen dan Klasifikasi topik pada Headline Berita Online

				Untuk mengklasifikasi apakah berita tersebut termasuk berita positif, negative, atau netral dan mengklasifikasi topik dari berita tersebut.
Yuna Sophia Dewi Febriant	2017	Twitter STMIK AKAKOM Yogyakarta	Naïve Bayes Classifier	Analisis dan klasifikasi sentiment terhadap twitter STMIK AKAKOM Yogyakarta
Rizky Maulana	2016	Twitter Tokoh Publik	Support Vector Machine	Analisis sentimen positif dan negatif pengguna twitter terhadap tokoh public secara online
Ahmad Fathan Hidayatullah	2014	Twitter Tokoh Publik	Support Vector Machine	Klaifikasi tweet berdasarkan sentimen dan kategori yang berasal dari fitur yang dimiliki oleh tokoh public
Penelitian yang dilakukan		Berita kemiskinan di Yogyakarta	Naïve Bayes Classifier	Analisis sentiment pada berita online tentang kemiskinan untuk mengklasifikasi apakah berita tersebut termasuk berita positif, negative, atau netral.

2.2 Dasar Teori

2.2.1 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 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 tertentu, dan membuang data pada frekuensi yang lain. Tahapan *Stemming* adalah proses pemetaandan 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.

2.2.2 *Sentiment Analisys*

Sentiment analysis 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 kecenderungan opini terhadap sebuah masalah atau objek oleh seseorang, apakah cenderung berpandangan atau beropini negatif atau positif (B. Liu. 2010).

Sentiment analysis adalah kegiatan melakukan analisis 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.

2.2.3 *Naïve Bayes Classifier*

Bayesian classification di dasarkan pada teorema bayes yang memiliki kemampuan hampir serupa dengan *decision tree* dan *neural network.* Teorema Bayes adalah teorema yang di gunakan 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 (ketidaktergantugan) yang kuat atau naif (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 tidk berkaitan dengan ada atau tidaknya fitur lain dalam data yang sama (Eko Prasetyo, 2012).

Bayesian classification adalah suatu metode pengklasifikasian data dengan model statik yang dapat di gunakan untuk menghitung probabilitas keanggotaan suatu kelas. Metode *Bayesian classification* digunakan menganalisis dalam membantu tercapainya pengambilan keputusan terbaik suatu permasalahan dari sejumlah alternatif. *Bayesian Clasification* 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 berita ini yang menjadi data uji adalah data headline. Ada dua tahap pada klasifikasi 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 dipresentasikan 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}). Dari rumus-rumus yang ada didapatlah rumus yang sudah disederhanakan yaitu sebagai berikut :

$$V_{MAP} = \underset{V\ j\ e\ V}{\arg max} \prod_{i=1}^{n} P(X_1 |Vj) P(Vj) \quad \text{……..……(2.1)}$$

Keterangan :

Vj = Kategori Headline

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

J1 = kategori headline sentiment positif

J2 = kategori headline sentiment negatif

J1 = 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:

$$P(Vj) = \frac{|docs\ j|}{|contoh|} \quad \text{…..……..……(2.2)}$$

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

$$P(xi|Vj) = \frac{nk + 1}{nk + |kosakata|} \quad \text{.....(2.3)}$$

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 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.

2.2.4 *Framework CodeIgniter*(CI)

CodeIgniter adalah sebuah *web application network* yang bersifat open source yang digunakan untuk membangun aplikasi php dinamis.

CodeIgniter menjadi sebuah *framework* PHP dengan model MVC *(Model, View, Controller)* untuk membangun website dinamis dengan menggunakan PHP yang dapat mempercepat pengembang untuk membuat sebuah aplikasi web. Selain ringan dan cepat, *CodeIgniter* juga memiliki dokumentasi yang super lengkap disertai dengan contoh implementasi kodenya. Dokumentasi yang lengkap inilah yang menjadi salah satu alasan kuat mengapa banyak orang memilih *CodeIgniter* sebagai framework pilihannya. Karena kelebihan-kelebihan yang dimiliki oleh *CodeIgniter*, pembuat PHP Rasmus Lerdorf memuji *CodeIgniter* di frOSCon (Agustus 2008) dengan mengatakan bahwa dia menyukai *CodeIgniter* karena *"it is faster, lighter and the least like a framework."*

CodeIgniter pertamakali dikembangkan pada tahun 2006 oleh Rick Ellis. Dengan logo api yang menyala, *CodeIgniter* dengan cepat "membakar" semangat para web developer untuk mengembangkan web dinamis dengan cepat dan mudah menggunakan framework PHP yang satu ini.