

## BAB II

### TINJAUAN PUSTAKA DAN LANDASAN TEORI

#### 2.1 Tinjauan Pustaka

Pada jurnal yang berjudul "Penerapan *Algoritma Naïve Bayes Classifier* untuk Mengklasifikasi Data Nasabah Asuransi" oleh (Bustami, 2014) mengatakan bahwa menggunakan *algoritma naïve bayes* untuk mengklasifikasi data nasabah asuransi mempunyai tingkat akurasi hasil yang lebih dari 80%. Pada jurnal ini penulis mencoba mengklasifikasikan sebuah data nasabah dengan status lancar, kurang lancar, dan tidak lancarnya nasabah dalam membayar premi asuransi. Data nasabah yang digunakan dalam penentuan klasifikasi ini mempunyai 7 kriteria yaitu jenis kelamin, usia, status, pekerjaan, masa pembayaran asuransi dan cara pembayaran asuransi. Dari 7 kriteria yang digunakan untuk menentukan hasil klasifikasi jenis data dari semua kriteria diantaranya merupakan data kategori.

Pada penelitian sebelumnya yang dilakukan oleh (Selvia, 2014) yang berjudul "Penerapan Metode Naive Bayes Pada Sistem Perpustakaan" metode Naive Bayes digunakan karena dalam pengolahan metode informasinya semakin mendalam dikarenakan banyaknya kategori yang dibutuhkan. Penulis pada jurnal ini mencoba memprediksi judul buku yang dicari, dan kategori dari buku tersebut dan ringkasan dari buku tersebut dalam sebuah perpustakaan. Pada penerapan *algoritma naïve bayes classifier* di jurnal ini tidak menggunakan banyak atribut untruk memprediksikan judul

buku, kategori dan ringkasan serta data yang digunakan melainkan data kategori.

Pada penelitian oleh (Effendi, 2015) yang berjudul "Rancangan Bangun Sistem Informasi Klasifikasi Status Gunung Berapi Menggunakan Metode Naïve Bayes Classifier". Pada Sistem informasi ini memiliki fitur yang dapat mengelola data status gunung berapi menggunakan metode algoritma Naïve bayes Classifier. Algoritma Naive Bayes Classifier dipilih karena dengan metode tersebut dapat mengolah bentuk data kategori dan numerik yang bersifat kontinu serta pada algoritma ini menggunakan konsep menghitung peluang terjadinya suatu kondisi dari setiap status gunung berapi yang ada sehingga setiap status dari gunung berapi dihitung jumlah peluangnya dan dicari nilai yang terbesar dari peluang tersebut hingga didapatkan hasil klasifikasi status dari gunung berapi. Pada hasil sistem informasi mempunyai tingkat akurasi 83,1% pada 9 inputan kriteria, yang menghasilkan klasifikasi status dari gunung berapi dengan cara membaca pola dengan konsep data mining menggunakan metode Naïve Bayes Classifier. Semakin banyak data training yang diinputkan maka semakin besar tingkat akurasi data yang didapatkan yang akan mempengaruhi hasil klasifikasi status gunung berapi.

Pada penelitian oleh (Nurrin & Paulus, 2016) yang berjudul "Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen di Twitter" yang menganalisis sentimen dari data *Twitter* dengan menggunakan 3 metode, antara lain *Naïve Bayes Classifier*, *Support*

*Vector Machine*, dan *Maximum Entropy*. Penelitian ini dilakukan dengan mengumpulkan data sejumlah 689 artikel terhadap tiga partai yang bersaing di pemilu India 2009 yaitu United Progressive Alliance (UPA), Telugu Desam Party (TPD), dan Telangana Rashtra Samithi (TRS). Data yang siap diolah dikumpulkan menggunakan Bahasa Inggris. Hasil penelitian ini menunjukkan bahwa secara umum Support Vector Machine memiliki skor parameter yang lebih baik dibandingkan dengan Naive Bayes Classifier, walaupun dalam beberapa perhitungan Naive Bayes Classifier tampak lebih unggul. Hasil akhir akurasi pada partai (UPA) untuk SVM 71.2%, NBC 65.2% dan ME 74.2%. Pada partai (TPD) untuk SVM 70%, untuk NBC 72.5% dan ME 60.1%. Pada partai (TRS) untuk SVM 66.6%, NBC 63.6%, dan ME 60.1%.

Pada penelitian oleh (Wayan & Saraswati, 2013) yang berjudul "*Naive Bayes Classifier dan Support Vector Machine Untuk Sentimen Analisis*" yang membahas tentang klasifikasi opini sebagai opini positif dan opini negative pada data berbahasa Indonesia dan data berbahasa Inggris yang menggunakan metode *Naive Bayes Classifier* dan *Support Vector Machine*. Baik metode NBC maupun metode SVM memberikan unjuk kerja yang baik dalam sentiment analysis pengklasifikasian opini berbahasa Inggris dan berbahasa Indonesia pada penelitian ini. Hasil percobaan menunjukkan bahwa metode SVM memberikan unjuk kerja yang lebih baik daripada metode NBC untuk mengklasifikasikan opini berbahasa Inggris dan opini positif berbahasa Indonesia. Sedangkan NBC memberikan unjuk kerja yang lebih baik dalam

mengklasifikasikan data uji opini negatif berbahasa Indonesia. Hasil rata-rata akurasi dari percobaan data berbahasa Inggris untuk NBC mencapai 78% dan SVM mencapai 88%, sedangkan berbahasa Indonesia untuk NBC mencapai 74% dan SVM mencapai 75%. metode SVM memberikan unjuk kerja yang lebih baik daripada metode NBC untuk mengklasifikasikan opini berbahasa Inggris dan opini berbahasa Indonesia. Perbedaan akurasi tidak terlalu besar yang menunjukkan NBC masih merupakan metode yang baik dalam pengklasifikasian teks. NBC dapat menjadi pilihan mengingat sifat komputasinya yang lebih ringan.

Pada penelitian selanjutnya oleh (Buntoro, 2017) yang berjudul "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter" yang menggunakan data sebanyak 300 tweet, berbahasa Indonesia dan menggunakan metode Lexicon Based dengan tiga class atribut yaitu positif, negatif, dan netral. Hasil dari proses klasifikasi diuji dengan menggunakan metode 10 fold cross validation dan menghasilkan nilai akurasi tertinggi didapat saat menggunakan metode klasifikasi Naïve 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%. Dalam penelitian ini juga dapat diketahui metode klasifikasi Naïve Bayes Classifier (NBC) lebih tinggi akurasinya untuk klasifikasi sentimen Tweet Bahasa Indonesia dibandingkan dengan metode klasifikasi Support Vector Machine (SVM).

Twitter merupakan media sosial yang di buat oleh Jack Dorsey pada tahun 2006. Pada Tahun 2013 berdasarkan *press-release* Twitter memiliki 500 juta *tweet* per harinya. Sebanyak 500 juta *tweet* akan percuma bila tidak dimanfaatkan karena terdapat berbagai macam opini masyarakat yang dapat diolah untuk mengklasifikasikan kedalam analisis sentimen menjadi opini positif, netral, dan negatif. Dengan menggunakan algoritma pengklasifikasian maka opini-opini tersebut dapat terklasifikasi. Pada penelitian yang dilakukan oleh (Pamungkas, Setiyanto, & Dolphina, 2015) hasil eksperimen untuk kategorisasi teks berbahasa Indonesia didapatkan bahwa metode *Naïve Bayes Classifier* memiliki akurasi 91%. Twitter merupakan sebuah indikator yang baik untuk memberikan pengaruh dalam penelitian. Namun masih belum banyak aplikasi dan metode analisa sentimen yang dikembangkan untuk bahasa Indonesia. Faktor-faktor keuntungan tersebut mendorong perlunya dilakukan penelitian analisis sentimen terhadap dokumen berbahasa Indonesia. Penelitian analisis sentimen ini dilakukan untuk mengetahui sentimen publik mengenai sesuatu dengan menggunakan pendekatan dalam machine learning yang dikenal dengan *Naïve Bayes Classifier* yang dikhususkan pada dokumentasi teks berbahasa Indonesia dengan fitur unigram. *Naive Bayes Classifier* dapat ditingkatkan untuk mencocokkan akurasi klasifikasi model yang lebih rumit untuk analisis sentimen dengan memilih jenis yang tepat dari fitur dan menghilangkan noise dengan pemilihan fitur yang sesuai. *Naive Bayes Classifier* dipilih karena mereka

sangat cepat untuk melatih dan dapat digunakan dengan dataset yang lebih besar. Mereka juga kuat terhadap gangguan dan kurang rentan terhadap overfitting. Kemudahan implementasi juga keuntungan besar dari *Naive Bayes Classifier*.

Pada penelitian yang dilakukan oleh (Pratama, & Sunengsih 2010) penelitian ini menggunakan layanan jejaring sosial Twitter sebagai sumber data teks berupa tweet selama bulan September 2017. Data tersebut berasal dari pengguna sosial media Twitter yang mengunggah opini tentang perusahaan Apple dan Samsung Mobile. Tujuan dari penelitian ini adalah melihat sentimen pengguna layanan jejaring sosial Twitter terhadap perusahaan Apple dan Samsung Mobile selama bulan September 2017, kemudian membandingkan hasil analisis sentimen tersebut dengan kondisi harga saham perusahaan Apple dan Samsung Mobile selama bulan September 2017. Hasil dari penelitian ini yaitu sentimen pengguna layanan jejaring sosial Twitter pada perusahaan Apple dan Samsung Mobile didominasi oleh sentimen positif yaitu 6530 *tweet* positif dalam 10000 *tweet* pada @Apple dan 7498 *tweet* dalam 10000 *tweet* pada @SamsungMobile.

Tabel Tinjauan Pustaka

Tabel 2. 1 Algoritma Naïve Bayes mengelola data kategori

Item Pembanding	Bustami, 2014	Selvia, 2014	Effendi, 2015
<b>Judul Penelitian</b>	Penerapan <i>Algoritma Naïve Bayes Classifier</i> untuk Mengklasifikasi Data Nasabah Asuransi	Penerapan Metode <i>Naïve Bayes Classifier</i> Pada Sistem Perpustakaan	Rancangan Bangun Sistem Informasi Klasifikasi Status Gunung Berapi Menggunakan Metode <i>Naïve Bayes Classifier</i>
<b>Tujuan</b>	Merancang perangkat lunak dalam menampilkan informasi status kelancaran dalam pembayaran asuransi nasabah menggunakan metode <i>Naïve Bayes Classifier</i>	Merancang perangkat lunak pencarian buku perpustakaan menggunakan metode <i>Naïve Bayes Classifier</i>	Mengetahui dan menganalisa data atribut klasifikasi status gunung berapi yang ada di Indonesia dengan menggunakan metode <i>Naïve Bayes Classifier</i>
<b>Metode</b>	<i>Naïve Bayes Classifier</i>	<i>Naïve Bayes Classifier</i>	<i>Naïve Bayes Classifier</i>

<p><b>Hasil</b></p>	<p>Sistem klasifikasi data nasabah ini digunakan untuk menampilkan informasi klasifikasi lancar, kurang lancar dan tidak lancar. Data nasabah yang digunakan melakukan penentuan klasifikasi yang terbagi menjadi 7 kriteria. Kriteria tersebut digunakan untuk menentukan hasil klasifikasi jenis data dari kriteria menggunakan data kategori. Data kategori diklasifikasi menggunakan metode <i>Naïve Bayes Classifier</i> menghasilkan nilai akurasi hingga lebih dari 80%</p>	<p>Pada sistem perpustakaan yang dibuat membutuhkan banyaknya kategori untuk memprediksi judul buku yang dicari. Penggunaan metode NBC dapat diimplementasikan pada pencarian perangkat lunak ini, karena berhasil mencari deskripsi yang berkaitan dengan kata yang dicari oleh user. Dengan menggunakan metode ini dapat melakukan pengklasifikasian dokumen dengan baik</p>	<p>Sistem Informasi klasifikasi status gunung berapi ini mampu memberikan hasil klasifikasi status dari gunung berapi dengan cara membaca pola data dengan menggunakan metode <i>Naive Bayes Classifier</i>. Klasifikasi status terbagi menjadi 4 kategori yaitu normal, waspada, siaga, dan was. Pada 9 inputan kriteria, sistem informasi dapat melakukan pengklasifikasian menurut status gunung berapi menggunakan metode NBC dengan akurasi 83.1%</p>
---------------------	--	--	--

**Tabel 2. 2 Perbandingan Algoritma Naïve Bayes dengan algoritma lain**

<b>Item Pembanding</b>	Nurrun, 2016	Saraswati, 2013	Buntoro, 2017
<b>Judul Penelitian</b>	Studi Literatur Tentang Perbandingan Metode Untuk Proses Analisis Sentimen di Twitter	<i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i> untuk Sentimen Analisis	Analisis Sentimen Calon Gubernur DKI Jakarta 2017 di Twitter
<b>Tujuan</b>	Mengetahui dan menganalisa metode yang terbaik dalam melakukan analisis sentimen dengan membandingkan metode <i>Naïve Bayes Classifier</i> , <i>Support Vector Machine</i> , dan <i>Maximum Entropy</i>	Mengetahui dan menganalisa metode yang terbaik untuk sentimen analisis dengan membandingkan metode <i>Naïve Bayes Classifier</i> dan <i>Support Vector Machine</i>	Mengetahui dan menganalisa metode yang terbaik antara <i>Naïve Bayes</i> dengan <i>Support Vector Machine</i> dalam memprediksi calon gubernur DKI Jakarta 2017 melalui sentimen analisis pada media sosial Twitter

<b>Metode</b>	<i>Naïve Bayes Classifier, Support Vector Machine, dan Maximum Entropy</i>	<i>Naïve Bayes Classifier dan Support Vector Machine</i>	<i>Naïve Bayes Classifier dan Support Vector Machine</i>
<b>Hasil</b>	<p>Hasil penelitian ini menunjukkan bahwa metode Support Vector Machine memiliki skor parameter yang lebih baik dibandingkan dengan Naive Bayes Classifier, walaupun dalam beberapa perhitungan Naive Bayes Classifier tampak lebih unggul. Terbukti dari hasil akurasi partai TDP dengan akurasi SVM 70% dan NBC 72,5%</p>	<p>Hasil penelitian ini menunjukkan bahwa metode SVM memberikan unjuk kerja yang lebih baik daripada metode NBC untuk mengklasifikasikan opini berbahasa Inggris dan opini berbahasa Indonesia. Perbedaan akurasi tidak terlalu besar yang menunjukkan NBC dapat menjadi pilihan mengingat sifat komputasinya yang lebih ringan.</p>	<p>Dari hasil pengujian didapatkan bahwa data melalui proses 10 fold cross validation menghasilkan data AHY memiliki nilai akurasi tertinggi pada metode Naïve Bayes Classifier dengan akurasi mencapai 95% dibandingkan dengan Support Vector Machine 90%.</p>

**Tabel 2. 3 Media sosial Twitter untuk analisis sentimen**

<b>Item Pembanding</b>	Pamungkas, 2015	Pratama, 2010
<b>Judul Penelitian</b>	Analisis Sentimen Pada Sosial Media Twitter Menggunakan Naïve Bayes Classifier Terhadap Kata Kunci "Kurikulum 2013"	Analisis Sentimen dengan Metode Naïve Bayes Pada Opini Pengguna Media Sosial Twitter
<b>Tujuan</b>	Mengetahui dan menganalisa sentimen dari opini masyarakat terhadap kata kunci "Kurikulum 2013" dengan menggunakan algoritma <i>Naïve Bayes Classifier</i> .	Mengetahui dan menganalisa sentimen dari opini pengguna media sosial Twitter menggunakan algoritma <i>Naïve Bayes Classifier</i>
<b>Metode</b>	<i>Naïve Bayes Classifier</i>	<i>Naïve Bayes Classifier</i>
<b>Hasil</b>	Pada hasil eksperimen untuk kategorisasi teks berbahasa Indonesia didapatkan bahwa metode <i>Naïve Bayes Classifier</i> memiliki akurasi 91% untuk	Pada penelitian ini menggunakan layanan jejaring sosial Twitter sebagai sumber data teks berupa tweet selama bulan September 2017. Data

data tweet sejumlah 3300 tweet dengan data training sejumlah 3000 dan data latih sejumlah 100. Klasifikasi tweet bersentimen lebih akurat jika data latih yang digunakan semakin banyak dalam data pengetahuan.

tersebut berasal dari pengguna sosial media Twitter yang mengunggah opini tentang perusahaan Apple dan Samsung Mobile. Hasil dari penelitian yaitu membandingkan hasil analisis sentimen tersebut dengan kondisi harga saham perusahaan Apple dan Samsung Mobile selama bulan September 2017. Hasil dari penelitian ini yaitu sentimen pengguna layanan jejaring sosial Twitter pada perusahaan Apple dan Samsung Mobile didominasi oleh sentimen positif yaitu 6530 tweet positif dalam 10000 tweet pada @Apple dan 7498 tweet dalam 10000 tweet pada @SamsungMobile.

## **2.2 Landasan Teori**

### **2.2.1 Penanggulangan Bencana**

Penanggulangan bencana merupakan segala upaya kegiatan yang dilakukan meliputi kegiatan pencegahan, penjinakan (mitigasi), penyelamatan, rehabilitasi, dan rekonstruksi baik sebelum, pada saat dan setelah bencana terjadi maupun menghindari dari bencana yang terjadi. Pemerintah dan pemerintah daerah bertanggung jawab dalam penyelenggaraan penanggulangan bencana. Sebagaimana didefinisikan dalam UU 24 Tahun 2007 tentang Penanggulangan Bencana, penyelenggaraan Penanggulangan Bencana adalah serangkaian upaya yang meliputi penetapan kebijakan pembangunan yang berisiko timbulnya bencana, kegiatan pencegahan bencana, tanggap darurat, dan rehabilitasi (BADAN NASIONAL PENANGGULANGAN BENCANA, 2008).

### **2.2.2 Pencegahan Bencana**

Pencegahan bencana adalah serangkaian kegiatan yang dilakukan untuk mengurangi atau menghilangkan risiko bencana, baik melalui pengurangan ancaman bencana maupun kerentanan pihak yang terancam bencana.

### **2.2.3 Kesiapsiagaan (*Preparedness*)**

Kesiapsiagaan adalah serangkaian yang dilakukan untuk mengantisipasi bencana melalui pengorganisasian serta melalui langkah yang tepat guna dan berdaya guna.

#### **2.2.4 Mitigasi (*Mitigation*)**

Mitigasi adalah serangkaian upaya untuk mengurangi risiko bencana, baik melalui pembangunan fisik maupun penyadaran dan peningkatan kemampuan menghadapi ancaman bencana

#### **2.2.5 Tanggap Darurat (*Response*)**

Tanggap darurat bencana adalah serangkaian kegiatan yang dilakukan dengan segera pada saat kejadian bencana untuk menangani dampak buruk yang ditimbulkan, yang meliputi kegiatan penyelamatan dan evakuasi korban, harta benda, pemenuhan kebutuhan dasar, perlindungan, pengurusan pengungsi, penyelamatan, serta pemulihan prasarana dan sarana.

#### **2.2.6 Pemulihan (*Recovery*)**

Pemulihan adalah suatu proses yang dilalui agar kebutuhan pokok terpenuhi. Proses Pemulihan terdiri dari rehabilitasi dan rekonstruksi.

#### **2.2.7 Rehabilitasi**

Rehabilitasi adalah perbaikan dan pemulihan semua aspek pelayanan publik atau masyarakat sampai tingkat yang memadai pada wilayah pascabencana dengan sasaran utama untuk normalisasi atau berjalannya secara wajar semua aspek pemerintahan dan kehidupan masyarakat pada wilayah pascabencana.

### **2.2.8 Rekonstruksi**

Rekonstruksi adalah pembangunan kembali semua prasarana dan sarana, kelembagaan pada wilayah pascabencana, baik pada tingkat pemerintahan maupun masyarakat dengan sasaran utama tumbuh dan berkembangnya kegiatan perekonomian, sosial dan budaya, tegaknya hukum dan ketertiban, dan bangkitnya peran serta masyarakat dalam segala aspek kehidupan bermasyarakat pada wilayah pascabencana.

### **2.2.9 Kebutuhan Khusus**

Kebutuhan khusus merupakan kebutuhan yang wajib dimiliki oleh korban bencana yang tempat tinggalnya terkena dampak akibat bencana alam yang terjadi yang diberikan oleh pemerintah dalam upaya membantu untuk memenuhi kebutuhan yang dipenuhi serta melindungi ke daerah yang lebih aman untuk ditinggali. Kebutuhan khusus ini terbagi menjadi 4 kebutuhan. Kebutuhan Konseling, Kebutuhan Psikologi, Kebutuhan Khusus Gen, dan Kebutuhan Fasilitas.

### **2.2.10 Kebutuhan Konseling**

Kebutuhan Konseling merupakan kebutuhan yang terkait dengan pendidikan. Kebutuhan konseling sangat dibutuhkan bagi para korban bencana yang memiliki kewajibannya dalam menuntut ilmu. Akibat dari dampak bencana Merapi maka para korban tidak bisa melanjutkan sekolah karena harus dievakuasikan ke daerah yang lebih aman. Didalam kebutuhan konseling terbagi menjadi beberapa kategori

yaitu kategori pendidikan, pelatihan, pengetahuan, dan penyuluhan. Dari keempat kategori tersebut tidak hanya mengajarkan pembelajaran sekolah namun mengajarkan kepada para korban untuk mendapatkan informasi atau menambahkan wawasan mereka dalam melakukan penanggulangan bencana agar warga bisa segera melakukan penyelamatan diri sendiri sewaktu terjadi bencana.

#### **2.2.11 Kebutuhan Psikologi**

Akibat dampak bencana Gunung Merapi meletus maka para korban akan mengalami gangguan psikologi. Kebutuhan Psikologi inilah yang merupakan kebutuhan yang terkait dengan penanganan korban bencana untuk meneguhkan atau menguatkan hati para korban agar kembali dengan normal. Kebutuhan Psikologi terbagi menjadi beberapa kategori yaitu kategori psikolog, terapi, hiburan, dan pemulihan. Dari keempat kategori tersebut mengajarkan untuk menghibur para korban, menguatkan perasaan dan membesarkan hatinya dengan cara memberikan permainan, melakukan terapi olahraga, dan memberikan perawatan kesehatan mental khusus untuk menjauhkan dari trauma, depresi stress agar jiwanya pulih seperti semula.

#### **2.2.12 Kebutuhan Khusus Kategori**

Kebutuhan khusus kategori merupakan kebutuhan yang diperlukan oleh korban dalam memenuhi kebutuhan pokok sehari-hari. Kebutuhan khusus gen terbagi menjadi beberapa kategori yaitu kebutuhan balita/bayi, kebutuhan anak-anak, kebutuhan perempuan, dan kebutuhan lansia.

Dari keempat kategori tersebut memberikan bantuan seperti makanan asupan gizi, tempat tinggal, pakaian, obat-obatan, alat mandi dan alat tidur.

#### **2.2.13 Kebutuhan Fasilitas**

Kebutuhan Fasilitas merupakan kebutuhan yang terkait dengan fasilitas yang dibutuhkan dalam penanggulangan bencana. Kebutuhan fasilitas terbagi menjadi beberapa kategori seperti toilet, akses, ruangpengungsi, dan fasilitas. Dari keempat kategori tersebut memberikan informasi terkait dengan akses jalan dalam melakukan evakuasi korban bencana, pembangunan toilet untuk lansia dengan jarak yang dekat dengan posko, mck dengan fasilitas air bersih layak digunakan, ruang khusus seperti ruang ibu menyusui, dan fasilitas yang dibutuhkan khusus untuk para korban yang membutuhkan.

#### **2.2.14 Data Mining**

Data mining adalah serangkaian proses untuk menggali nilai tambah berupa informasi dari suatu data dengan melakukan penggalian pola-pola untuk menemukan pola menarik dari sejumlah data yang besar dengan tujuan untuk memanipulasi data menjadi informasi yang lebih berisi atau berharga, dimana data dapat tersimpan di dalam database, warehouse dan repositori informasi lainnya. Bidang lain yang berkontribusi dengan data mining adalah saraf jaringan, pengenalan pola, dan yang lainnya (Gorunescu, 2011). Dalam data mining terdapat beberapa metode yang bisa digunakan. Metode tersebut

digunakan untuk menemukan pengetahuan yang nantinya digunakan. Metode yang ada seperti classification, clustering, regression, dependency modeling, deviation change detection, dan summarization.

#### **2.2.15 Sentimen Analisis**

Sentimen analisis 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 (Rozi, Hadi, & Achmad, 2012).

#### **2.2.16 Metode Naïve Bayes Classifier**

*Naïve Bayes* merupakan sebuah metode klasifikasi yang berakar pada Teorema Bayes. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yang dikemukakan oleh ilmuwan Inggris Thomas Bayes. Ciri utama dari *Naïve Bayes Classifier* adalah asumsi yang sangat kuat akan independensi dari masing-masing kondisi atau kejadian. *Naïve Bayes Algorithm* sangat berguna untuk pembelajaran algoritma, pengetahuan masa lalu dan pengamatan data yang bisa digabungkan, ini dapat membantu dalam memberikan perspektif yang bermanfaat untuk memahami dan juga mengevaluasi banyak pembelajaran. Ini membantu menentukan probabilitas yang

tepat untuk hipotesis dan juga kuat terhadap noise pada data input (Parveen & Pandey, 2017). *Naïve Bayes Classifier* menggunakan data yang sudah ada yang bernama data training yang digunakan untuk membentuk sebuah model classifier. Model ini merupakan representasi pengetahuan yang akan digunakan untuk prediksi kelas data baru yang belum pernah ada. Kemudian membutuhkan data latih yang bernama data testing yang digunakan untuk mengukur sejauh mana classifier berhasil melakukan klasifikasi dengan benar. Kemudian rumus dasar dari *Naïve Bayes* yang digunakan, yaitu :

$$P(H_i | X) = \frac{P(X|H_i) P(H_i)}{P(X)}$$

Keterangan :

X : Kriteria suatu kasus berdasarkan masukan.

H<sub>i</sub>: Kelas solusi pola ke-i, dimana i adalah jumlah label kelas.

P(H<sub>i</sub>|X) : Probabilitas kemunculan label kelas H<sub>i</sub> dengan kriteria masukan X.

P(X|H<sub>i</sub>) : Probabilitas kriteria masukan X dengan label kelas H<sub>i</sub>.

P(H<sub>i</sub>) : Probabilitas label kelas H<sub>i</sub>.

### **2.2.17 Pre-proccesing**

*Pre-proccesing* adalah suatu proses perubahan bentuk data yang belum terstruktur menjadi data yang terstruktur sesuai dengan kebutuhan, untuk proses mining yang lebih lanjut seperti sentimen analysis,

peringkasan, clustering dokumen, dan lainnya (Sarjana, Statistika, Matematika, Ilmu, & Alam, 2017). adapun tahapan dalam proses *pre-proccesing* teks adalah sebagai berikut :

a. *Case Folding* : proses dimana mengubah semua karakter pada teks menjadi huruf kecil dan menghilangkan angka atau bentuk tanda baca sehingga data yang didapat hanya mengandung karakter huruf a sampai z.

b. *Cleansing* : proses membersihkan *tweet* dari kata yang tidak diperlukan atau untuk mengurangi *noise*. Kata yang dihilangkan adalah karakter HTML, *emoticon*, *hashtag* (#), *username* dan *url*.

c. *Tokenizing* : proses memecah kalimat menjadi potongan kata yang setelah itu diberikan bobot masing-masing.

d. *Stopwords* : proses menghilangkan kata-kata yang tidak berguna yang dimana dapat menghambat proses analisis, seperti "dan", "oleh", "pada" dan sebagainya.

e. *Stemming* : proses menyederhanakan kata yang berisi imbuhan, awalan, sisipan, akhiran dan kombinasi kembali ke kata dasarnya.

### **2.2.18 WEKA**

WEKA (Waikato Environment for Knowledge Analysis) adalah sebuah perangkat lunak yang menyediakan implementasi algoritma pembelajaran yang dapat diterapkan dengan mudah pada dataset. WEKA memiliki beberapa tools untuk masalah pengelolaan data mining, mulai dari *regression*, *pre-processing*, *classification*, *clustering*, *association rules*, dan *visualization*. WEKA merupakan perangkat lunak yang dibangun menggunakan Bahasa pemrograman Java yang kemudian didistribusikan menjadi perangkat lunak *open source* (Frank, Hall, & Witten, 2016).

### **2.2.19 K-fold Cross Validation**

K-Fold Cross Validation merupakan salah satu metode dalam menentukan data training dan data testing dari keseluruhan data. K-fold cross validation mengulang k-kali untuk membagi sebuah himpunan data secara acak menjadi k-subset yang saling bebas, setiap ulangan disisakan satu subset untuk pengujian dan subset lainnya untuk pelatihan. Nilai K yang disarankan adalah 10 karena lebih akurat dalam ukuran estimasinya (Zega, 2014).

K-fold cross validation juga disebut 10 k-fold cross validation, karena k mengambil nilai 10 yang telah menjadi praktek paling umum. Bahkan, studi empiris menunjukkan bahwa sepuluh tampaknya menjadi jumlah yang optimal dari fold, karena bias dan varians relatif rendah. Evaluasi performansi dilakukan untuk menguji

hasil dari klasifikasi dengan mengukur nilai performansi dari sistem yang telah dibuat. Parameter pengujian yang digunakan untuk evaluasi yaitu akurasi. Metode pengukuran yang digunakan adalah confusion matrix (classification matrix or a contingency table) (Zuhri, Alamsyah, Si, & Sc, 2017).

### 2.2.20 Confusion Matrix

Confusion matrix merupakan suatu metode yang digunakan untuk melakukan perhitungan akurasi pada konsep data mining (Rosandy, 2016). Confusion matrix berisikan informasi mengenai hasil klasifikasi aktual dan telah di prediksi oleh sistem klasifikasi. Performa dari sistem tersebut biasanya dievaluasi menggunakan data dalam sebuah matrix. Dibawah ini menampilkan sebuah confusion matrix untuk pengklasifikasian ke dalam dua kelas.

**Tabel 2. 4 Visualisasi Confusion Matrix**

<b>Kelas</b>	<b>Terklasifikasi Positif</b>	<b>Terklasifikasi Negatif</b>
Positif	TP (True Positive)	FN (False Negatif)
Negatif	FP (False Positive)	TN (True Negatif)

Keterangan :

1. Nilai True Negatif (TN) merupakan jumlah data negatif yang terdeteksi dengan benar.

2. False Positive (FP) merupakan data negatif namun terdeteksi sebagai data positif.
3. True Positive (TP) merupakan data positif yang terdeteksi benar.
4. False Negatif (FN) merupakan kebalikan dari True Positive, sehingga data positif, namun terdeteksi sebagai data negatif.

Berdasarkan nilai True Negatif (TN), False Positive (FP), False Negatif (FN), dan True Positive (TP) dapat diperoleh nilai akurasi, presisi, recall dan f-measure.

a. Presisi

Presisi adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. Dengan rumus sebagai berikut :

$$\text{Presisi} = \frac{TP}{TP+FP}$$

b. Recall

Recall adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. Dengan rumus sebagai berikut :

$$\text{Recall} = \frac{TP}{TP+FN}$$

c. Akurasi

Akurasi adalah tingkat kedekatan antara nilai prediksi dengan nilai aktual. Dengan rumus sebagai berikut :

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN}$$

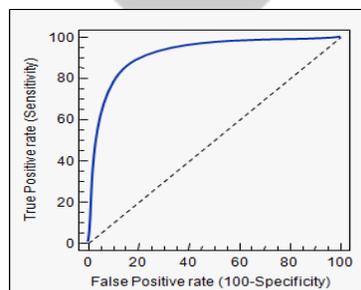
#### d. F-measure

F-measure merupakan salah satu perhitungan evaluasi dalam temu kembali informasi yang mengkombinasikan recall dan precision. Nilai recall dan precision pada suatu keadaan dapat memiliki bobot yang berbeda. Ukuran yang menampilkan timbal balik antara recall dan precision adalah F-measure yang merupakan bobot harmonic mean dari recall dan precision. Dengan rumus sebagai berikut :

$$F1 = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$$

Range dari nilai F-Measure adalah 0 sampai dengan 1.

Setelah mengetahui nilai akurasi, error rate, presisi dan recall, penulis menggunakan kurva ROC untuk memvisualisasikan akurasi metode yang digunakan. Kurva ini menggambarkan TPR pada sumbu y dan FPR pada sumbu x. Dengan menggunakan kurva ROC kita dapat mengetahui besar daerah *area under curve (AUC)*. Lalu nilai AUC akan digunakan untuk menguji performa model klasifikasi. AUC yang memiliki nilai 0.5 atau mendekati 0.5 menandakan metode yang digunakan mengklasifikasikan dengan tidak benar. AUC yang memiliki nilai 1.0 menandakan metode yang digunakan dapat mengklasifikasikan dengan benar.



**Gambar 2. 1 kurva Receiver Operating Characteristic**

### **2.2.21 SentiStrength**

Metode *SentiStrength* adalah suatu metode yang dapat mendapatkan skor kekuatan sentiment dan dapat mengelompokkan sentimen atas positif, netral, dan negatif sesuai dengan nilai-nilai yang terkandung dari tweet yang disampaikan (Wahid & SN, 2016). Metode ini mengukur nilai sentimen dari kamus-kamus yang sudah diberi bobot. Bobot yang diberi nilai -5 (memiliki sentimen negatif yang sangat kuat) sampai 5 (memiliki nilai sentimen positif yang sangat kuat). Kamus-kamus ini didapat dari hasil translasi kamus sentimen berbahasa inggris yang melibatkan 3 responden ahli, mahasiswa S2 Linguistik Universitas Gadjah Mada (UGM) untuk memberi nilai bobot pada sentimen (Wahid & SN, 2016). Kamus-kamus yang digunakan sebagai patokan dalam menentukan bobot sentimen adalah :

#### **a. Kamus Sentimen**

Pada kamus ini terdapat beberapa kata-kata yang memiliki nilai sentimen positif, negatif atau netral yang sudah diberi bobot. Contoh beberapa kata dalam kamus sentimen dapat dilihat pada tabel 2.5.

**Tabel 2. 5 Contoh kata dalam kamus sentimen**

Kata	Bobot
Adil	5
Anjing	-4
Beruntung	4
Gemes	3

b. Kamus Idiom / Kata Kiasan

Kamus idiom merupakan suatu kata ungkapan yang memiliki arti yang berbeda dari kata aslinya (Wahid & SN, 2016). Contoh kata dalam kamus idiom dapat dilihat pada tabel 2.6

**Tabel 2. 6 Contoh kata dalam kamus Idom**

Kata	Bobot
Main duitan	-4
Oke deh	1
Turun tangan	3
Selamat pagi	1

c. Boosterwords

Boosterwords adalah kata yang dapat meningkatkan atau mengurangi nilai makna sentimen dari suatu kata yang ada di sebelahnya (Wahid & SN, 2016). Contoh dari kata boosterwords dapat dilihat pada tabel 2.7

**Tabel 2. 7 Contoh kata dalam kamus Boosterword**

<b>Kata</b>	<b>Bobot</b>
Hanya	-1
Banget	1
Sempurna	1
Kurang	-2

d. Kata Negasi

Kata negasi adalah kata yang dapat mengubah nilai opini dari suatu kata. Misalnya kata "jahat" yang memiliki nilai sentimen negatif, didahului dengan kata negasi. Misalnya menjadi "Tidak jahat". Kata ini berubah menjadi sentimen yang mengandung nilai positif. Contoh kata negasi dapat dilihat pada tabel 2.8.

**Tabel 2. 8 Contoh beberapa kata negasi**

<b>Kata</b>	<b>Bobot</b>
Hanya	-1
Banget	1
Sempurna	1
Kurang	-2

e. Kata Tanya

Kata tanya dapat mengubah suatu makna nilai opini seseorang dalam menyampaikan tweet di Twitter. Contohnya "apakah kamu jahat?". Walaupun pada kalimat tersebut terdapat kata "jahat", tetapi kalimat tersebut tidak termasuk ke dalam positif atau negatif, melainkan masuk ke dalam kategori netral. Contoh beberapa kata tanya dapat dilihat pada tabel 2.9.

**Tabel 2. 9 Contoh beberapa kata tanya**

No	Kata
1	Mengapa
2	Berapa
3	Bagaimana
4	Apakah

Dengan menggunakan metode sentistrength maka hasil yang diperhitungkan dengan membandingkan nilai positif tertinggi di kurangkan dengan nilai negatif yang tertinggi.

Hasil positif jika nilai positif > nilai negatif maka sentimen positif.

Hasil negative jika nilai positif < nilai negatif maka sentimen negatif.

Hasil netral jika nilai positif = nilai negatif maka sentimen netral.