

## BAB 3. LANDASAN TEORI

Pada landasan teori penulis menjelaskan tentang teori-teori yang digunakan untuk mendukung penelitian ini. Teori tersebut akan menjadi panduan dalam analisis sentimen masyarakat terhadap Covid-19, penyelenggaraan vaksin Covid-19 dan vaksin Covid-19 di Indonesia melalui Twitter. Teori tersebut terdapat analisis sentimen, media sosial Twitter, Covid-19, vaksin Covid-19, data *mining*, *Multinomial Naïve Bayes*, *Support Vector Machine*, *SentiStrength*, *Confusion Matrix* dan *python3*.

### 3.1. Analisis Sentimen

*Opinion Mining* atau sering disebut dengan analisis sentimen dalam penelitian digunakan untuk mencari tahu pandangan masyarakat terhadap kasus Covid-19 yang terus meningkat setelah dilakukannya Vaksin. Dimana analisis sentimen merupakan proses dimana data-data berbentuk teks dipahami, diekstrak, kemudian diolah secara otomatis agar dapat memperoleh informasi sentimen yang terdapat pada sebuah kalimat yang mengandung opini [27]. Analisis ini dilakukan agar opini atau pendapat dari suatu masalah ataupun objek tertentu. Analisis ini akan melihat apakah opini tersebut cenderung beropini positif maupun opini negatif.

Analisis sentimen merupakan sebuah penelitian dimana hal yang diteliti adalah opini, sentiment, dan emosi yang dituangkan kedalam bentuk teks oleh masyarakat. Analisis sentiment memiliki tujuan untuk melakukan ekstrak pada atribut dan komponen yang berasal dari objek yang sudah dikomentari. Objek yang sudah dikomentari ini digunakan untuk menentukan apakah komentar yang ada memiliki makna positif atau negatif.

### 3.2. Media Sosial Twitter

Media sosial Twitter dalam penelitian ini digunakan penulis sebagai tempat untuk mencari data training analisis sentimen. Twitter sendiri didirikan sejak tahun 2006 oleh seseorang yang bernama Jack Dorsey dan kini telah memperoleh banyak

peminat dari berbagai penjuru yang ada di dunia. Twitter pada saat ini telah mempunyai ratusan juta users. Twitter dapat dikatakan sebuah “SMS dari Internet”. Logo dari twitter adalah seekor burung yang memiliki warna biru di tubuhnya yang diberi nama “*Larry the Bird*”.

Twitter adalah salah satu media sosial yang berguna sebagai alat komunikasi yang akan membagikan berbagai informasi mengenai banyak hal. Tweets adalah tulisan berupa teks pada suatu halaman yang mengandung 140 karakter pada bagian profil dari pengguna. Tweets dapat dilihat oleh semua orang, tetapi orang yang mengirim tweets bisa memberi batas agar tweets yang dikirimkan hanya kepada friendlist dari pengirim itu sendiri. Twitter sebagai media sharing memiliki potensi besar karena dapat memberikan akses yang mudah kepada penggunaannya.

### **3.3. Covid-19**

*CoronaVirus* atau sering disebut dengan Covid-19 digunakan penulis dalam penelitian ini sebagai acuan ataupun judul untuk mencari opini masyarakat terkait kasus Covid-19 yang terus meningkat dan sering dibahas baik dalam media sosial dan lainnya. Virus Corona atau disebut *Severce Acute Respiratory Syndrome* atau sering disebut Covid-19. Virus ini menyerang ke bagian pernapasan dalam tubuh manusia. Virus ini hadir pertama kali di kota Wuhan-*China*. Virus ini muncul di tahun akhir 2019 dan pada bulan Maret 2020 pihak dari WHO telah memberitakan bahwa virus ini menjadi dinyatakan sebagai pandemi penyerangan virus secara global [1]. Sejak saat itu dan akhir bulan Juli 2021 kasus positif virus Covid-19 ini sudah menyerang 188 juta jiwa di seluruh dunia. Dan Sekitar 4,06 juta lebih jiwa yang telah meninggal dunia dalam kasus virus covid-19 ini.

Sedangkan untuk kasus Covid-19 di Indonesia muncul pertama kali di bulan Maret tahun 2020. Saat ini kasus virus ini sudah ada di Indonesia 1 tahun lebih, dan untuk kasus positif di Indonesia pada Juli 2021 ada sekitar 3 juta jiwa lebih yang positif dan sudah 70 ribu lebih jiwa yang meninggal dunia. Kasus ini akan terus menambah sejalan waktu yang berjalan. Indonesia sendiri sudah melakukan penyelenggaraan vaksin ke masyarakat di bulan Januari 2021. Upaya pemerintah

dalam menangani kasus pandemi ini pun sudah dibentuk dalam sebuah tim. Tim tersebut diberi nama Gugus Tugas yang dibuat oleh Presiden Indonesia. Tugas dari Gugus Tugas untuk meningkatkan ketahanan nasional dibidang kesehatan dan pencegahan penyebaran virus lebih banyak.

#### **3.4. Vaksin Covid-19**

Vaksin Covid-19 digunakan penulis dalam penelitian ini sebagai acuan ataupun judul untuk mencari opini masyarakat. Meningkatnya terus penyebaran virus Covid-19 ini membuat setiap negara harus membuat Vaksin termasuk Indonesia. Indonesia sudah melakukan penyelenggaraan vaksin ke masyarakat mulai dari Januari 2021. Vaksin yang dibuat ini bukan lah obat, melainkan sebuah pembentukan kekebalan dalam sistem tubuh manusia agar terhindar maupun tertular dari sakit. Pemberian vaksin di Indonesia sudah berjalan selama 6 bulan lebih. Untuk sampai saat ini sudah 29 juta lebih jiwa yang menerima vaksin dosis pertama dan untuk vaksin dosis kedua sebanyak 13 juta lebih jiwa [3]. Jenis vaksin yang di sediakan di Indonesia ada tujuh vaksin dan yang digunakan di Indonesia ada tiga yaitu buatan *Sinovac* dari *China* , PT Bio Farma dari Indonesia dan *Oxford AstraZeneca* dari Inggris [5]. Meskipun sudah adanya vaksin tim Gugus Tugas dan selaku tim lain yang berhubungan dalam penanganan kasus Covid-19 ini selalu menyampaikan kepada seluruh masyarakat Indonesia untuk melakukan 3M (memakai masker, mencuci tangan dengan sabun dan menjaga jarak) agar tidak tertular oleh virus Covid-19.

#### **3.5. Data Mining**

Data *Mining* merupakan sebuah proses analisis yang bekerja dengan cara meninjau kumpulan data agar dapat memperoleh sebuah relasi yang tak terduga. Hal ini dilakukan agar mendapatkan manfaat untuk pemilik data. Hasilnya dapat dijadikan perbaikan dalam hal pengambilan keputusan kedepannya nanti.

Terdapat beberapa teknik yang ada pada data *Mining*. Tugas ini terbagi berdasarkan fungsi yang dapat dilakukan yaitu:

1. Deskripsi

Pada bagian ini dilakukan penggambaran dari pola yang ada serta kecenderungan yang ada pada data.

2. Estimasi

Pada bagian ini record yang telah lengkap akan menyediakan value yang berasal dari sebuah variabel target sebagai *value* dari prediksi.

3. Prediksi

Pada bagian ini dilakukan prediksi dari nilai yang merupakan hasil yang akan muncul di masa yang akan datang.

4. Klasifikasi

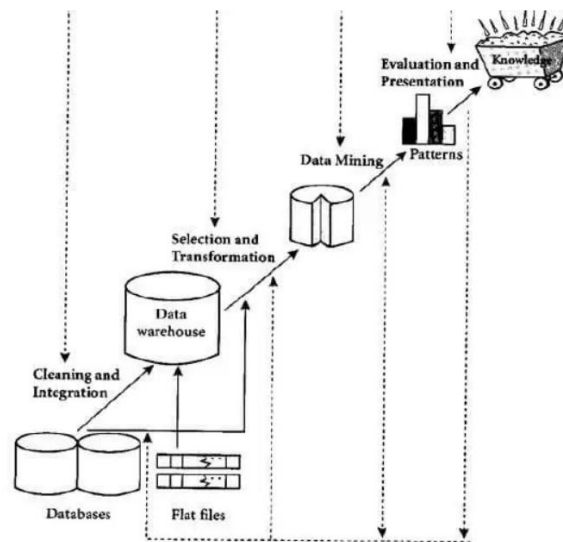
Pada bagian ini dilakukan pengelompokan menurut kelas-kelas ataupun ciri-ciri yang sama ataupun beda.

5. *Clustering*

Pada bagian ini dilakukan pengelompokan dari *record*, yaitu mengamati atau memperhatikan objek-objek yang mempunyai hal yang mirip.

6. Asosiasi

Pada bagian ini ditemukan atribut-atribut yang terdapat pada satu waktu. Hal ini lebih dikenal sebagai analisis keranjang belanja pada dunia bisnis. Pada penelitian ini digunakan fungsi klasifikasi karena akan melakukan pengelompokan berdasarkan karakteristik alternatif. Data mining sering kali disebut juga *knowledge discovery in database* atau KDD yang berguna untuk menjelaskan suatu proses dimana terdapat informasi-informasi yang tersimpan atau tersembunyi pada database yang besar. Penjelasan dari KDD ini dapat dilihat pada gambar 3.1 Proses KDD dalam Data Mining.



Gambar 3. 1. Proses KDD Dalam Data Mining [28]

Penjelasan gambar diatas terbagi menjadi beberapa tahapan yaitu:

a) Data *Cleaning* (Pembersihan data)

Pada tahap ini dilakukan penghilangan noise serta data-data yang tidak memiliki konsisten.

b) Data *Integration* (Integrasi data)

Pada tahap ini data dari berbagai database akan digabung agar menjadi database yang baru.

c) Data *Selection* (Seleksi data)

Pada tahap ini dicari data yang sesuai dari database yang kemudian akan dilakukan analisis. Ini disebabkan oleh terdapat data-data yang tidak relevan dan perlu diolah agar sesuai dengan data yang dibutuhkan.

d) Data *Transformation* (Transformasi data)

Pada tahap ini dilakukan penggabungan berdasarkat format yang sesuai pada data mining.

e) Proses Data *Mining*

Pada tahap ini dilakukan proses yang akan menemukan informasi-informasi yang berharga pada database.

f) *Pattern Evaluation* (Evaluasi pola)

Pada tahap ini dilakukanlah evaluasi terhadap pola-pola yang telah ditemukan.

g) *Knowledge Presentation* (Presentasi pengetahuan)

Pada tahap ini dilakukan proses dimana pola-pola ataupun sifat-sifat yang penting pada topik tertentu disimpan sehingga mendapatkan solusi dari masalah yang ada.

Pada tahap *pre-processing* diperlukan beberapa tahapan yaitu Stemming. Stemming digunakan sebagai fungsi yang akan mengenali kata-kata yang diinputkan, melakukan pemetaan serta penguraian terhadap kata-kata menjadi bentuk dasar atau bisa disebut juga *conflation*.

### 3.6. Multinomial Naïve Bayes

*Multinomial Naïve Bayes* adalah salah satu dari model yang ada pada *Naïve Bayes* yang berguna untuk melakukan klasifikasi terhadap teks. Metode yang digunakan oleh *Multinomial Naïve Bayes* adalah *supervised learning*, dimana setiap data yang ada perlu diberikan pelabelan sebelum melakukan *training* [15].

Penjelasan dasar dari teorema ini yaitu:

$$P(C) = \frac{\text{count}(c) + K}{N + K \cdot |\text{classes}|}$$

- P: Merupakan probabilitas yang berasal dari variabel c
- *Count(c)*: Merupakan jumlah dari kemunculan sampel c yang ada
- K: Merupakan nilai dari parameter
- *|classes|*: Merupakan jumlah kelas yang ada pada sampel

Kemudian melakukan penghitungan terhadap nilai dari atribut. Jumlah dari atribut yang telah diberi kelas ditambahkan satu, setelah itu dibagi dengan nilai dari jumlah kelas tertentu, ditambahkan lagi dengan hasil dari perkalian antara jumlah dari atribut dikali satu. Berikut adalah persamaan yang digunakan untuk melakukan perhitungan terhadap nilai dari  $p(w/c)$  dengan menggunakan laplacian smoothing.

### 3.7. Support Vector Machine

*Support Vector Machine* (SVM) merupakan salah satu metode dalam supervised learning yang biasanya digunakan untuk klasifikasi dan regresi. SVM memiliki konsep yang lebih matang dibandingkan dengan *engineering*. SVM juga dapat menyelesaikan masalah klasifikasi dan regresi dengan linier dan nonlinier [15].

Penjelasan dasar dari teorema ini yaitu:

$$f(x) = wx + b \quad \text{atau} \quad f(x) = \sum_{i=1}^m a_i y_i K + b$$

- $w$ : parameter hyperplane
- $x$ : inputan
- $b$ : parameter *hyperplane*
- $w$ : 1 (jumlah hyperline ke atas)
- $x$ : 3 (jumlah hyperline ke kanan)
- $b$ : 1

### 3.8. SentiStrength

*SentiStrength* merupakan program opinion *mining* yang dikembangkan oleh *CyberEmotion*. *SentiStrength* adalah contoh *classifier* yang menggunakan pendekatan berbasis leksikon yang berfokus pada pendeteksian kekuatan sentimen. Dengan memanfaatkan leksikal dengan sistem dual *scale*, *SentiStrength* ingin menunjukkan bahwa manusia dapat merasakan positif *emotion* dan negatif *emotion* secara bersamaan, sampai di batas tertentu secara independen.

*SentiStrength* akan menghasilkan *positive value* dan *negative value*. Range value/score dimulai dari angka 1 sampai 5. Nilai 1 menandakan data tersebut tidak memiliki sentimen positif ataupun negatif, sedangkan nilai 5 menandakan data tersebut memiliki sentimen yang “sangat positif” atau “sangat negatif”. Hanya value terbesar dari masing-masing emotion yang ditampilkan sebagai final value. Final value ini masih akan diberi tambahan nilai berupa 1 untuk positif dan -1 untuk negatif untuk mencegah emotion bernilai 0 atau kosong [29] [16] [30]. Kondisi ini

masih tetap berlaku meskipun emotion-nya sudah memiliki value. Keputusan akhir didasarkan pada algoritma sebagai berikut [26]:

- a) *if positive value > negative value* → Positif Sentimen
- b) *if positive value < negative value* → Negatif Sentimen
- c) *if positive value = negative value* → Netral Sentimen

Elemen yang akan dipakai adalah sebagai berikut:

- a) Kamus sentimen berikut nilainya.
- b) Pengoreksian penulisan.
- c) Intensifikasi/kata pendorong.
- d) Negasi.
- e) Sentimen negatif pada kalimat pertanyaan dianggap tidak memiliki sentiment

*SentiStrength* dapat dimodifikasi untuk bahasa berbeda dengan menerjemahkan isi leksikon default-nya, menyesuaikan opsi pengaturannya untuk mengatasi feature pada struktur jenis bahasa tertentu. Aturan-aturan yang ada di *SentiStrength* dapat dimodifikasi bahkan dihilangkan sesuai kebutuhan bahasa. Pengaplikasian ke bahasa selain English cenderung menghasilkan nilai akurasi yang lebih rendah karena analisis dan proses pembangunan memang lebih sering berfokus pada bahasa Inggris.

### 3.9. Confusion Matrix

*Confusion matrix* digunakan penulis untuk menentukan metode mana yang mendapatkan nilai terbaik dari data training untuk analisis tersebut. *Confusion matrix* merupakan sebuah metode untuk menguji suatu objek yang hasilnya untuk mencari hasil yang benar dan salah. Pada permasalahan yang menyangkut *Multy classification*, gunakanlah *precision*, *recall*, dan *accuracy* sebagai kriteria untuk melakukan evaluasi. Dan juga untuk mendapatkan nilai rata-rata dari *precision* dan *recall* digunakanlah *f1 – score*.

Berikut penjelasan dari masing-masing kriteria tersebut, yaitu [15]:



a) *Precision*

*Precision* merupakan suatu prediksi untuk membandingkan rasio *True Positive* (TP) dengan keseluruhan dari hasil yang memiliki prediksi positif.

$$Precision = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FP_i)} \times 100\%$$

b) *Recall*

*Recall* merupakan suatu prediksi untuk membandingkan rasio *True Positive* (TP) dengan keseluruhan data yang *True Positive* (TP).

$$Recall = \frac{\sum_{i=1}^l TP_i}{\sum_{i=1}^l (TP_i + FN_i)} \times 100\%$$

c) *Accuracy*

*Accuracy* merupakan suatu prediksi yang membedakan rasio *True Positive* (TP) dan *True Negative* (TN) dengan keseluruhan data yang ada.

$$Accuracy = \frac{\sum_{i=1}^l \frac{TP_i + TN_i}{TP_i + TN_i + FP_i + FN_i}}{l} \times 100\%$$

d) *f1-Score*

*f1-Score* merupakan perbandingan dari rata-rata nilai yang didapatkan dari *precision* dan *recall*.

$$F1 = \frac{2 \times Precision \times Recall}{Precision + Recall}$$

Penjelasan untuk TP, FP, FN, dan TN sebagai berikut:

TP (*True Positive*) adalah kata positif yang diklasifikasikan sebagai positif.

FP (*False Positive*) adalah kata negatif yang diklasifikasikan sebagai positif.

FN (*False Negative*) adalah kata positif yang diklasifikasikan sebagai negatif.

TN (*True Negative*) adalah kata negatif yang diklasifikasikan sebagai negatif.

### 3.10. Python 3

*Python* merupakan sebuah Bahasa pemrograman yang interpretatif multiguna, dimana kemampuan serta kapabilitas digabungkan sehingga dapat menghasilkan syntax yang amat cukup besar dan menyeluruh. *Python3* memiliki fitur yang sangat berguna yaitu manajemen memori secara otomatis. *Python* biasanya digunakan untuk mengembangkan suatu perangkat lunak tetapi juga mampu atau dapat berjalan di berbagai macam sistem operasi yang ada. Kode-kode pada python ketika di compile akan menghasilkan *byte code*, itulah kenapa *Python* sangat cocok jika digunakan sebagai Bahasa pemrograman pada *scripting*, aplikasi yang ada pada web, dan juga lainnya.