

BAB II

LANDASAN TEORI

2.1. Landasan / Kerangka Teori

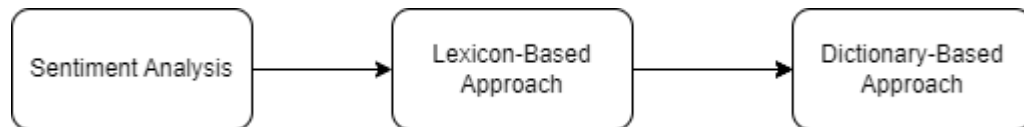
2.1.1. Kecerdasan Buatan

Penelitian yang akan dilakukan adalah *Text Mining* Bahasa manusia / *Natural Language Processing* yang merupakan bagian dari kecerdasan buatan terhadap postingan/komentar/cuitan masyarakat terhadap kinerja Pemerintah Republik Indonesia atas banyak kebocoran data publik yang telah terjadi.

Kecerdasan Buatan (*Artificial Intelligence*) merupakan salah satu bagian dari ilmu computer yang mempelajari tentang bagaimana membuat mesin dapat melakukan suatu pekerjaan seperti dan sebaik bahkan lebih baik dengan yang dilakukan oleh manusia (Dahria, 2008). Sementara (Siswanto, 2010) menyatakan bahwa *Artificial Intelligence* merupakan proses dimana peralatan mekanik dapat melaksanakan kejadian-kejadian dengan menggunakan pemikiran atau kecerdasan seperti manusia.

Kecerdasan buatan memiliki beberapa cabang, antara lain *Text Mining*. *Text Mining* sendiri merupakan proses menemukan informasi dalam kumpulan teks besar, dan secara otomatis mengidentifikasi pola dan hubungan yang menarik dalam data tekstual. Ini juga merupakan sebuah area penelitian yang relatif baru dan baru-baru ini menarik banyak minat di antara komunitas penelitian dan industri, terutama karena jumlah informasi yang tersedia di Web dan di tempat lain terus meningkat (Feldman dkk., 2007). Salah satu bagian dari *Text Mining* yang dapat digunakan untuk menghimpun opini masyarakat atas suatu kejadian adalah *Sentiment Analysis* (SA).

SA (*Sentiment Analysis*) dapat dilakukan salah satunya menggunakan *Lexicon-Based Approach*. Pada penelitian ini dilakukan implementasi SA (*Sentiment Analysis*) berbasis *Dictionary-Based Approach* terdapat dalam (Gambar 2.1).



Gambar 2. 1 *Sentiment Analysis*.

Sentiment analysis, juga dikenal sebagai *Opinion Mining* adalah riset komputasional dari opini, sentimen dan emosi yang diekspresikan secara tekstual. Analisis sentimen juga dilakukan untuk menentukan apakah opini atau komentar terhadap suatu permasalahan atau isu terkait topik tertentu memiliki kecenderungan positif, negatif, atau netral dan dapat dijadikan sebagai acuan dalam meningkatkan suatu pelayanan, ataupun meningkatkan kualitas produk (Liu, 2010). Untuk mengumpulkan data sentiment tersebut digunakan sebuah teknik yaitu *Natural Language Processing*.

Natural Language Processing secara teori adalah sebuah jangkauan teknik komputasi yang digunakan untuk menganalisa berdasarkan representasi teks yang terjadi secara alami pada satu atau lebih tingkat analisis linguistik untuk tujuan mencapai pemrosesan bahasa seperti manusia untuk berbagai tugas atau aplikasi (Liddy, 2001). Untuk memproses dan mengklasifikasikan sentiment yang berasal dari Bahasa manusia tersebut maka akan dipergunakan Bahasa pemrograman *Python*.

Python adalah bahasa pemrograman yang pemrograman tingkat tinggi yang bisa melakukan eksekusi sejumlah perintah secara langsung (interpretative), interaktif dengan metode OOP (Object Oriented Programming). Ini menggabungkan *modules*, *exceptions*, *dynamic typing*, *very high-level dynamic data types*, and *classes* (*General Python FAQ — Python 3.10.8 documentation*, t.t.; Nitnaware, 2019).

2.2. Penelitian Terdahulu

Telah banyak peneliti yang melakukan penelitian mengenai sentiment analysis dengan menggunakan algoritma yang berbeda-beda. Pada hasil penelitian yang dilakukan oleh peneliti terdahulu, didapatkan hasil akurasi yang berbeda disetiap algoritma yang digunakan. Berikut beberapa referensi dari penelitian terdahulu (table 2.1).

Table 2. 1 Penelitian Terdahulu

Judul, Penulis, Tahun	Algoritma	Open-Source dataset	Akurasi
<i>Sentiment Analysis algorithm and applications</i> (Medhat dkk., 2014)	Machine learning, lexicon, hybrid	-	-
Text Mining Dengan Metode <i>Lexicon Based</i> Untuk <i>Sentiment Analysis</i> Pelayanan PT. POS INSONESIA Melalui Media Social Twitter (Matulatuwa dkk., 2017)	Lexicon Based Approach	Media social <i>Twitter</i> PT Pos Indonesia	Opini positif 31%, Opini negatif 58%, dan Opini nihil 1%
<i>Lexicon Based Sentiment Analysis of Indonesian Language Student Feedback Evaluation</i>	Lexicon Based	Telkom Universit y student feedback data	Penelitian ini mendapatkan 91.2% akurasi dan performa terbaik adalah

(Firdaus dkk., 2021)			sentiments positif karena memiliki ketinggian prediksi tertinggi. Untuk akurasi dokumen didapatkan sebesar 90.9%
Penerapan <i>Lexicon Based</i> Untuk Analisis Sentiment Pada Twitter Terhadap Isu Covid-19 (Prasetya & Winarso, 2021)	Lexicon Based	Twitter API	opini sentimen kategori positif sebesar 58.08%, opini sentimen kategori negatif sebesar 37.61%, dan opini sentimen kategori netral sebesar 4.31%
<i>Sentiment Analysis of Twitter Data Using Naïve Bayes Algorithm</i> (Malik & Kumar, 2018)	Naïve Bayes	Crowdflo wer's	81.64%
Analisis Sentiment Masyarakat Terhadap Kasus Covid-19 Pada Media Sosial <i>YouTube</i> Dengan Metode <i>Naïve Bayes</i> (Ahmadi dkk., 2021)	Naïve Bayes	Komenta r Youtube pada channel Kompas Tv tentang kasus Covid-19.	74%

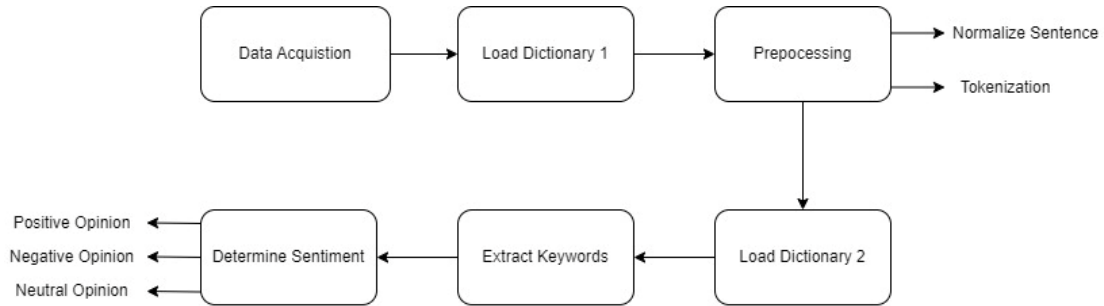
<i>Rahul Gandhi on Twitter: An Analysis of Brand Building Through Twitter by the Leader of the Main Opposition Party in Ind</i> (Antil & Verma, 2021)	-	<i>Tweet</i> Rahul Ghandi	46% very positive, 16% moderately positive, 8% moderately negative, 30% very negative.

Berdasarkan penelitian terdahulu, dapat disimpulkan bahwa *Lexicon-Based Approach* merupakan pendekatan yang dapat digunakan untuk menganalisa data *Sentiment* yang diambil dari opini masyarakat yang ada di media sosial.

Keterbaruan dari penelitian ini adalah menghimpun opini public atas kinerja Pemerintah dalam menangani keamanan data dan kebocorannya. Topik ini merupakan hal yang belum banyak di teliti berdasarkan data yang di dapat dari google scholar dengan kata kunci pencarian: sentimen analisis kinerja Pemerintah terhadap keamanan data dan sentimen analisis kinerja pemerintah terhadap kebocoran data yang dilakukan pencariannya pada periode 20 Oktober – 2 November 2022. Beberapa penelitian adalah membahas mengenai vaksin (Rachman & Pramana, 2020; Yulita dkk., 2021), kebocoran data (Wibowo dkk., 2021), analisis terhadap aplikasi (AINIYAH, 2022; Zaidah, t.t.).

2.3. Kerangka Pemikiran

Kerangka pemikiran dari penelitian ini digambarkan pada alur berikut (Gambar 2.2).



Gambar 2. 2 Tahapan Penelitian

Ada 6 tahapan yang akan dilakukan dari *Data Acquisition* hingga menentukan sentimen. Pada tahap pertama yaitu *Data Acquisition*, data berupa kalimat-kalimat yang diperoleh dari bahasa Indonesia mengenai hasil persepsi masyarakat di media sosial. Setelah data terkumpul, dilakukan *Load Dictionary 1* yang berisikan data dari Kamus Serangan. Kemudian dilakukan tahap *Preprocessing*. Tahap selanjutnya *Load Dictionary 2* yang berupa *positive keyword*, *negative keyword* dan *negation keyword*. Kemudian melakukan tahap *Extract Keyword* yang mencocokkan tiap token dengan jenis kata positif dan juga negative. Tahap yang terakhir adalah *Determine Sentiment* yang merupakan hasil dari pencocokan pada tahap *Extract Keyword* dengan jenis kata kunci yang menghasilkan *Data Sentiment*. Berikut adalah penjelasan tahap tahap dari diagram.

2.3.1. *Data Acquisition*

Pada *Data Acquisition* dilakukan pengambilan data yang diambil dari data *tweets* pada media sosial <https://twitter.com> dengan menggunakan *query* pencarian mengenai kasus kebocoran data yang terjadi di Indonesia. *Tweets* yang di ambil mengandung hastag #bjorka, #Bjorka, #BlokirKominfo, #Hacker, #Jebol, #Bocor, #Tersebar, #Tersiar, #Terbongkar, #Terekspos, #Terungkap, #Terkuak, #Tembus.

2.3.2. *Load Dictionary 1*

Pada tahap *Load Dictionary 1* dilakukan pemuatan kamus Serangan. Kamus serangan sendiri berisikan kata- kata yang berkaitan dengan kebocoran data.

2.3.3. *Preprocessing*

Pada tahap *Preprocessing* dilakukan normalisasi data pada data yang telah diambil sebelumnya. Normalisasi yang dilakukan pada penelitian ini ada beberapa tahap yaitu:

1) Penghapusan karakter

Pada data yang diambil dari *Twitter* sebelumnya masih terdapat banyak kata yang perlu dihapus seperti: hastag (#), username (@), email, dan juga tanda baca. Penghapusan pada kata yang berlebihan juga dilakukan untuk memperpendek data contohnya “Jangannn” menjadi “Jangan”.

2) Pergantian kata

Pada data yang sudah melalui proses penghapusan karakter, selanjutnya mengubah seluruh kalimat menjadi huruf kecil.

3) *Tokenization*

Pada tahap ini dilakukan pemecahan kalimat menjadi sebuah kata.

4) *Stemming*

Pada tahap *Stemming Data* setiap kata dilakukan perubahan data menjadi kata dasar, pada *Stemming* ini peneliti menggunakan library python *Sastrawi*.

2.3.4. *Load Dictionary 2*

Pada tahap *Load Dictionary 2* dilakukan load kamus yang berisikan *Positive Keyword*, *Negative Keyword* dan *Negation Keyword*.

2.3.5. *Extract Keywords*

Dalam tahapan ini dilakukan ekstrak kata kunci dari kalimat dengan menggunakan beberapa token yang sudah disediakan. Pada tahapan ini kalimat yang cuitan dipecah menjadi kata per kata atau toke. Contohnya: main rtp maxwin tinggi rilis kokotogel diproses menjadi [main, rtp, maxwin, tinggi, rilis, kokotogel].

2.3.6. *Determine Sentiment*

Dalam tahapan ini merupakan tahap dimana dilakukan penentuan sentiment yang dilakukan dengan menghitung probabilitas kemunculan kata kunci positif dan kata kunci negatif.