

**IMPLEMENTASI NAIVE BAYES DAN LEXICON BASED PADA
SENTIMEN ANALISIS LOCKDOWN PADA TWITTER
MENGUNAKAN BAHASA R**

TUGAS AKHIR

Program Studi DIV Teknik Informatika

Diajukan Untuk Memenuhi Kelulusan Matakuliah Tugas Akhir
Pada Program Studi D4 Teknik Informatika



Oleh :

Sulpadiani Bunyamin 1.16.4.096

D4 Teknik Informatika 4B

**PROGRAM DIPLOMA IV TEKNIK INFORMATIKA
POLITEKNIK POS INDONESIA
BANDUNG
2020**

LEMBAR PENGESAHAN

IMPLEMENTASI NAIVE BAYES CLASSIFIER DAN LEXICON BASED PADA SENTIMEN ANALISIS LOCKDOWN PADA TWITTER MENGGUNAKAN BAHASA R

Sulpadianti Bunyamin

1.16.4.096

Laporan Program Tugas Akhir ini telah diperiksa, disetujui dan disidangkan

Di Bandung, 25 Agustus 2020

Oleh :

Penguji Pendamping



Mohamad Harry Khomas Saputra,

S.T., M.T.I

NIK. 118.88.246

Penguji Utama



Cahyo Prianto, S.Pd., M.T.

NIK. 117.84.222

Menyetujui,
Koordinator Tugas Akhir



Woro Isti Rahayu, S.T., M.T.

NIK. 105.79.081

LEMBAR PENGESAHAN

IMPLEMENTASI NAIVE BAYES CLASSIFIER DAN LEXICON BASED PADA SENTIMEN ANALISIS LOCKDOWN PADA TWITTER MENGGUNAKAN BAHASA R

Sulpadianti Bunyamin

1.16.4.096

Laporan Program Tugas Akhir ini telah diperiksa, disetujui dan disidangkan

Di Bandung, 25 Agustus 2020

Oleh :

Pembimbing Utama,



Cahyo Prianto, S.Pd., M.T.

NIK. 117.84.222

Pembimbing Pendamping



Roni Andarsyah, S.T., M.Kom.

NIK. 115.88.193

Menyetujui,

Ketua Program Studi D4 Teknik Informatika,



M. Yusril Helmi S, S.Kom., M.Kom.

NIK. 113.74.163

ABSTRAK

Twitter merupakan salah satu media sosial yang digunakan di Indonesia untuk mengeluarkan pendapat/ opini. Salah satunya opini mengenai *Covid-19* yang sedang menggemparkan dunia. Banyak opini masyarakat yang menyarankan untuk melakukan “*lockdown*”. *Lockdown* merupakan kondisi seseorang tidak diperbolehkan meninggalkan tempat tinggal (17/4). Kebijakan *Lockdown* sendiri banyak menuai pro-kontra di media sosial salah satunya *twitter*. Pada penelitian ini “*lockdown*” akan dijadikan sebagai kata kunci untuk melakukan sentimen analisis. Sentimen analisis merupakan proses memahami, mengestrak mengolah suatu data textual secara otomatis untuk mendapatkan informasi yang terdapat dalam suatu kalimat opini. Metode *Lexicon Based* digunakan sebagai pemberian nilai sentimen positif atau negatif dan metode *Naïve Bayes Classifier* digunakan untuk melakukan klasifikasi dan menghitung jumlah akurasi dari *class* yang telah didapatkan. Berdasarkan hasil dari penelitian penulis mendapatkan total 1000 dokumen tweet dengan kata kunci “*Lockdown*” pada tanggal 5-Mei-2020 dengan 785 dokumen tweet yang telah melalui proses cleaning dan metode *Lexicon Based* melakukan pelabelan terhadap proses pembobotan yang sudah dilakukan dengan menghasilkan 370 tweet negatif, 257 tweet netral, dan 158 tweet positif. Kemudian metode *Naïve Bayes Classifier* data set sebanyak 785 data, kemudian membagi menjadi data training sebanyak 700 data, dan data testing sebanyak 85 data dan menghasilkan akurasi sebesar 0.6353.

Kata Kunci : Sentimen analisis, media sosial, *twitter*, R, *Lockdown*.

ABSTRACT

Twitter is one of the social media used in Indonesia to issue opinions / opinions. One of them is the opinion about Covid-19 which is taking the world by storm. Many public opinions suggest doing a "lockdown". Lockdown is a condition when a person is not allowed to leave their residence (17/4). The Lockdown policy itself has reaped many pros and cons on social media, one of which is Twitter. In this study, "lockdown" will be used as a keyword for analyzing sentiment. Sentiment analysis is a process of understanding, extracting, automatically processing a textual data to obtain information contained in an opinion sentence. The Lexicon Based method is used as a positive or negative sentiment value and the Naïve Bayes Classifier method is used to calculate the amount of accuracy of the class that has been obtained. Based on the results of the research, the authors obtained a total of 1000 tweet documents with the keyword "Lockdown" on 10-May-2020 with 785 tweet documents that had gone through the cleaning process and the Lexicon Based method labeled the weighting process that had been carried out by producing 370 negative tweets, 257 neutral tweets, and 158 positive tweets. Then the Naïve Bayes Classifier data set method is 785 data, then divides it into training data of 700 data, and testing data as much as 85 data and produces an accuracy of 95%.

Keywords: Sentiment Analysis, Social Media, Twitter, R, Lockdown.

KATA PENGANTAR

Puji syukur penulis ucapkan kepada Tuhan YME atas berkat, rahmat dan karunia-Nya sehingga laporan Tugas Akhir dengan judul **“Implementasi *Naive Bayes* dan *Lexicon Based* Pada Sentimen Analisis *Lockdown* Pada Twitter Menggunakan Bahasa R”** ini dapat terselesaikan. Banyak kendala yang dihadapi dalam penyusunan laporan ini dan penulis menyadari bahwa penyusunan laporan ini masih belum sempurna. Mengingat keterbatasan pengetahuan, pengalaman serta kemampuan penulis, penulis mengharapkan kritik dan saran yang sifatnya membangun dari pembaca.

Dalam penyusunan laporan Tugas Akhir ini tidak terlepas dari bantuan, bimbingan dan dorongan dari berbagai pihak. Untuk itu penulis mengucapkan terima kasih kepada:

1. Kedua orang tua dan keluarga penulis yang telah mendorong dan memberi semangat kepada penulis.
2. Cahyo Prianto, S.Pd., M.T. selaku dosen pembimbing I dan Roni Andarsyah, S.T., M.Kom. yang telah memberikan ilmu pengetahuan serta memberi masukan yang sangat bermanfaat dalam menyelesaikan laporan Tugas Akhir ini.

Akhirnya penulis berharap semoga apa yang penulis dikerjakan pada Tugas Akhir semester delapan ini dapat bermanfaat bagi pihak-pihak yang berkepentingan.

Bandung, 22 Agustus 2020

Penulis

DAFTAR ISI

ABSTRAK.....	i
<i>ABSTRACT</i>	ii
KATA PENGANTAR	iii
DAFTAR ISI.....	iv
DAFTAR GAMBAR	vi
DAFTAR SINGKATAN	viii
BAB I PENDAHULUAN	I-1
1.1 Latar Belakang.....	I-1
1.2 Identifikasi Masalah	I-2
1.3 Tujuan.....	I-2
1.4 Ruang Lingkup	I-2
BAB II LANDASAN TEORI	II-4
2.1 Analisis Sentimen	II-4
2.2 Media Sosial	II-4
2.2.1 <i>Twitter</i>	II-5
2.3 <i>Covid-19</i>	II-6
2.3.1 <i>Lockdown</i>	II-6
2.4 <i>Lexicon Based</i>	II-7
2.5 Klasifikasi	II-7
2.6 <i>Naïve Bayes Clasifier</i>	II-8
2.7 Bahasa Pemrograman R.....	II-8
BAB III GAMBARAN OBJEK STUDI.....	III-14
3.1 Penganalisaan Opini Media Sosial (Twitter)	III-14
3.2 Langkah Mendapatkan Data	III-16
3.3 Hasil <i>Crawling</i> Data Twitter.....	III-16
BAB IV METODOLOGI PENELITIAN	IV-17
4.1 Diagram Alur Metodologi Penelitian.....	IV-17
4.2 Tahapan dalam Penelitian	IV-18
4.2.1 Merumuskan Masalah.....	IV-18

4.2.1.1 Studi Literatur	IV-18
4.2.2 Pengumpulan Data	IV-18
4.2.2.1 <i>Request API Twitter</i>	IV-18
4.2.2.2 <i>Crawling Data</i>	IV-19
4.2.3 <i>Preprocessing Data</i>	IV-20
4.2.4 <i>Term Weighting</i> (Pembobotan Kata)	IV-22
4.2.5 <i>Lexicon Based</i>	IV-22
4.2.6 Klasifikasi	IV-23
4.2.6.1 <i>Naïve Bayes Classifier</i>	IV-23
4.2.7 Hasil	IV-23
4.2.8 Kesimpulan	IV-24
BAB V EXPERIMENT AND RESULT	V-25
5.1 <i>Experiment</i>	V-25
5.1.1 Implementasi RShiny	V-25
5.1.1.1 Server	V-25
5.1.1.2 UI	V-30
5.2 Result.....	V-35
BAB VI KESIMPULAN	VI-37
6.1 Kesimpulan Masalah.....	VI-37
6.2 Kesimpulan Metode	VI-37
6.3 Kesimpulan Pengujian	VI-37
BAB VII DISKUSI	VII-38
7.1 Diskusi.....	VII-38
DAFTAR PUSTAKA	
LAMPIRAN	

DAFTAR GAMBAR

Gambar 3.1 Hasil <i>Crawling Data</i>	III-15
Gambar 4.1 Alur Penelitian	IV-17
Gambar 4.2 <i>API Keys</i> dan <i>Access Token Secret</i>	IV-19
Gambar 4.3 Daftar <i>library</i> yang digunakan	IV-19
Gambar 4.4 <i>Data Crawling</i>	IV-17
Gambar 4.5 Menampilkan Hasil <i>Case folding</i>	IV-21
Gambar 4.6 Menampilkan Hasil <i>Stemming</i>	IV-21
Gambar 4.7 Menampilkan Hasil <i>Filtering</i> menggunakan <i>Stopword</i>	IV-22
Gambar 5.1 <i>Syntax</i> membaca <i>directory</i> kamus positif dan negatif	V-25
Gambar 5.2 <i>Syntax</i> data bersih	V-26
Gambar 5.3 <i>Syntax</i> Proses <i>score</i> (pembobotan)	V-26
Gambar 5.4 <i>Syntax</i> Proses <i>Cleaning</i> data	V-26
Gambar 5.5 <i>Syntax</i> Pemecahan Tweet	V-27
Gambar 5.6 <i>Syntax</i> mengubah list ke vector	V-27
Gambar 5.7 <i>Syntax</i> membandingkan kata dengan kamus	V-27
Gambar 5.8 <i>Syntax</i> pengubahan kata.....	V-28
Gambar 5.9 <i>Syntax</i> penjumlahan bobot kata	V-28
Gambar 5.10 <i>Syntax</i> membuat data frame.....	V-28
Gambar 5.11 <i>Syntax</i> penentuan hasil penjumlahan	V-29
Gambar 5.12 <i>Syntax</i> pembuatan file csv	V-29
Gambar 5.12 File CSV.....	V-30
Gambar 5.13 <i>Library</i> yang digunakan pada UI.....	V-30
Gambar 5.14 <i>Syntax</i> untuk membaca data.....	V-31
Gambar 5.15 <i>Syntax</i> tampilan pada <i>UI RShiny</i>	V-32
Gambar 5.16 <i>Library</i> yang digunakan <i>Naïve Bayes Classifier</i>	V-32
Gambar 5.17 <i>Syntax</i> menentukan <i>directory</i>	V-32
Gambar 5.18 <i>Syntax</i> baca data.....	V-32
Gambar 5.19 <i>Syntax</i> mengatur nomor awal	V-32
Gambar 5.20 <i>Syntax</i> mengubah jadi corpus	V-32
Gambar 5.21 <i>Syntax</i> pembagian data	V-33
Gambar 5.22 <i>Syntax</i> membuat jadi dtm.....	V-33
Gambar 5.23 <i>Syntax</i> DTM Data Training dan Data Testing	V-33
Gambar 5.24 <i>Syntax</i> Perhitungan NBC	V-33

Gambar 5.25 Apply Data Training dan Data Testing.	V-34
Gambar 5.26 Klasifikasi Data Training	V-34
Gambar 5.27 Proses prediksi data testing.	V-34
Gambar 5.28 Proses Confussion Matrix	V-34
Gambar 5.29 Membuat akurasi.....	V-34
Gambar 5.30 Hasil Sentimen Analisis	V-35
Gambar 5.31 Grafik Hasil Sentimen Analisis.....	V-35
Gambar 5.32 Akurasi Hasil Klasifikasi	V-36

DAFTAR SINGKATAN

No	Singkatan	Keterangan
1	<i>NBC</i>	<i>Naïve Bayes Classifier</i>
2	<i>API</i>	<i>Application Programming Interface</i>
3	<i>UI</i>	<i>User Interface</i>

BAB I

PENDAHULUAN

1.1 Latar Belakang Masalah

Keadaan dunia saat ini digemparkan oleh wabah virus *Covid-19* yang menyebar pesat ke beberapa wilayah di dunia termasuk Indonesia. *Covid-19* adalah virus jenis baru yang menular ke manusia dengan cara menyerang sistem pernapasan hingga akhirnya berujung pada kematian. Banyak kebijakan yang diambil pemerintah untuk melakukan pencegahan terhadap virus *Covid-19* seperti *physically distancing*. Banyak juga yang menyarankan untuk melakukan *lockdown*. *Lockdown* merupakan kondisi seseorang tidak diperbolehkan meninggalkan tempat tinggal. Presiden sendiri belum memilih untuk melakukan proses *Lockdown* di seluruh wilayah Indonesia. (18/3) Kebijakan *Lockdown* sendiri banyak menuai pro-kontra di media sosial. Media sosial digunakan masyarakat Indonesia untuk mengutarakan opini. Salah satu media sosial yang digunakan untuk mengutarakan opini seseorang adalah *twitter*. *Twitter* merupakan jejaring sosial yang banyak digunakan termasuk di Indonesia sehingga bisa dijadikan bahan untuk melakukan klasifikasi sentimen.

Sentimen analisis adalah bidang interdisipliner, sebuah bidang dimana pendekatan pemecahan masalahnya dengan menggunakan tinjauan dari berbagai sudut pandang ilmu serumpun secara relevan dan terpadu.[17] Adapun opini-opini tersebut didapatkan melalui penambangan data dokumen, *website*, atau media sosial[25]. Sentimen analisis merupakan proses untuk memahami, kemudian mengestrak, dan terakhir mengolah data textual secara otomatis yang digunakan untuk mendapatkan informasi sentimen yang terdapat dalam suatu kalimat opini. Sentimen analisis dilakukan untuk melihat kecenderungan opini terhadap suatu masalah atau objek oleh seseorang[17]

Pada penelitian ini peneliti menggunakan *API twitter* untuk mengambil data twitter. Kemudian menggunakan metode *Naïve Bayes Classifier* sebagai proses klasifikasi data yang telah melalui proses *term weighting* (pembobotan), sedangkan metode *Lexicon Based* digunakan untuk penentuan orientasi sentiment keseluruhan data. Seluruh proses dalam penelitian ini menggunakan bahasa pemrograman R i386 4.0.1 dan *IDE* yang digunakan adalah *Rstudio* 1.1.463. Hasil dari penelitian kali ini akan berupa klasifikasi opini tentang “lockdown” yang bersifat positif, netral, dan negatif. Dengan adanya penelitian ini, akan disimpulkan bagaimana sentimen twitter pendapat masyarakat Indonesia terhadap “lockdown”. Penelitian ini diharapkan dapat bermanfaat membantu untuk melakukan riset atas opini masyarakat yang mengandung sentimen positif, netral atau negatif

1.2 Identifikasi Masalah

Identifikasi Masalah yang timbul pada sentimen analisis twitter *Lockdown* menggunakan R, yaitu:

1. Bagaimana mengimplementasikan metode *Naïve Bayes Classifier (NBC)* dan *Lexicon Based* pada penelitian ini?
2. Bagaimana hasil klasifikasi opini masyarakat tentang *lockdown* pada media sosial(*twitter*)?

1.3 Tujuan

Tujuan dari penelitian ini, yaitu:

1. Mengimplementasikan metode *Naïve Bayes Classifier (NBC)* dan *Lexicon Based* untuk melakukan sentimen analisis.
2. Menghasilkan klasifikasi opini masyarakat tentang *lockdown* pada media sosial (*twitter*)

1.4 Ruang Lingkup

Ruang lingkup yang akan dibahas dalam penelitian ini, yaitu:

1. Sentimen analisis di twitter tentang *lockdown* dilakukan menggunakan aplikasi *RStudio*
2. Penelitian ini hanya melakukan analisis terhadap *tweet* Bahasa Indonesia
3. Data twitter dengan kata kunci *lockdown Indonesia* yang diambil sebanyak 1000 data pada tanggal 5 Mei 2020

BAB II

LANDASAN TEORI

2.1 Deskripsi Teori yang sama

2.1.1 *Sentiment Analysis on Twitter Data-set using Naïve Bayes Algorithm*

Penelitian ini membahas analisis sentiment dari web microblogging yaitu Twitter. Dengan menganalisis sentimen pada tweet yang membantu memberikan beberapa prediksi pada intelijen bisnis. Digunakan *Hadoop Framework* pada proses set data. Hasil analisis tersebut menyajikan sentimen positif, negatif, dan netral.[1]

2.1.2 *Sentiment Analysis of Movie Reviews using Machine Learning Techniques*

Penelitian ini menggunakan Metode *Naïve Baiyes*, *K-Nearest Neighbor*, dan *Random Fores* untuk proses sentiment analisis review film pada media sosial.[2]

2.1.3 *Sentiment analysis using deep learning architectures: a review*

Penelitian ini merangkum tentang dataset yang populer dengan fitur utama menggunakan *deep learning*. Akurasi diperoleh dari data yang dikumpulkan dari opini- opini masyarakat dan perbandingan dari beberapa model *deep learning*. Tujuan utama dari penelitian ini adalah untuk melihat kekuatan dari arsitektur *deep learning* itu sendiri untuk menjadi suatu *problem solving* pada masalah analisis sentiment.[3]

2.1.4 *Sentiment Analysis on Twitter Data*

Penelitian ini ditemukan bahwa media sosial *Twitter* dapat digunakan untuk memprediksi sentimen. Terdapat 3 algoritma *machine learning* pada penelitian ini yaitu *unigram*, *featured based model*, dan *tree kernel model* dengan menggunakan tool Weka. [4]

2.1.5 *Combining Naive Bayes and Adjective Analysis for Sentiment Detection on Twitter*

Kesimpulan dari penelitian ini adalah analisis sentiment bisa ditingkatkan dengan menggunakan kombinasi antara metode *naïve bayes* dan *adjective analysis*. Metode ini bekerja dengan baik dengan jenis kalimat yang telah ditentukan yang telah ditentukan. [5]

2.1.6 *Context-Sensitive Lexicon Features for Neural Sentiment Analysis*

Lexicon Sentimen digunakan sebagai fitur yang berguna untuk analisis sentiment yang mengarah ke akurasi akhir. sebagian besar metode yang ada menggunakan leksikon *sentimen* tanpa mempertimbangkan konteks, biasanya mengambil hitungan, jumlah kekuatan, atau skor sentimen maksimum atas seluruh input.[6]

2.1.7 *Sentiment Analysis of Movie Reviews: A study on Feature Selection & Classification Algorithms*

Penelitian ini menggunakan metode computation linguistic pada reprocessing data. Kemudian menampilkan analisis dampak fitur dengan menghitung perolehan informasi untuk setiap fitur dalam set fitur dan menggunakannya untuk memperoleh set fitur yang dikurangi. Di antara enam teknik klasifikasi, kami menemukan bahwa akurasi tertinggi diberikan oleh Random Forest dengan akurasi 88,95%.[7]

2.1.8 *Sentiment Analysis of Movie Reviews Based on CNN-BLSTM*

Dalam karya ini, kami menggambarkan serangkaian percobaan dengan *CNN* dan *BLSTM* dan memperoleh akurasi memuaskan yang bisa mencapai hingga 89,7%. Seluruh proses mengandung dua langkah. Langkah pertama adalah pra-pemrosesan data dan setiap kata diubah menjadi vektor kata 50D. Langkah kedua adalah menyelesaikan analisis sentimen dengan *CNN* dan *BLSTM*. Pada langkah ini, penelitian mengubah jumlah lapisan konvolusi dan penyatuan dalam *CNN*

untuk menemukan *CNN* sederhana dengan satu lapisan konvolusi dan satu lapisan penyatuan yang berkinerja terbaik dengan perbandingan.[8]

2.1.9 *Sentiment analysis of movie reviews: finding most important movie aspects using driving factors*

Penelitian ini dilakukan untuk menemukan aspek film yang mempengaruhi orientasi ulasan menggunakan faktor pendorong. Ini menyimpulkan dengan aspek Film, Akting dan Plot menghasilkan akurasi 79,372% untuk dataset saat ini dalam pertimbangan. Seperti yang dapat kita lihat dari hasil yang diperoleh untuk genre-spesifik, metode ini memberikan akurasi tinggi untuk beberapa genre, sementara itu memberikan akurasi yang lebih rendah untuk yang lain. Dengan demikian jelas bahwa metode yang digunakan untuk klasifikasi tinjauan campuran tidak baik untuk ulasan genre tertentu [9]

2.1.10 Deteksi Emosi Media Sosial Menggunakan Pendekatan Leksikon dan *Natural Language Processing*.

Hasil penelitian ini menunjukkan bahwa leksikon emosi mampu mendeteksi 55.45% atau 15.357 dari 27.696 kata yang diperoleh dari update status pengguna Facebook. *NLP* dapat dimanfaatkan untuk memperbaiki teks yang berasal dari update status. Hasil perbaikan dari *NLP* dicocokkan dengan leksikon yang telah dibuat untuk mengetahui label emosi dari suatu update status. Sebanyak 26 update status dapat terdeteksi label emosinya dan 61,53% di antaranya akurat. [10]

2.2 Deskripsi Metode yang sama

2.2.1 Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode *Naïve Bayes Classifier* (Studi Kasus Samsat Kota Malang)

Kesimpulan dari penelitian ini adalah membangun aplikasi analisis sentimen yang menerapkan pendekatan *Naïve Bayes Classifier* untuk mengklasifikasikan kata-kata dan difokuskan pada tweet dalam bahasa Indonesia. Data diperoleh melalui cara web scrapping dan sumber teks yang digunakan sebagai topik bahasan adalah Sistem Administrasi Manunggal Satu Atap (SAMSAT) Malang Kota. Proses klasifikasi dilakukan melalui serangkaian tahapan seperti

preproses (case folding, cleaning, tokenizing, dan stopword) serta proses klasifikasi dengan algoritma *Naïve Bayes Classifier* itu sendiri untuk mendapatkan hasil klasifikasi dengan kategori positif, negatif atau netral. Dari hasil uji akurasi klasifikasi yang dilakukan oleh aplikasi menghasilkan nilai akurasi tertinggi pada setiap kategori positif, negatif, netral masing-masing sebesar 82%, 92%, 80% dengan jumlah data latih 200 tweet negatif, 200 tweet positif, dan 200 tweet netral. [11]

2.2.2 Analisis Sentimen Opini Film Menggunakan Metode *Naïve Bayes* dan *Lexicon Based Features*

Pada penelitian ini, penggunaan metode *Naïve Bayes* dan *lexicon based features* untuk analisis sentimen opini film dapat dilakukan dengan melalui beberapa tahapan dimulai dari *text preprocessing*, menghitung nilai term weighting, dan klasifikasi *Naïve Bayes* dengan pembobotan nilai sentimen berdasarkan kamus Barasa. Pengujian metode *Naïve Bayes* dan *lexicon based features* untuk analisis sentimen opini film menghasilkan nilai accuracy sebesar 0,8, precision sebesar 0,8, recall sebesar 0,8, dan fmeasure sebesar 0,8. Hal tersebut dikarenakan adanya suatu kata yang memiliki nilai sentimen positif jauh lebih besar dibandingkan dengan nilai sentimen negatif pada dokumen yang kelas sebenarnya negatif, dan sebaliknya. Adanya perbedaan antara nilai sentimen positif dengan nilai sentimen negatif yang sangat besar dapat mengakibatkan kesalahan klasifikasi. Sedangkan penggunaan metode *Naïve Bayes* tanpa *lexicon based features* untuk analisis sentimen opini film menghasilkan nilai accuracy sebesar 0.95, precision sebesar 1, recall sebesar 0,9, dan f-measure sebesar 0,9474 yang mana hasil tersebut lebih baik dibandingkan dengan menggunakan *lexicon based features*. [12]

2.2.3 Analisis Sentimen Terhadap Kandidat Presiden Republik Indonesia Pada Pemilu 2019 di Media Sosial Twitter

Penelitian ini memfokuskan pada penggalian data berupa teks yang dihasilkan dari media sosial twitter yang merespon akun kandidat presiden RI pada pemilu 2019. Analisis sentimen diperoleh melalui klasifikasi tweet dengan

menggunakan sentiment analysis Tools seperti *NRC lexicon* dan *Bing Lexicon* sehingga diperoleh informasi berupa polaritas positif dan polaritas negatif dari tweet masyarakat terhadap para kandidat Presiden di pemilu 2019 dengan menggunakan data bulan Maret sebelum pemilu 2019.[13]

2.2.4 Analisis Sentimen Berdasarkan Opini Masyarakat Pada Twitter Menggunakan *Naïve Bayes*

Berdasarkan dari hasil penelitian yang telah dilakukan, dapat disimpulkan bahwa penelitian ini telah menghasilkan sebuah sistem analisis sentimen jasa pengiriman barang menggunakan naïve bayes. Sistem yang dibangun telah mampu mentransformasi sentimen yang berupa teks opini pada fanpage Tiki sehingga dapat menampilkan informasi sentimen masyarakat terhadap jasa pengiriman barang yang bersifat positif, negatif dan netral. Penggunaan metode naïve bayes classifier pada penelitian ini tidak menjamin ketepatan dalam proses klasifikasi. Akurasi pada sistem yang dibangun mencapai 84%. Tingkat akurasi akan semakin baik bila data latih yang digunakan lebih banyak, jumlah kategori pada setiap kelas sentimen pun akan berpengaruh.[14]

2.2.5 *Twitter Sentiment Analysis for Product Review Using Lexicon Method*

Penelitian ini telah mencoba menerapkan metodologi analisis sentimen berdasarkan kamus dan mengembangkan algoritma yang digunakan untuk sejumlah data yang besar untuk memperkirakan sentimen publik. Dengan cara mengganti kata akronim dengan membuat kamus akronim dan juga mendeteksi emotikon di tweet. Selain itu, dengan melakukan analisis level dokumen dan level aspek berdasarkan metodologi yang diusulkan, yang membantu dalam pengambilan keputusan.[15]

2.3 Teori Pendukung Penelitian

2.3.1 Analisis Sentimen

Analisis sentimen adalah bidang interdisipliner, sebuah bidang dimana pendekatan pemecahan masalahnya dengan menggunakan tinjauan dari berbagai sudut pandang ilmu serumpun secara relevan dan terpadu. Analisis sentimen terdiri dari pemrosesan bahasa alami, analisis teks dan komputasi linguistik untuk mengidentifikasi sentimen dari suatu dokumen[16]

Informasi tekstual secara umum dapat dibagi menjadi informasi fakta dan opini[17]. Tugas dasar dalam analisis sentimen adalah mengelompokkan polaritas dari teks yang ada dalam dokumen, apakah pendapat yang dikemukakan dalam dokumen bersifat positif, negatif atau netral. Penelitian mengenai analisis sentimen telah berkembang sejak tahun 2003 dan merupakan bagian dari text mining yang merupakan penelitian komputasi berdasarkan sentimen, emoticon, pendapat, komentar dan setiap ekspresi yang diungkapkan oleh teks. Analisis sentimen difokuskan untuk review klasifikasi berdasarkan polaritas. Berdasarkan klasifikasi, analisis sentimen dibagi menjadi dua kelompok utama. Yaitu dokumen klasifikasi ke pendapat atau fakta, atau dikenal sebagai klasifikasi subjektivitas (subjectivity classification) dan dokumen klasifikasi ke dalam positif atau negatif, atau dikenal sebagai analisis sentimen. Hal ini adalah proses yang penting untuk menentukan dokumen yang memiliki opini dan dokumen yang menyimpulkan opini bernilai positif, negatif maupun netral.[17]

2.3.2 Media Sosial

Media sosial adalah sebuah media online, dengan para penggunanya bisa dengan mudah berpartisipasi, berbagi, dan menciptakan isi meliputi blog, jejaring sosial, wiki, forum dan dunia virtual. Blog, jejaring sosial dan wiki merupakan bentuk media sosial yang paling umum digunakan oleh masyarakat di seluruh dunia.[18]

Dampak positif dari media sosial adalah memudahkan kita untuk berinteraksi dengan banyak orang, memperluas pergaulan, jarak dan waktu bukan lagi masalah, lebih mudah dalam mengekspresikan diri, penyebaran informasi dapat berlangsung secara cepat, biaya lebih murah. Sedangkan dampak negatif dari media sosial adalah menjauhkan orang-orang yang sudah dekat dan sebaliknya, interaksi

secara tatap muka cenderung menurun, membuat orang-orang menjadi kecanduan terhadap internet, menimbulkan konflik, masalah privasi, rentan terhadap pengaruh buruk orang lain. [19]

2.3.3 *Twitter*

Twitter merupakan salah satu layanan jejaring sosial, yang memungkinkan penggunanya untuk saling berkomunikasi. Pengguna dapat mengirim serta membaca pesan yang terdiri dari maksimal 140 karakter, yang disebut dengan *tweet*. *Tweet* dari pengguna lain yang diikuti atau di-*follow* akan muncul pada halaman home untuk dibaca. Pengguna bisa melakukan *retweet* atau mengirim ulang pesan yang dikirim oleh pengguna lain. Pada pesan yang dikirim, apabila menyebutkan nama pengguna lain maka pada *tweet* tersebut tertulis tanda @ diikuti nama pengguna. Pengguna dapat menggunakan tanda # (hashtag) untuk menulis pesan berdasarkan topik. *Twitter* didirikan pada Maret 2006 oleh *Jack Dorsey*. Pertumbuhan jumlah pengguna *Twitter* naik dengan cepat, sehingga kini terdapat 300 juta orang pengguna *Twitter*. *Twitter* naik ke peringkat kedua sebagai situs jejaring sosial yang paling sering dikunjungi di dunia, dari yang sebelumnya menempati peringkat dua puluh dua. Tingginya popularitas *Twitter* menyebabkan layanan ini telah dimanfaatkan untuk berbagai keperluan dalam berbagai aspek, misalnya sebagai sarana protes, kampanye politik, sarana pembelajaran, dan sebagai media komunikasi darurat[19]

2.3.4 *Covid-19*

Corona Virus Disease (COVID-19) adalah jenis virus baru yang menular pada manusia dan menyerang gangguan system pernapasan sampai berujung pada kematian. Tanda-tanda umum orang terinfeksi virus ini adalah demam di atas 38°C, batuk, sesak, dan susah bernapas. Virus ini berawal dari kota *Wuhan, China* yang diduga ditularkan melalui hewan kepada manusia, Virus tersebut menyebar sangat cepat hingga sampai pada Indonesia. Pada 2 Maret 2020, Presiden Joko 'Jokowi' Widodo mengumumkan kasus pertama (dan kedua) positif terinfeksi virus korona baru, atau terkena penyakit *COVID-19*, di Indonesia. Belakangan diketahui bahwa

dua orang (perempuan berusia 31 tahun dan ibunya berusia 64 tahun) mengetahui status mereka yang terinfeksi dari berita, dan bahwa Presiden mengumumkan hal tersebut kepada publik sebelum petugas kesehatan memberitahukan kepada mereka secara langsung[20].

2.3.4.1 Lockdown

Penyebaran virus *corona* atau Covid-19 di Indonesia terus meluas. Saat ini setidaknya sudah ada 117 kasus positif di tanah air, (16/3) dengan delapan orang sembuh dan lima orang meninggal. Informasi terbaru, virus sudah sampai ke kalangan pejabat negara. Menteri Perhubungan Budi Karya Sumadi dinyatakan positif virus *corona*. Tidak ingin kasus positif terus bertambah, pemerintah daerah hingga pusat pun mulai menerapkan kebijakan beraktivitas dan bekerja di rumah. Kebijakan mulai dari Gubernur DKI Jakarta Anies Baswedan hingga Presiden Joko Widodo (Jokowi), sampai pada kebijakan tiap daerah. Usulan *Lockdown* di tengah kekhawatiran penyebaran virus *corona*, mencuat gagasan untuk memberlakukan *lockdown* (penutupan suatu kota) di Indonesia. Berdasarkan kamus Bahasa Inggris, *lockdown* artinya kunci. Maksudnya, negara yang terinfeksi virus *corona* mengunci akses masuk dan keluar sebagai pengamanan ketat untuk mencegah penyebaran virus *corona*. [21]

Berdasarkan *Cambridge English Dictionary*, *lockdown* berarti sebuah kondisi isolasi atau pembatasan akses sebagai langkah pengamanan. Terkait virus korona, *lockdown* dapat diartikan upaya menutup wilayah alias karantina wajib untuk mencegah penyebaran virus korona. *Lockdown* juga dilakukan dengan larangan mengadakan pertemuan yang melibatkan banyak orang, penutupan sekolah, hingga tempat-tempat umum. Dengan begitu, risiko penularan virus *corona* pada masyarakat di luar wilayah lockdown bisa berkurang. Pemerintah menyatakan belum ada rencana memberlakukan *lockdown*. Meskipun, faktanya melansir dari cnnindonesia.com (16/3), jumlah kasus positif *Covid-19* terus bertambah. Konsekuensi yang sangat besar apabila *lockdown* dilakukan. *Lockdown* akan sangat memukul laju ekonomi Indonesia secara jangka pendek. Hal ini karena kota dengan kasus terbanyak virus *corona* ialah Jakarta, yang merupakan pusat

pemerintahan, bisnis, dan perdagangan Indonesia. Berangkat dari pertimbangan itulah, Indonesia belum juga menerapkan kebijakan tersebut. Saat ini Indonesia masih pada fase untuk memperkuat deteksi, pencegahan, serta memastikan tersedianya logistik medis seperti alat pelindung diri (APD). [21]

Kesimpulannya, kebijakan self distancing memang mau tidak mau sudah harus diterapkan. Sebab, rata-rata penyebaran virus corona di dalam negeri setidaknya mencapai 2,91 persen setiap jam. Kita bersama tentu tidak ingin seperti Italia dan Iran, yang ‘telat’ mengantisipasi penyebaran virus *corona* sampai akhirnya mau tidak mau menerapkan *lockdown*. Sementara itu, Indonesia bisa belajar dari China, yang merelakan laju perekonomiannya demi memutus mata rantai virus[21]

2.3.5 Pembobotan Kata

Pembobotan kata ialah proses mengubah indeks atau fitur dari hasil *preprocessing* yang berupa data kata-kata menjadi data angka dengan memberi nilai atau bobot untuk setiap kata. [9]. Hasil dari proses *term weighting* dapat digunakan untuk proses klasifikasi

2.3.6 Klasifikasi

Klasifikasi adalah proses pengidentifikasian obyek ke dalam sebuah kategori, kelas atau kelompok berdasarkan prosedur, definisi dan karakteristik yang telah ditentukan sebelumnya. Klasifikasi bertujuan untuk menempatkan objek yang ditugaskan hanya ke salah satu kategori yang disebut kelas [23]

2.3.6.1 Naïve Bayes Classifier

Naïve Bayes Classifier atau disebut juga dengan *Bayesian Classification* merupakan metode pengklasifikasian statistik yang didasarkan pada teorema bayes yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas.

Bayesian Classification terbukti memiliki akurasi dan kecepatan yang tinggi saat diaplikasikan ke dalam *database* yang besar.[23]

Bentuk umum *teorema bayes* adalah sebagai berikut:[10]

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

Keterangan :

X : Kriteria suatu kasus berdasarkan masukan

C_i : Kelas solusi pola ke-i, dimana i adalah jumlah label kelas

P(C_i|X) : Probabilitas kemunculan label kelas C_i dengan kriteria masukan X

P(X|C_i) : Probabilitas kriteria masukan X dengan label kelas

C_i P(C_i) : Probabilitas label kelas C_i

2.3.7 *Lexicon Based*

Lexicon based ialah kata yang telah diberi bobot berdasarkan kamus. Pemberian bobot dilakukan untuk setiap kata yang termasuk sentimen positif, negatif, atau netral. Tujuan penggunaan *Lexicon Based* pada penelitian ini yaitu untuk menentukan orientasi sentimen keseluruhan data berdasarkan hasil klasifikasi yang telah dibuat oleh metode sebelumnya. [22]

2.3.8 Bahasa Pemrograman R

R merupakan bahasa pemrograman statistika yang dapat digunakan untuk analisis dan manipulasi data statistika (pemodelan statistika), dan grafik. R diciptakan (*was created*) oleh *Ross Ihaka* dan *Robert Gentleman* (nama depan sama-sama diawali dengan huruf *R* & *R*, *Ross* dan *Robert*) dari departemen statistika, di *Universitas Auckland, New Zealand*. Saat ini R dikembangkan oleh *R Development Core Team*. Munculnya R terinspirasi oleh bahasa statistika dengan nama *S* (*statistical language S*). Bahasa statistika *S* diciptakan oleh *John Chambers*

dan rekan, di *Bell Labs* atau *AT&T Bell Laboratories*. Selain bahasa statistika *S*, bahasa pemrograman *C* juga dikembangkan di *AT&T Bell Laboratories*. [24]

Beberapa software seperti *SPSS*, *Minitab*, *SAS*, *EViews*, dan sebagainya, perlu membayar sejumlah uang, untuk mendapatkan izin menggunakan, sementara *R* gratis. Di samping itu, *R* juga tersedia untuk *di-install* di sistem operasi *Windows*, *Mac OS X*, dan *Linux*. [24]

BAB III

GAMBARAN OBJEK STUDI

3.1 Penganalisaan Opini Media Sosial (*Twitter*)

Pada sentimen analisis dibutuhkan proses untuk memahami, kemudian mengestrak, dan terakhir mengolah data *textual* secara otomatis sehingga menghasilkan informasi sentimen yang terdapat dalam suatu kalimat opini. Analisis sentimen dilakukan untuk melihat kecenderungan opini terhadap suatu masalah atau objek oleh seseorang, apakah cenderung beropini negatif atau positif.

3.2 Langkah Mendapatkan Data

1. *Crawling* data

Data didapatkan dengan cara melakukan *crawling data* pada twitter dengan menggunakan *R*. Pada proses *crawling data* memanfaatkan penggunaan *API Twitter*.

Proses untuk mendapatkan data dari *twitter* yaitu dengan *login* ke <https://developer.twitter.com> dan masuk ke <https://developer.twitter.com/en/apps> untuk *create an app*. Setelah mengisi *form* pembuatan aplikasi, maka *twitter* akan menyetujui pembuatan aplikasi twitter tersebut. Pada aplikasi *twitter* tersebut terdapat keys dan tokens yang akan digunakan dalam proses *crawling data*.

2. Studi Literatur

Studi Literatur sebagai dasar penulis untuk mendapatkan referensi yang mengenai teori – teori yang berkaitan dengan masalah – masalah yang bersumber dari buku literatur. Berdasarkan hasil studi pustaka yang telah dilakukan penulis, ditemukan solusi pemecahan masalah tentang analisis sentimen menggunakan metode *Lexicon Based Method* dan algoritma *Naïve Bayes Classifier*.

3.3 Hasil *Crawling Data Twitter*

Berikut tampilan beberapa hasil *crawling* data twitter pada tanggal 6 Mei 2020 dengan kata kunci “*lockdown* Indonesia”. Terdapat 1000 data yang akan digunakan pada penelitian ini.

Luhut: Presiden Jernih Melihat, Kalau Lockdown Ekonomi Mati! <https://cnbcindonesia.com/news/20200503220937-4-155924/luhut-presiden-jernih-melihat-perlukah-indonesia-menetapkan-darurat-covid-19-dan-menerapkan-lockdown?wpw>
 Perlukah Indonesia Menetapkan Darurat COVID-19 dan Menerapkan Lockdown? - <https://lapakfbku.com/perlukah-indonesia-menerapkan-lockdown?wpw>
 Nah itu pokok masalahnya, baru skala PSBB aja yg kena phk bnyk, gmn lockdown? Baru PSBB aja yg nyunat dana bansos banyak, gmn nanti lockdown? Makan
 Teriak lockdown melulu nih cucu presiden.. Ketahuilah nak.. Negara lagi bangkrut.. Ga punya duit.. Klo misalnya lockdown mau di kasih makan apa rakyat i
 Inggris Sebut Akan Longgarkan Lockdown Bertahap <https://cnnindonesia.com/internasional/20200504052411-134-499638/inggris-sebut-akan-longgarkan-lockdown>
 Kalau Jakarta Sekitarnya Lockdown, Wapadai Hal Ini - <https://lapakfbku.com/kalau-jakarta-sekitarnya-lockdown-wapadai-hal-ini?wpwautoposter=15885435>
 Ada unsur pemaksaan pada anak ga ya?? Ini jenis orang tua apa ya? Well seluruh indonesia menyaksikan. Bangga atau malu skrg dengan hal ini? Seperti kata
 Gmn mau berhasil, selama rezim @jokowi & Luhut menjegal kebijakan PSBB dg cara :
 Warganya disuruh lockdown tapi kenapa TKA CHINA 500orang boleh masuk indonesia???
 Baca berita, tentang Uni Emirat Arab selesai lockdown. seneng bgt sih semoga indonesia menyusul, dan negara lain pun juga Aamiin
 Wah ngeri deh pemerintah indonesia...
 Semua wish gue di 2020 sekarang gue pendem dan cuma ada satu wish yang gue pengen banget terwujud. Dapat berita kalo "lockdown indonesia sekarang s
 Negara lain udah buka lockdown eh indonesia makin parah
 LOCKDOWN BELUM PANTAS UNTUK INDONESIA!
 Mungkin nanti bakalan ada perayaan besar di INDONESIA stelah kelar LOCKDOWN..
 Krn itu Indonesia TIDAK DI LOCKDOWN. Klu IND LOCKDOWN, TUH TKA CHINA GA BISA MASUK SEDANGKAN MEREKA CARI DUIT DI IND. SEDANGKAN PSBB K

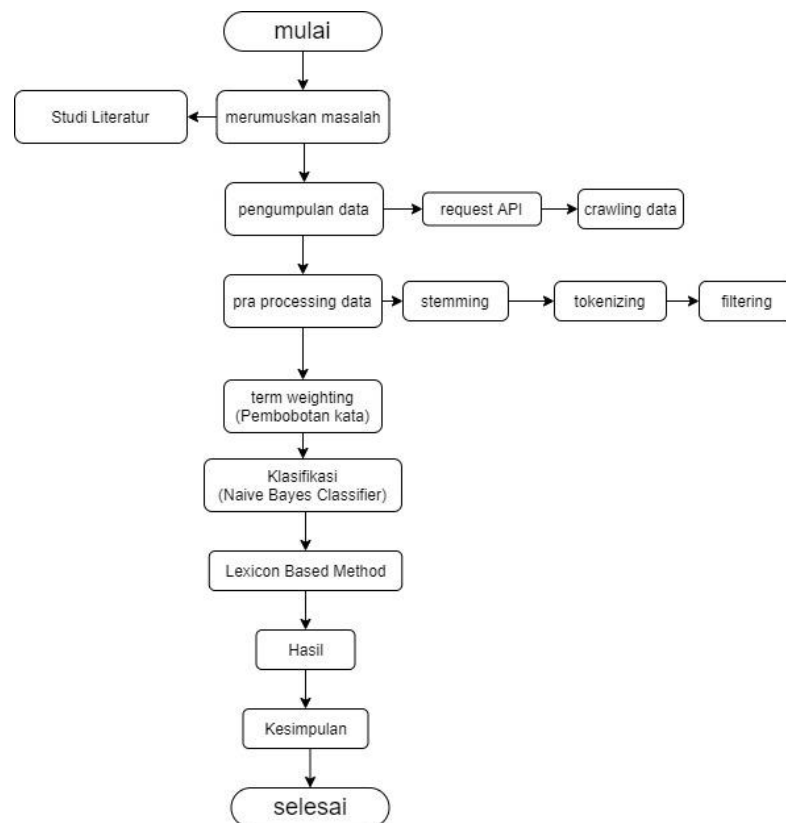
Gambar 3. 1 Hasil *Crawling Data*

BAB IV

METODOLOGI PENELITIAN

4.1 Diagram Alur Metodologi Penelitian

Pada diagram alur metodologi penelitian ini, penulis menjelaskan alur proses penelitian yang dilakukan. Penelitian ini akan diberikan ruang lingkup/pembahasan agar pembahasan yang dilakukan dapat terstruktur dan terorganisir dengan baik serta dapat mengefektifkan penelitian yang dilakukan karena proses awal hingga akhir telah ditentukan. Secara bertahap kerangka kerja penelitian ini dapat dijelaskan pada Gambar 4.1 berikut.



Gambar 4.1 Alur Penelitian

4.2 Tahapan dalam Penelitian

4.2.1 Merumuskan Masalah

Dalam merumuskan masalah terlebih dahulu dilakukan tahapan yaitu Studi Literatur

4.2.1.1 Studi Literatur

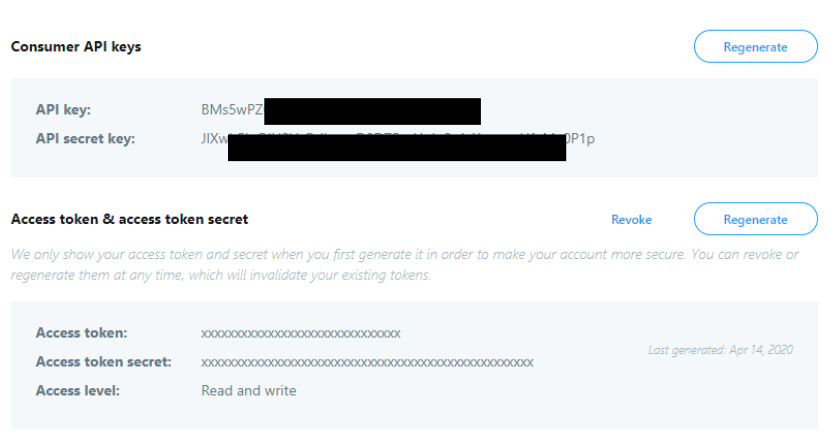
Studi Literatur sebagai dasar penulis untuk mendapatkan referensi yang mengenai teori – teori yang berkaitan dengan masalah yang bersumber dari buku literatur. Berdasarkan hasil studi pustaka yang telah dilakukan penulis, ditemukan solusi pemecahan masalah tentang analisis sentiment menggunakan algoritma *Naïve Bayes Classifier* dan *Lexicon Based*.

4.2.2 Pengumpulan Data

Dalam Pengumpulan data penulis dalam proses awal sebelum pengolahan data harus memiliki *API twitter* untuk melakukan sentimen analisis *twitter*, maka penulis melakukan *request API twitter* sebagai berikut :

4.2.2.1 Request API Twitter

1. Tahapan pertama dilakukan dengan login ke <https://developer.twitter.com> dan masuk ke <https://developer.twitter.com/en/apps> untuk *create an app*
2. Selanjutnya dengan mengisi *form Create an Application*.
3. Setelah mengisi *form*, maka *twitter* akan menyetujui pembuatan aplikasi *twitter* tersebut dan *key twitter* dapat dilihat pada *keys and tokens* yang terdapat pada aplikasi *developer twitter*.



Gambar 4.2 API Keys dan Access Token Secret

4.2.2.2 Crawling Data

Crawling data yaitu mengumpulkan data opini di dalam *twitter* yang mana di dalam *tweet* tersebut terdapat kata “*Lockdown*” sebagai kata kunci yang akan di olah.

Pada proses *crawling data*, *library* yang digunakan sebagai berikut:

```
library(twitter)
library(ROAuth)
library(RCurl)
```

Gambar 4.3 Daftar *library* yang digunakan

Keterangan:

1. *Library (twitterR)* adalah *package* yang digunakan sebagai penyedia akses API twitter sehingga dapat terhubung dan melakukan *crawling data twitter* dengan R
2. *Library (ROAuth)* berfungsi untuk membaca *key twitter* dan bisa menghubungkan R dengan *Twitter*.

3. *Library RCurl* yang digunakan memiliki fungsi sebagai tampilan dari R yang menghubungkan antara R dan *Library* dimana *Library* tersebut menyediakan fasilitas HTTP

Berikut data sampel hasil *crawling*:

Luhut: Presiden Jernih Melihat, Kalau Lockdown Ekonomi Mati! <https://cnbcindonesia.com/news/20200503220937-4-155924/luhut-presiden-jernih-melihat-kalau-lockdown-perlukah-indonesia-menetapkan-darurat-covid-19-dan-menerapkan-lockdown?wpwautoposter=15885>
 Nah itu pokok masalahnya, baru skala PSBB aja yg kena phk bnyk, gmn lockdown? Baru PSBB aja yg nyunat dana bansos banyak, gmn nanti lockdown? Makanya gw bilang lock
 Teriak lockdown melulu nih cucu presiden.. Ketahuilah nak.. Negara lagi bangkrut.. Ga punya duit.. Klo misalnya lockdown mau di kasih makan apa rakyat indonesia.. Masa
 Inggris Sebut Akan Longgarkan Lockdown Bertahap <https://cnnindonesia.com/internasional/20200504052411-134-499638/inggris-sebut-akan-longgarkan-lockdown-bertahap>
 Kalau Jakarta Sekitarnya Lockdown, Wapadai Hal Ini - <https://lapakfjbku.com/kalau-jakarta-sekitarnya-lockdown-wapadai-hal-ini?wpwautoposter=1588543970&pic.twitter>

Gambar 4.4 *Data Crawling*

4.2.3 Pra processing Data

Tahap preprocessing data adalah tahap permulaan pada proses text mining untuk memperoleh data yang siap untuk diproses[26]. Berikut tahap preprocessing data:

1. Proses *Case Folding*.

Proses *Case folding*, yaitu penyeragaman bentuk huruf agar data hanya menjadi huruf latin dan menjadi huruf kecil dengan *code* “*tolower*”.
 Contoh: “bingkai Foto” menjadi “bingkai foto”

```
dok_casefolding<-tm_map(corpusdok, content_transformer(tolower))
inspect(dok_casefolding[1:10])
```

```
[1] luhut presiden jernih melihat kalau lockdown ekonomi mati  â

[2] perlukah indonesia menetapkan darurat covid dan menerapkan lockdown  âpictwittercomffbtngie

[3] nah itu pokok masalahnya baru skala psbb aja yg kena phk bnyk gmn lockdown baru psbb aja yg nyunat dana bans
os banyak gmn nanti lockdown makanya gw bilang lockdown itu cm omong kosong belaka karena rakyat indonesia itu bn
yk yg bandel
[4] teriak lockdown melulu nih cucu presiden ketahuilah nak negara lagi bangkrut ga punya duit klo misalnya
lockdown mau di kasih makan apa rakyat indonesia masa harus ngutang lagi ini aja negara pusing ngurusnya akiba
t pemimpinnya terlalu pintar pintar ngibul
[5] inggris sebut akan longgarkan lockdown bertahap  â

[6] kalau jakarta sekitarnya lockdown wapadai hal ini  âpictwittercomqcyjzelg

[7] ada unsur pemaksaan pada anak ga ya ini jenis orang tua apa ya well seluruh indonesia menyaksikan bangga ata
u malu skrg dengan hal ini seperti kata seorang sultan yg bijaksana yg hrs kasih makan rakyat saat lockdown adala
h mereka yg meminta hal itu\mpaham tollollllll  â
[8] gmn mau berhasil selama rezim jokowi luhut menjegal kebijakan psbb dg cara \n tdk menghentikan transportasi
masal\n memasukkan tka cina byk\n alergi dg karantina\lockdown\n\nnrk ungkap penumpang krl bogorjakarta positif c
orona psbb bisa gagal \n
[9] warganya disuruh lockdown tapi kenapa tka china orang boleh masuk indonesia\napa gunanya nyuruh warga ngedek
em dihome hampir bulan  â

[10] baca berita tentang uni emirat arab selesai lockdown seneng bgt sih semoga indonesia menyusul dan negara lai
n pun juga aamiin
```

Gambar 4.5 Menampilkan Hasil *Case folding*

2. Proses *Stemming*

```
stem_text<-function(text,mc.cores=1)
{
  stem_string<-function(str)
  {
    str<-tokenize(x=str)
    str<-sapply(str,katadasaR)
    str<-paste(str,collapse = "")
    return(str)
  }
  x<-mclapply(X=text,FUN=stem_string,mc.cores=mc.cores)
  return(unlist(x))
}
dok_stemming<-tm_map(dok_slangword,stem_text)
inspect(dok_stemming[1:10])
```

```
[1] luhut presiden jernih lihat kalau lockdown ekonomi mati  â

[2] perlu indonesia tetap darurat covid dan terap lockdown  âpictwittercomffbtngie

[3] nah itu pokok masalah baru skala psbb aja yg kena phk bnyk gmn lockdown baru psbb aja yg nyunat dana bansos
banyak gmn nanti lockdown makanya gw bilang lockdown itu cm omong kosong belaka karena rakyat indonesia itu bnyk
yg bandel
[4] teriak lockdown melulu nih cucu presiden tahu nak negara lagi bangkrut ga punya duit kalau misal lockdow
n mau di kasih makan apa rakyat indonesia masa harus ngutang lagi ini aja negara pusing ngurusnya akibat pemim
pinnya lalu pintar pintar ngibul
[5] inggris sebut akan longgar lockdown tahap  â

[6] kalau jakarta kitar lockdown wapadai hal ini  âpictwittercomqcyjzelg

[7] ada unsur pemaksaan pada anak ga ya ini jenis orang tua apa ya well seluruh indonesia saksi bangga atau malu
skrg dengan hal ini seperti kata orang sultan yg bijaksana yg hrs kasih makan rakyat saat lockdown adalah mereka
yg minta hal itu\mpaham tollollllll  â
[8] gmn mau hasil lama rezim jokowi luhut jegal bijak psbb dengan cara \n tidak henti transportasi masal\n masu
k tka cina banyak\n alergi dengan karantina\lockdown\n\nnrk ungkap tumpang krl bogorjakarta positif corona psbb bi
sa gagal \n
[9] warga suruh lockdown tapi kenapa tka china orang boleh masuk indonesia\napa guna nyuruh warga ngedekem dihom
e hampir bulan  â

[10] baca berita tentang uni emirat arab selesai lockdown neng banget sih moga indonesia susul dan negara lain pu
n juga aamiin
```

Gambar 4.6 Menampilkan Hasil *Stemming*

3. Proses *Filtering*

Tahap *filtering* merupakan tahap untuk mengambil kata-kata penting dari hasil token tersebut. Berikut implementasinya: Selain itu digunakan *Stopwords* untuk menghapus kata-kata yang tidak terlalu penting.

```
library(tidytext)
cStopwordID<-readLines("stopwords.csv")
dok_stopword <- tm_map(dok_stemming, removeWords, cStopwordID)
inspect(dok_stopword[1:10])
```

```
[1] luhut presiden jernih lihat lockdown ekonomi mati a
[2] indonesia darurat covid terap lockdown apictwittercomffbtngie
[3] pokok skala psbb aja yg kena phk bnyk gmn lockdown psbb aja yg nyunat dana bansos gmn lockdown gw bi
lang lockdown cm omong kosong belaka rakyat indonesia bnyk yg bandel
[4] teriak lockdown melulu nih cucu presiden nak negara bangkrut ga duit lockdown kasih makan rakyat
indonesia ngutang aja negara pusing ngurusnya akibat pemimpinnya pintar pintar ngibul
[5] inggris longgar lockdown tahap a
[6] jakarta kitar lockdown wapai apictwittercomqcyze1g
[7] unsur pemaksaan anak ga ya jenis orang tua ya well indonesia saksi bangga malu skrg orang sultan
yg bijaksana yg hrs kasih makan rakyat lockdown yg npaham tollollllll a
[8] gmn hasil rezim jokowi luhut jegal bijak psbb n henti transportasi masal\n masuk tka cina n alergi
karantina\lockdown\n\nrk tumpang kr1 bogorjakarta positif corona psbb gagal\n
[9] warga suruh lockdown tka china orang masuk indonesia\n nyuruh warga ngedekem dihome a
[10] baca berita uni emirat arab selesai lockdown neng banget sih moga indonesia susul negara aamiin
```

Gambar 4.7 Menampilkan Hasil *Filtering* menggunakan *Stopword*

4.2.4 Pembobotan Kata

Pembobotan kata ialah proses mengubah indeks atau fitur dari hasil pra proses yang berupa data kata-kata menjadi data angka dengan memberi nilai atau bobot untuk setiap kata. [23]. Hasil dari proses *Pembobotan kata* dapat digunakan untuk proses klasifikasi.

4.2.5 Klasifikasi

Klasifikasi merupakan proses memasukan suatu data yang kelasnya belum diketahui berdasarkan data yang kelasnya sudah didefinisikan sebelumnya. Proses klasifikasi terbagi dalam dua fase, yaitu *learning step* dan *classification step*. Pada *learning step*, dilakukan pembuatan model klasifikasi dengan menggunakan algoritma yang mempelajari data yang kelasnya sudah didefinisikan sebelumnya

atau data latih. Karena kelasnya sudah diketahui sebelumnya, tahap ini disebut *supervised learning*. Selanjutnya model hasil *learning step* digunakan pada *classification step* untuk memprediksi suatu data yang kelasnya belum diketahui. [23]

4.2.5.1 Naïve Bayes Classifier

Naïve bayes classifier merupakan metode klasifikasi berbasis statistik berdasarkan teorema *Bayes* untuk mengklasifikasikan suatu data ke dalam kelas yang sudah ditentukan sebelumnya[23].

Proses pengujian data dilakukan dengan melakukan proses pengujian terhadap data yang diperoleh dari hasil sentimen *twitter* menggunakan R dengan cara menghitung data *training* dan menguji kinerja algoritma *Naïve Bayes Classifier*.

$$p(class + data) = \frac{p(data|class) p(class)}{p(data)}$$

4.2.6 Lexicon Based

Lexicon based ialah kata yang telah diberi bobot berdasarkan kamus. Pemberian bobot dilakukan untuk setiap kata yang termasuk sentimen positif, negatif, atau netral. Tujuan penggunaan *Lexicon Based* pada penelitian ini yaitu untuk menentukan orientasi sentimen keseluruhan data.

4.2.7 Hasil

Hasil implementasi nantinya akan menjawab masalah pada penelitian ini yaitu bagaimana mengimplementasikan metode *Naïve Bayes Classifier (NBC)* dan *Lexicon Based* dan hasil klasifikasi opini masyarakat tentang *lockdown* pada media sosial(*twitter*).

BAB V

EKSPERIMEN DAN HASIL

5.1 Eksperimen

Pada tahap ini penulis melakukan percobaan terhadap implementasi penelitian yang dilakukan berdasarkan data yang diperoleh dan memberikan hasil analisis data eksperimen yang dilakukan.

5.1.1 Implementasi *RShiny*

Pada pengimplementasiannya, Rstudio yang digunakan merupakan versi 4.0.1, kemudian digunakan *package* Shiny untuk mengizinkan penggunaanya membangun *web apps* yang interaktif. Rshiny sendiri memiliki 2 struktur, yaitu *server* dan *UI*. Berikut penjelasannya:

5.1.1.1 *Server*

Server berfungsi sebagai pendefinisian atas logika kerja analisis dari sisi *server* pada aplikasi.

1. Membaca kamus positif dan negatif yang terdapat pada *directory*.

```
43 pos = scan('E:/SEMESTER 8/R PROGRAMMING/Twitter-Sentiment-Analysis-using-R-Shiny-WebApp-master/positive_words.txt',what='character',
44           comment.char=';')
45 neg = scan('E:/SEMESTER 8/R PROGRAMMING/Twitter-Sentiment-Analysis-using-R-Shiny-WebApp-master/negative_words.txt',what='character',
46           comment.char=';')
47 pos.words = c(pos)
48 neg.words = c(neg)
```

Gambar 5.1 *Syntax* membaca *directory* kamus positif dan negatif

Keterangan:

1. *Syntax* nomor 43 - 46 merupakan code yang digunakan untuk membaca data pada *directory*.
2. *Syntax* nomor 47 dan 48 digunakan untuk mengcopy data dengan definisi bahwa “*pos.words*” memiliki isi yang sama dengan “*pos*”. begitu pula dengan “*neg.words*” memiliki isi yang sama dengan “*neg*”.

3. Membaca file data bersih

```
sentences<-read.csv('E:/SEMESTER 8/R PROGRAMMING/Twitter-Sentiment-Analysis-using-R-shiny-webApp-master/data_bersih.csv',header=TRUE)
```

Gambar 5.2 Syntax data bersih

Keterangan:

Data yang sebelumnya telah dibersihkan, akan dibaca lagi di R untuk melakukan proses pengolahan data.

4. Proses Perhitungan *Score* (Pembobotan)

```
score.sentiment = function(sentences, pos.words, neg.words, .progress='none')
{
  require(plyr)
  require(stringr)
```

Gambar 5.3 Syntax Proses *score* (pembobotan)

Keterangan:

Gambar diatas merupakan syntax mengartikan bahwa “sentences”, “pos.words”, dan “neg.words” akan diolah dengan membutuhkan package “plyr” dan “stringr”. Package plyr itu sendiri adalah package untuk merangkum data dengan cepat. Sedangkan package stringr adalah package untuk memanipulasi data string.

5. *Cleaning Data*

```
59 scores = laply(sentences, function(sentence, pos.words, neg.words) {
60   #gsub() function replaces all matches of a string
61   # clean up sentences with R's regex-driven global substitute, gsub():
62   sentence = gsub('[[punct:]]', '', sentence)
63   sentence = gsub('[[cntrl:]]', '', sentence)
64   sentence = gsub('\\d+', '', sentence) #angka
65   # and convert to lower case:
66   sentence = tolower(sentence)
```

Gambar 5.4 Syntax Proses *Cleaning* data

Keterangan:

1. Fungsi *syntax* nomor 59 yaitu akan menjumlahkan tweets positif dan *tweet* negatif, untuk sebuah kata positif bernilai +1 dan sebuah kata negatif bernilai -1, sedangkan kata yang normal bernilai 0 .

2. Sebelum melakukan perhitungan, untuk memastikan bahwa data yang digunakan telah bersih maka peneliti melakukan proses cleaning kedua yang ditunjukkan pada syntax nomor 62-66 diatas.

6. Pemecahan kalimat menjadi kata

```
# split into words. str_split is in the stringr package
word.list = str_split(sentence, '\\s+')
```

Gambar 5.5 Syntax Pemecahan Tweet

Keterangan:

Gambar diatas merupakan *syntax* yang digunakan untuk memecah tweet perkata kedalam sebuah list sehingga memudahkan proses pemberian score (pembobotan).

7. Perubahan *list*

```
# sometimes a list() is one level of hierarchy too much
words = unlist(word.list)
```

Gambar 5.6 Syntax mengubah list ke vector

Keterangan:

Gambar diatas merupakan *syntax* yang digunakan untuk mengubah list data kedalam *vector*

8. Perbandingan kata

```
pos.matches = match(words, pos.words)
neg.matches = match(words, neg.words)
```

Gambar 5.7 Syntax membandingkan kata dengan kamus

Keterangan:

Gambar diatas merupakan *syntax* yang digunakan untuk membandingkan kata yang digunakan dengan kamus kata negatif dan positif yang sudah dibuat sebelumnya.

9. Pengubahan kata

```
pos.matches = !is.na(pos.matches)
neg.matches = !is.na(neg.matches)
```

Gambar 5.8 Syntax pengubahan kata

Keterangan:

Gambar diatas merupakan *syntax* yang digunakan untuk mengubah kata yang cocok kedalam bentuk *TRUE* atau *FALSE*. Maksudnya *True* yang berarti 1 dan *False* -1. Kemudian akan ditambahkan fungsi *sum* untuk menjumlahkan.

10. Penjumlahan bobot kata

```
posscore = sum(pos.matches)
negscore = sum(neg.matches)
score=posscore-negscore
return(score)
```

Gambar 5.9 Syntax penjumlahan bobot kata

Keterangan:

Gambar diatas merupakan *syntax* yang digunakan untuk menjumlahkan bobot kata yang telah dicocokkan dengan kamus yang telah dibuat.

11. Membuat data *frame*

```
scores.df = data.frame(score=scores, text=sentences)
```

Gambar 5.10 Syntax membuat data frame

Keterangan:

Fungsi *syntax* diatas yaitu untuk mengubah data yang telah diolah menjadi data *frame*. Data *frame* itu sendiri merupakan data yang berisikan variable dengan karakter matriks.

12. Penentuan hasil penjumlahan

```
hasil = score.sentiment(sentences$text, pos.words, neg.words)
sum(hasil$score)
#View(hasil)
hasil$klasifikasi<- ifelse(hasil$score>0,'Positif', ifelse (hasil$score<0,'Negatif', 'Netral'))
```

Gambar 5.11 Syntax penentuan hasil penjumlahan

Keterangan:

Gambar diatas merupakan *syntax* yang digunakan untuk membuat seperti kolom baru yang terdapat pada data dengan nama klasifikasi dengan keterangan bahwa jika pada hasil yang dimaksudkan adalah kalimat/*tweet* memiliki *score* lebih besar daripada 0 maka kalimat/*tweet* yang terdapat pada data tersebut merupakan kalimat yang positif. Sedangkan jika kurang dari 0 maka kalimat/*tweet* tersebut merupakan kalimat negatif. Kemudian, untuk kalimat netral didefinisikan kepada kalimat/*tweet* yang tidak mengandung kata positif atau negatif samasekali. Kalimat netral juga bisa berartikan bahwa *score* kalimat/*tweet* tersebut memiliki jumlah yang imbang sehingga hasil perhitungannya 0.

13. Pembuatan *file csv*

```
write.csv(data, file = 'data_label.csv')
```

Gambar 5.12 Syntax pembuatan file *csv*

Keterangan:

Untuk melihat data yang telah diberi label positif, negatif dan netral, peneliti membuat data tersebut menjadi file *csv* dengan nama *data_label*. Gambaran file *csv*nya sebagai berikut.

14. Hasil *file csv*

klasifikasi score	text
Negatif	-1 luhut presiden lihat lockdown ekonomi mati
Negatif	-1 indonesia darurat covid terap lockdown
Negatif	-3 pokok skala psbb kena phk lockdown psbb sunat dana bansos gimana lockdown bilang lockdown omong kosong belaka rakyat indonesia bandel
Negatif	-1 teriak lockdown melulu cucu presiden negara bangkrut duit lockdown makan rakyat indonesia hutang negara ngurus akibat pemimpin pintar kibil
Negatif	-1 inggris longgar lockdown tahap
Netral	0 jakarta lockdown waspada
Negatif	-3 unsur paksa anak orang tua well indonesia saksi bangga malu orang sultan bijaksana makan rakyat lockdown paham tolol
Netral	0 gimana hasil rezim jokowi luhut bijak psbb henti transportasi masal masuk tenaga kerja asing cina alergi karantina lockdown tumpang krl bogor jakarta positif corona psbb gagal
Netral	0 warga suruh lockdown tenaga kerja asing china orang masuk indonesia suruh warga suruh dekam
Negatif	-2 baca berita uni emirat arab selesai lockdown banget semoga indonesia susul negara minimal
Negatif	-1 ngeri pemerintah indonesia larang mudik terap psbb lockdown tanggung rakyat dosa pemimpin adil rakyat menang lapar dosa ngalir
Positif	1 wish pendem wish banget wujud berita lockdown indonesia cabut warga indonesia bebas corona
Negatif	-1 negara buka lockdown indonesia parah
Positif	1 lockdown pantai indonesia unlock indonesia bebas mudik rezim taubat
Negatif	-1 raya indonesia kelar lockdown covid lockdown rumah
Positif	1 indonesia lockdown indonesia lockdown tenaga kerja asing china masuk cari duit indonesia psbb khusus kerja suruh rumah makan rakyat jokowi tenaga kerja asing china bebas n
Positif	3 negeri lockdown biaya negara indonesia psbb sengsara rakyat lockdown bodoh dunia psbb larang pasar tahan mall penuh minimalmarket buka dagang suruh tutup besaran rakya
Negatif	-5 indonesia darurat virus corona keluarga lapar akibat lockdown tinggal rumah cari rezeki keluarga butuh makan tolong bantu allah
Negatif	-2 lockdown cepat outus rantai sebar wilavah indonesia damoak neeatif skala luas oilih toko tutuo beli iarane phi iatuh

Gambar 5.12 File CSV

Keterangan:

Untuk melihat hasil sentimen *tweet* positif, netral, negatif. Selain itu terdapat *score*(bobot) pada masing- masing *tweet*.

5.1.1.2 UI

Userinterface merupakan tampilan antarmuka yang digunakan untuk menghubungkan antara *server* dan pengguna. *User interface* dimanfaatkan sebagai frontend dari sistem yang dibangun sehingga memberikan informasi atas pembangunan sistem yang lebih baik dan langsung kepada pengguna.

1. *Package/ library* yang digunakan

```
library(shiny)
library(ggplot2)
library(dplyr)
library(tidyverse)
```

Gambar 5.13 Library yang digunakan pada UI

Keterangan:

1. *Library shiny* yang digunakan untuk memudahkan pembangunan aplikasi web yang lebih interaktif, langsung terhadap bahasa *R*.
2. *Library Ggplot2* merupakan salah satu *package* yang sangat banyak digunakan oleh pengguna *R* untuk kebutuhan visualisasi.

3. *Library dplyr* digunakan *package* yang sangat berguna untuk melakukan manipulasi/transformatasi data menggunakan R.
4. *Library tidyverse* adalah *package* yang dirancang untuk ilmu *data science*.

2. Membaca data pada *directory*

```
housing <- read.csv("E:/SEMESTER 8/R PROGRAMMING/Twitter-Sentiment-Analysis-using-R-Shiny-webApp-master/data_label.csv",
                  header = TRUE)
```

Gambar 5.14 Syntax untuk membaca data

Keterangan:

Gambar diatas merupakan *syntax* yang digunakan untuk membaca data pada *directory*.

15. Tampilan *fluidpage* di *Rshiny*

```
ui <- fluidPage(
  mainPanel(
    tabsetPanel(
      tabPanel("Grafik", imageOutput("team_img")),
      tabPanel("Summary", checkboxGroupInput("regioninputID",
                                              "Pilih Sentimen",
                                              choices = unique(housing$klasifikasi)
                                              ),verbatimTextOutput("regionoutputID"), width = "auto", height = "auto"
    )
  )
)
```

Gambar 5.15 Syntax tampilan pada UI RShiny.

5.1.2 Penerapan Metode *Naïve Bayes Classifier*

Pada tahap ini peneliti menggunakan metode *Naïve Bayes Classifier* untuk melakukan klasifikasi terhadap data *set*. Data *set* adalah data yang diolah yang terdiri dari data training dan data testing. Selain itu metode ini digunakan untuk menghitung jumlah akurasi yang terdapat pada data set.

Adapun library yang digunakan:

```
library(tm)
library(RTextTools)
library(e1071)
library(dplyr)
library(caret)
```

Gambar 5.16 Library yang digunakan Naïve Bayes Classifier

Tahap penerapan *Naïve Bayes Classifier* pada penelitian ini yaitu:

1. Menentukan directory dengan menggunakan *code*:

```
setwd("E:/SEMESTER 8/R PROGRAMMING/Twitter-Sentiment-Analysis-using-R-Shiny-webApp-master")
```

Gambar 5.17 Syntax menentukan directory

2. Kemudian, membaca *file csv* dan mengubah data menjadi data *frame*. Fungsi data *frame* itu sendiri adalah membuat data agar tersusun menjadi baris dan kolom.

```
df<- read.csv("data_label.csv", stringsAsFactors = FALSE)
glimpse(df)
```

Gambar 5.18 Syntax baca data

3. Mengatur nomor awal yang digunakan untuk menghasilkan urutan nomor acak.

```
set.seed(1)
df <- df[sample(nrow(df)), ]
```

Gambar 5.19 Syntax mengatur nomor awal

4. Membuat *corpus*. *Corpus* merupakan data frame dari objek yang kolomnya dinamakan “text” tipe dari *corpus text*.

```
corpus <- Corpus(VectorSource(df$text))
```

Gambar 5.20 Syntax mengubah jadi corpus

5. Pembagian data *training* dan data *testing*. Pada penelitian ini data *training* terdiri dari 700 data, yaitu data nomor urut 1 sampai dengan 700. Sedangkan data *testing* terdiri dari 85 data, yaitu data nomor urut 701 sampai dengan 785, dengan menggunakan *script* seperti dibawah ini:

```
df.train <- df[1:700,]
df.test <- df[701:785,]

dtm.train <- dtm[1:700,]
dtm.test <- dtm[701:785,]

corpus.train <- corpus[1:700]
corpus.test <- corpus[701:785]
```

Gambar 5.21 Syntax pembagian data

6. Mengubah data menjadi *document term matrix*. *Document Term Matrix* merupakan dasar dari *matrix* dimana dokumennya didesain berdasarkan baris dan kata- kata pada kolom.

```
fivefreq <- findFreqTerms(dtm.train, 10)
length((fivefreq))
```

Gambar 5.22 Syntax membuat jadi dtm

7. Pembuatan data *training* dan data *testing*.

```
dtm.train.nb <- DocumentTermMatrix(corpus.train, control=list(dictionary = fivefreq))
dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.test, control=list(dictionary = fivefreq))
dim(dtm.test.nb)
```

Gambar 5.23 Syntax DTM Data Training dan Data Testing

8. Proses pelabelan seperti berikut:

```
dtm.train.nb <- DocumentTermMatrix(corpus.train, control=list(dictionary = fivefreq))
dim(dtm.train.nb)
dtm.test.nb <- DocumentTermMatrix(corpus.test, control=list(dictionary = fivefreq))
dim(dtm.test.nb)

# Function to convert the word frequencies to yes (presence) and no (absence) labels
convert_count <- function(x) {
  y <- ifelse(x > 0, 1, 0)
  y <- factor(y, levels=c(0,1), labels=c("No", "Yes"))
  y
}
```

Gambar 5.24 Syntax Perhitungan NBC

9. Mengaplikasikan *data training* dan *data testing*.

```
trainNB <- apply(dtm.train.nb, 2, convert_count)
testNB <- apply(dtm.test.nb, 2, convert_count)|
```

Gambar 5.25 Apply Data Training dan Data Testing.

10. Proses klasifikasi *data training*

```
# Train the classifier
system.time(classifier<-naiveBayes(trainNB,df.train$klasifikasi,laplace=1))
```

Gambar 5.26 Klasifikasi Data Training

11. Proses prediksi dari data testing

```
# Use the NB classifier we built to make predictions on the test set.
system.time(pred<-predict(classifier, newdata=testNB) )
```

Gambar 5.27 Proses prediksi data testing.

12. Proses *Confussion Matrix*. *Confusion matrix* itu untuk melihat berapa banyak data yang sesuai dengan class positif, negatif dan netral berdasarkan hasil perhitungan *class prediction*.

```
# Create a truth table by tabulating the predicted class labels with the actual class labels
table("Predictions"= pred, "Actual" = df.test$kategori )
str(pred)
str(df.test)
# Prepare the confusion matrix
conf.mat <- confusionMatrix(factor(pred, levels = 1:785),
                             factor(df.test$final, levels = 1:785))
```

Gambar 5.28 Proses Confussion Matrix

13. Hasil akurasi. Hasil akurasi diperoleh berdasarkan perhitungan *confussion matrix*.

```
conf.mat$byClass

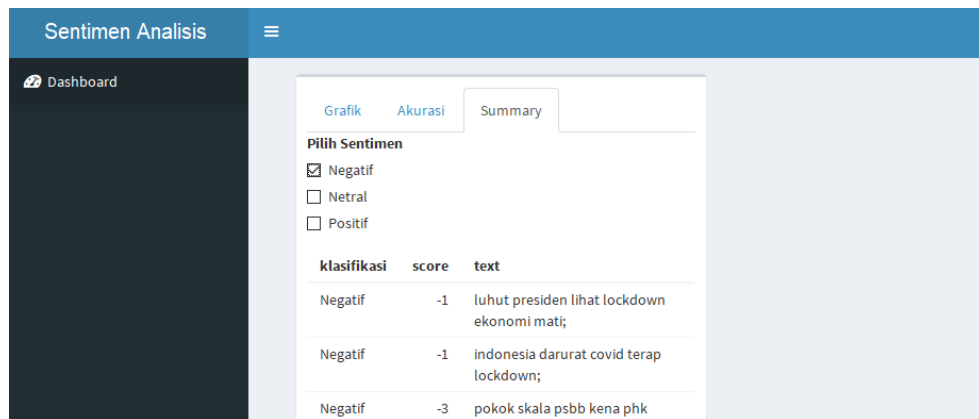
conf.mat$overall
conf.mat$overall['Accuracy']
```

Gambar 5.29 Membuat akurasi

5.2 Hasil

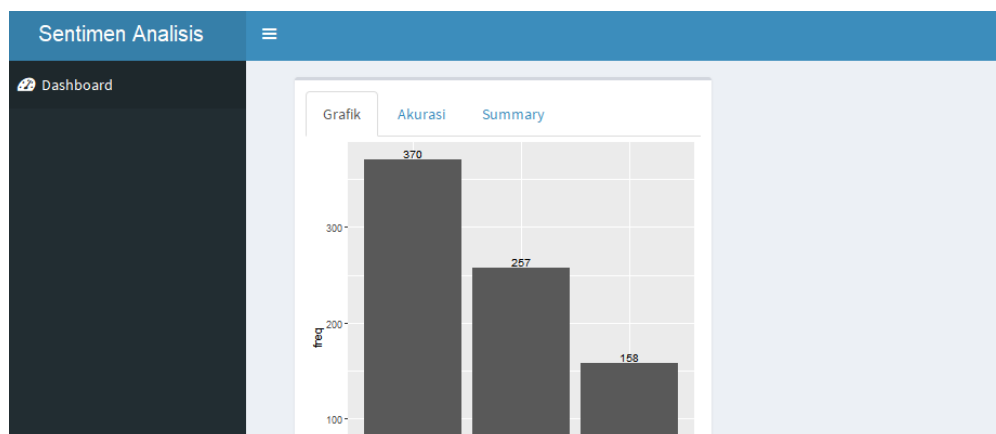
1. Hasil Sentimen Analisis

Pada pembobotan yang telah dilakukan, *metode lexicon based* menghasilkan tweet sebanyak 158 kalimat positif, 257 kalimat netral, dan 370 kalimat negatif.



Gambar 5.30 Hasil Sentimen Analisis

2. Grafik hasil sentimen



Gambar 5.31 Grafik Hasil Sentimen Analisis

3. Kemudian hasil dari *Metode Naïve Bayes Classifier* yaitu akurasi sebesar 0.6353.

Prediction	Negatif	Netral	Positif
Negatif	31	3	8
Netral	10	17	4
Positif	2	4	6

Overall Statistics

Accuracy : 0.6353

95% CI : (0.5238, 0.7371)

No Information Rate : 0.5059

P-Value [Acc > NIR] : 0.01103

Kappa : 0.4091

Mcnemar's Test P-Value : 0.06102

Gambar 5.32 Akurasi Hasil Klasifikasi

BAB VI

KESIMPULAN

6.1 Kesimpulan Masalah

Berdasarkan hasil analisis dan uraian yang telah dikemukakan pada bab-bab sebelumnya, maka penulis dapat mengambil keputusan untuk kesimpulan masalah dalam penelitian ialah sebagai berikut :

1. Implementasi metode *Lexicon Based* dan *Naïve Bayes Classifier* dapat digunakan untuk melakukan sentimen analisis dengan menggunakan R.
2. Hasil sentimen analisis dapat ditampilkan kedalam *user interface R shiny* dengan baik dan jelas

6.2 Kesimpulan Metode

Dari metode yang digunakan penulis dalam penelitian berikut, dapat disimpulkan bahwa :

1. Metode *Naïve Bayes Classifier* berhasil melakukan proses sentimen analisis dengan baik terhadap tweet “*lockdown*” dengan data set sebanyak 785 data, kemudian membagi menjadi data *training* sebanyak 700 data, dan data *testing* sebanyak 85 data dan menghasilkan akurasi sebesar 95%.
2. Hasil klasifikasi menunjukkan bahwa 370 tweet adalah kalimat yang bersifat negatif, 257 kalimat netral, dan 158 kalimat positif. Sehingga, metode *Lexicon Based* menyimpulkan bahwa tweet “*lockdown*” ditanggapi negatif oleh pengguna *twitter* atau masyarakat berdasarkan tweet yang telah didapatkan.

6.3 Kesimpulan Pengujian

Berdasarkan pengujian terhadap tujuan dan metode yang sudah dilakukan, maka dapat disimpulkan bahwa :

Dengan menghasilkan klasifikasi yang lebih didominasi oleh sentimen negatif, pemberlakuan “*lockdown*” dinilai tidak cocok untuk diberlakukan di Indonesia.

Setelah mengetahui hasilnya, dituangkan dalam bentuk artikel, koran, atau jurnal ilmiah yg bisa dibaca oleh masyarakat luas

BAB VII

DISKUSI

7.1 Diskusi

Berdasarkan kesimpulan diatas, maka penulis merekomendasikan berupa saran-saran berikut :

1. Pada penerapannya, penelitian ini hanya menggunakan R, diharapkan kedepannya, agar penerapannya dijalankan pada basis lainnya seperti web, mobile, dan lain-lain.
2. Pada penelitian berikutnya, diharapkan agar bisa dikomparasi dengan menggunakan metode yang berbeda guna menghasilkan suatu perbandingan atau penelitian yang lebih baik lagi.

DAFTAR PUSTAKA

- [1] Parveen, Huma, and Shikha Pandey. "Sentiment analysis on Twitter Data-set using Naive Bayes algorithm." *2016 2nd international conference on applied and theoretical computing and communication technology (iCATccT)*. IEEE, 2016.
- [2] Baid, Palak, Apoorva Gupta, and Neelam Chaplot. "Sentiment analysis of movie reviews using machine learning techniques." *International Journal of Computer Applications* 179.7 (2017): 45-49.
- [3] Ain, Qurat Tul, et al. "Sentiment analysis using deep learning techniques: a review." *Int J Adv Comput Sci Appl* 8.6 (2017): 424.
- [4] Sahayak, Varsha, Vijaya Shete, and Apashabi Pathan. "Sentiment analysis on twitter data." *International Journal of Innovative Research in Advanced Engineering (IJIRAE)* 2.1 (2015): 178-183.
- [5] Mertiya, Mohit, and Ashima Singh. "Combining naive bayes and adjective analysis for sentiment detection on Twitter." *2016 International Conference on Inventive Computation Technologies (ICICT)*. Vol. 2. IEEE, 2016.
- [6] Teng, Zhiyang, Duy-Tin Vo, and Yue Zhang. "Context-sensitive lexicon features for neural sentiment analysis." *Proceedings of the 2016 conference on empirical methods in natural language processing*. 2016.
- [7] Sahu, Tirath Prasad, and Sanjeev Ahuja. "Sentiment analysis of movie reviews: A study on feature selection & classification algorithms." *2016 International Conference on Microelectronics, Computing and Communications (MicroCom)*. IEEE, 2016.
- [8] Shen, Qianzi, Zijian Wang, and Yaoru Sun. "Sentiment analysis of movie reviews based on cnn-blstm." *International Conference on Intelligence Science*. Springer, Cham, 2017.

- [9] Parkhe, Viraj, and Bhaskar Biswas. "Sentiment analysis of movie reviews: finding most important movie aspects using driving factors." *Soft Computing* 20.9 (2016): 3373-3379.
- [10] Rohman, Arif Nur, Ema Utami, and Suwanto Raharjo. "Deteksi Kondisi Emosi pada Media Sosial Menggunakan Pendekatan Leksikon dan Natural Language Processing." *Jurnal Eksplora Informatika* 9.1 (2019): 70-76.
- [11] Rozi, Imam Fahrur, Elok Nur Hamdana, and Muhammad Balya Iqbal Alfahmi. "Pengembangan Aplikasi Analisis Sentimen Twitter Menggunakan Metode Naïve Bayes Classifier (Studi Kasus Samsat Kota Malang)." *Jurnal Informatika Polinema* 4.2 (2018): 149-149.
- [12] Kurniawan, Arifin, and Sigit Adinugroho Indriati. "Analisis Sentimen Opini Film Menggunakan Metode Naïve Bayes dan Lexicon Based Features." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer e-ISSN 2548: 964X*.
- [13] Prianto, Cahyo, Nisa Hanum Harani, and Indra Firmansyah. "Analisis Sentimen Terhadap Kandidat Presiden Republik Indonesia Pada Pemilu 2019 di Media Sosial Twitter." *JURNAL MEDIA INFORMATIKA BUDIDARMA* 3.4 (2019): 405-413.
- [14] Lorosae, Teguh Ansyor, et al. "Analisis Sentimen Berdasarkan Opini Masyarakat pada Twitter Menggunakan Naive Bayes." *Semnasteknomedia Online* 6.1 (2018): 1-10.
- [15] Ray, Paramita, and Amlan Chakrabarti. "Twitter sentiment analysis for product review using lexicon method." *2017 International Conference on Data Management, Analytics and Innovation (ICDMAI)*. IEEE, 2017.
- [16] Vinodhini, G. & Chandrasekaran, R.M. 2016. A comparative performance evaluation of neural network based approach for sentiment classification of online reviews. *Journal of King Saud University - Computer and Information Sciences*, 28(1): 2–12.

- [17]Raharja, Yosoa Putra. "Rancang Bangun Sistem Rekomendasi Beasiswa Menggunakan Algoritma Klasifikasi C4. 5 Pada Universitas Dian Nuswantoro." Universitas Dian Nuswantoro, Semarang, skripsi (2014).
- [18] Utami, Agustin Dyah, and Ramadian Agus Triyono. "Pemanfaatan Blackberry Sebagai Sarana Komunikasi Dan Penjualan Batik Online Dengan Sistem Dropship Di Batik Solo 85." *Speed-Sentra Penelitian Engineering dan Edukasi* 3.3 (2013).
- [19] Cahyono, Anang Sugeng. "Pengaruh media sosial terhadap perubahan sosial masyarakat di Indonesia." *Jurnal Publiciana* 9.1 (2016): 140-157.
- [20] Yusuf, Thalia. 2020. "Gaya hidup orang percaya berlandaskan Mazmur 91 : 1-16 dalam menyikapi masalah virus corona (Covid-19) masa kini". Institut Agama Kristen Negeri Toraja
- [21]Lestari, Novi Puji. "Menimbang Mahalnya Biaya 'Lockdown'." *Arsip Publikasi Ilmiah Biro Administrasi Akademik* (2020).
- [22] Rofiqoh, Umi, Rizal Setya Perdana, and M. Ali Fauzi. "Analisis sentimen tingkat kepuasan pengguna penyedia layanan telekomunikasi seluler indonesia pada twitter dengan metode Support Vector Machine dan Lexicon Based Features." *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer* e-ISSN 2548 (2017): 964X.
- [23] Muhamad, Husin, et al. "Optimasi Naïve Bayes Classifier Dengan Menggunakan Particle Swarm Optimization Pada Data Iris." *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, p-ISSN (2017): 2355-7699.
- [24] Gio, Prana Ugiana, and Adhitya Ronnie Effendie. "Belajar Bahasa Pemrograman R." (2018).
- [25] Pudjajana, Andre Maureen, and Danny Manongga. "Sentimen Analisis Tweet Pornografi Kaum Homoseksual Indonesia Di Twitter dengan Naive Bayes." *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer* 9.1 (2018): 313-318.

- [26] Alamsah, Anggah Suhendra. Implementasi Sistem Temu Kembali Informasi Untuk Pencarian Buku Pada Toko Buku Online Menggunakan Metode Vector Space Model. Diss. Universitas Muhammadiyah Gresik, 2018