

PROGRAM STUDI SISTEM INFORMASI

**ANALISIS SENTIMEN TERHADAP VIRUS CORONA
BERDASARKAN OPINI MASYARAKAT BERBASIS WEB**



**ALPIA ILWANASARI
171410257**

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana
Komputer di Universitas Bina Darma**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINADARMA
PALEMBANG
2022**



**ANALISIS SENTIMEN TERHADAP VIRUS CORONA
BERDASARKAN OPINI MASYARAKAT BERBASIS WEB**

SKRIPSI

**ALPIA ILWANASARI
171410257**

**Skripsi ini diajukan sebagai syarat memperoleh gelar Sarjana
Komputer di Universitas Bina Darma**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER
UNIVERSITAS BINADARMA
PALEMBANG
2022**

HALAMAN PENGESAHAN

Analisis Sentimen Terhadap Virus Corona Berdasarkan Opini Masyarakat Berbasis Web

ALPIA JLWANASARI

171410257

Telah diterima sebagai salah satu syarat untuk memperoleh gelar
Sarjana Komputer pada Program Studi Sistem Informasi

Palembang, 10 Maret 2022

Fakultas Ilmu Komputer

Universitas Bina Darma

Dekan,

Pembimbing

A. Haidar Mirza, S.T., M.Kom.



Dedy Syamsuar, S.Kom., M.I.T., Ph.D.

HALAMAN PERSETUJUAN

Skripsi Berjudul "Analisis Sentimen Terhadap Virus Corona Berdasarkan Opini Masyarakat Berbasis Web" Oleh "Alpia Ilwanasari", telah dipertahankan di depan komisi penguji pada hari Kamis tanggal 10 Maret 2022.

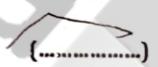
Komisi Penguji

1. Ketua : A. Haidar Mizra, S.T., M.Kom.



(.....)

2. Anggota : Linda Atika, M.Kom



(.....)

3. Anggota : Novri Hadinata, M.Kom.



(.....)

Mengetahui,
Program Studi Sistem Informasi
Fakultas Ilmu Komputer
Universitas Bina Darma
Ketua,



Zaid Amin, M.Kom., Ph.D.

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama : Alpia Ilwanasari

NIM : 171410257

Dengan ini menyatakan bahwa:

1. Karya tulis saya (Skripsi) adalah asli dan belum pernah diajukan untuk mendapatkan gelar akademik (Sarjana) di Universitas Bina Darma atau perguruan tinggi lainnya ;
2. Karya tulis ini murni gagasan, rumusan dan penelitian saya dengan arahan dari tim pembimbing ;
3. Di dalam karya tulis ini tidak terdapat karya atau pendapat yang telah ditulis atau di publikasikan orang lain, kecuali secara tertulis dengan jelas dikutip dengan mencantumkan nama pengarang dan memasukkan ke dalam daftar rujukan ;
4. Saya bersedia tugas skripsi, di cek keasliannya menggunakan plagiarism checker serta di unggah ke internet, sehingga dapat diakses secara daring ;
5. Surat pernyataan ini saya tulis dengan sungguh-sungguh dan apabila terbukti melakukan penyimpangan atau ketidakbenaran dalam pernyataan ini maka saya bersedia menerima sanksi dengan peraturan dan perundang-undangan yang berlaku ;

Demikian surat pernyataan ini saya buat agar dapat dipergunakan sebagaimana mestinya.

Palembang, Maret 2022

Yang membuat pernyataan,



ALPIA ILWANASARI

NIM: 171410257

MOTTO DAN PERSEMBAHAN

Motto:

- ❖ **Do the best and pray, good will take care of the rest.** (Lakukan yang terbaik, kemudian berdoalah. Tuhan yang akan mengurus sisanya.)
- ❖ Allah tidak membebani seseorang itu melainkan sesuai kesanggupannya (QS. Al Baqarah:286).
- ❖ Sesungguhnya Allah tidak akan merubah keadaan suatu kaum sehingga mereka merubah keadaan yang ada pada diri mereka sendiri (QS. Ar Rad: 11).
- ❖ Dan hanya kepada Tuhanmulah hendaknya kamu berharap (QS. Al Insyirah: 8).
- ❖ Kamu tak bisa apa-apa tanpa Allah, TAPI kamu bisa meraih segalanya dengan izin Allah SWT.

Persemaahan:

Skripsi ini kupersembahkan untuk:

- ✚ Kedua Orang tuaku tercinta Bapak Supratman (Alm) dan Ibu Nurlaila yang selalu memberikan doa, nasehat, kasih sayang serta dukungan baik moral maupun material untuk anaknya.
- ✚ Saudara-saudaraku tersayang yang telah memberikan dukungan dan semangat untuk terus berusaha.
- ✚ Dosen pembimbingku Bapak A.Haidar Mirza, S.T., M.Kom, terima kasih atas bimbingan, nasehat, arahan dan bantuannya selama ini.
- ✚ Fiki Kurniawan, S.Kom, terima Kasih karena telah banyak melakukan hal-hal luar biasa bagi saya hingga sampai di titik yang baik ini.
- ✚ Sahabat terbaikku Asniarti & Amelia Tya Miranti yang selalu ada untuk terus memberikan semangat agar bisa menyelesaikan Skripsi ini dengan baik.
- ✚ Almamaterku

Abstract

Corona viruses are a large family of viruses that can infect humans and animals that usually cause respiratory tract infections, ranging from flu to serious illnesses such as MERS (Middle-East respiratory syndrome) which first appeared in the Middle East in 2012 and SARS (Severe Acute Respiratory Syndrome). Syndrome) was first epidemic in China in 2002. (Hanati, 2020). Coronavirus Disease 2019 abbreviated as COVID-19 is an infectious disease caused by SARS-CoV-2. Basically, sentiment analysis has two types of approaches, namely machine learning and lexicon based. Several machine learning classification methods such as nave Bayes, support vector machines, and lexicon-based are often used to get the best results. This study aims to find out how the impact of the corona virus in Indonesia is according to public opinion. data collection is done by using the Crawling technique on the Twitter API key. The data used in this research is Twitter user post data found on the Twitter.com site. The data collected is in the form of text data taken using a crawling technique using the Twitter API Key. crawling data, using the Twitter API Key, data is obtained with a total of 2501 Tweet records, provided that the training data for the tweet document is 80% and the training data for the tweet document is 20%. The scraped data is then analyzed. The results can be seen in the precision and recall values. from each class it can be seen that the level of system processing ability in finding the level of accuracy between the information desired by the user as a positive class is "0.88%", and for a negative class it is "0.00%". The success rate of system processing in retrieving positive class information is "1.00%", for negative class it is "0.00%".

Keywords: Sentiment analysis, Corona, Nave bayes, KDD

Abstrak

Virus Corona merupakan keluarga besar virus yang dapat menyerang manusia dan hewan yang biasanya menyebabkan penyakit infeksi saluran pernapasan, mulai dari flu hingga penyakit serius seperti MERS (*Middle-East respiratory syndrome*) pertama kali muncul di Timur Tengah pada tahun 2012 dan SARS (*Severe Acute Respiratory Syndrome*) pertama kali mewabah di China pada tahun 2002. (Hanati, 2020). Coronavirus Disease 2019 disingkat COVID-19 adalah penyakit menular yang disebabkan oleh SARS-CoV-2. Pada dasarnya sentimen analisis memiliki dua tipe pendekatan yaitu machine learning dan lexicon based. Beberapa metode klasifikasi machine learning seperti naïve bayes, support vector machine, dan lexicon-based sering digunakan untuk mendapatkan hasil yang terbaik. Penelitian ini bertujuan untuk mengetahui bagaimana dampak virus corona di indonesia sesuai opini masyarakat. Pengumpulan data dilakukan dengan teknik *Crawling* pada *API key twitter*. Data yang digunakan dalam penelitian ini adalah data postingan pengguna *Twitter* yang terdapat pada situs *Twitter.com*. Data yang dikumpulkan berupa data teks yang diambil menggunakan teknik *crawling* menggunakan *API Key Twitter*. *crawling* data, menggunakan *API Key Twitter* maka didapatkan data dengan jumlah 2501 *record Tweet*, dengan ketentuan data latih dokumen *tweet* sebesar 80% dan data *training* dokumen *tweet* sebesar 20%. Data yang telah di scraping kemudian dianalisis Hasil dapat dilihat nilai *precision*, dan *recall* dari setiap kelas dapat dilihat tingkat kemampuan pemrosesan sistem dalam mencari tingkat ketepatan antara informasi yang diinginkan oleh pengguna sebagai kelas positif adalah “0.88%”, dan untuk kelas *negatif* adalah “0.00%”. Tingkat keberhasilan dari pemrosesan sistem dalam memperoleh kembali informasi kelas positif adalah “1.00%”, untuk kelas *negatif* adalah “0.00%”.

Kata Kunci: *Analisis sentiment, Corona, Naïve bayes, KDD*

KATA PENGANTAR



Puji dan syukur penulis panjatkan kehadirat Tuhan Yang Maha Esa atas berkat dan anugerahnya, sehingga penulis dapat menyelesaikan skripsi berjudul "Analisis Sentimen Terhadap Virus Corona Berdasarkan Opini Masyarakat Berbasis WEB". Skripsi ini disusun sebagai salah satu syarat menyelesaikan program studi Sistem Informasi di Universitas Binadarma Palembang.

Keberhasilan penulis dalam menyusun skripsi ini tidak terlepas atas bantuan dan dorongan dari berbagai pihak. Dengan segala kerendahan hati penulis menyampaikan terimakasih kepada:

1. Kedua Orang tuaku tercinta Bapak Supratman (Alm) dan Ibu Nurlaila yang selalu memberikan doa, nasehat, kasih sayang serta dukungan baik moral maupun material & terima kasih juga untuk saudara-saudara tersayang yang selalu memberikan suport.
2. Ibu Dr. Sunda Ariana,M.Pd.,M.M, selaku Rektor Universitas Binadarma Palembang.
3. Bapak Dedy Syamsuar,M.IT.,Ph.D, selaku Dekan Fakultas Ilmu Komputer Universitas Binadarma Palembang.
4. Bapak Zaid Amin, M.Kom.,Ph.D selaku Ketua Program Studi Sistem Informasi Universitas Binadarma Palembang.
5. Bapak A. Haidar Mirza, S.T., M.Kom, selaku dosen pembimbing yang telah memberikan petunjuk dan saran sehingga skripsi ini terselesaikan dengan baik.
6. Fiki Kurniawan, S.Kom., Asniarti, Amelia Tya M, terimakasih atas doa dan dukungan untuk keberhasilan penulis dalam menyelesaikan laporan skripsi ini.

7. Teman teman seperjuangan terutama kelas SII dan Semua pihak yang telah memberikan semangat luar biasa dan membantu dalam penelitian ini.
8. Penulis menyadari skripsi ini tidak luput dari berbagai kekurangan. Penulis mengharapkan saran dan kritik demi kesempurnaan dan perbaikannya sehingga akhirnya laporan skripsi ini dapat memberikan manfaat bagi bidang pendidikan dan bisa dikembangkan lebih lanjut lagi dan dapat bermanfaat bagi semua pihak.

Terima kasih.

Palembang, Maret 2022



Penulis

DAFTAR ISI

HALAMAN PENGESAHAN.....	i
HALAMAN PERSETUJUAN	ii
SURAT PERNYATAAN.....	iii
Abstract.....	v
Abstrak	vi
KATA PENGANTAR	vii
DAFTAR ISI.....	ix
DAFTAR GAMBAR.....	xi
DAFTAR TABLE	xii
BAB I.....	1
PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	4
1.4 Tujuan dan Manfaat Penelitian	4
1.4.2 Manfaat Penelitian	4
1.5 Luaran	5
1.6 Sistematika Penulisan	5
BAB II.....	6
TINJAUAN PUSTAKA.....	6
2.1 Landasan Teori	6
2.1.1 Text mining.....	6
2.1.2 Analisis Sentimen	6
2.1.3 Virus Corona.....	8
2.1.4 Website.....	9
2.1.5 Naïve bayes.....	9
2.1.6 Confusion Matrix.....	14
2.1.7 Knowledge Discovery in Database (KDD)	15
BAB III.....	17
METODOLOGI PENELITIAN.....	17
3.1 (Data Selection) Pengumpulan Data	17

3.1.1 Crawling Twitter.....	18
3.1.2 Pelabelan	20
3.2 Preprocessing	21
3.2.1 Tahapan Proses <i>Preprocessing</i>	21
3.3 Transformation	24
3.3.1 Ekstraksi Fitur	25
3.3.2 Klasifikasi.....	25
3.3.3 Uji Model.....	25
3.3.4 Data Mining	26
3.3.5 <i>Evaluation (Evaluasi)</i>	26
BAB IV	27
HASIL DAN PEMBAHASAN	27
4.1 Hasil.....	27
4.1.1 Crawling dengan Phyton	27
4.1.2 Preprocessing dengan Phyton	29
4.1.3 Algoritma Naïve Bayes Classifier	33
4.1.4 Ekstraksi Fitur dengan Phyton.....	36
4.2 Pembahasan.....	38
4.2.1 mplementasi Klasifikasi <i>Naïve Bayes Pada Phyton</i>	38
4.2.2 Uji Model	40
BAB V	44
KESIMPULAN DAN SARAN	44
5.1 Kesimpulan	44
5.2 Saran.....	44
DAFTAR PUSTAKA	

DAFTAR GAMBAR

Gambar 2. 1 Skema Naive Bayes	14
Gambar 3. 1 Tahapan Peneltian KDD	17
Gambar 3. 2 Halaman Awal Twitter	18
Gambar 3. 3 Developer Twitter.....	18
Gambar 3. 4 API Twitter	19
Gambar 3. 5 Hasil Crawling Data Twitter.....	19
Gambar 4. 1 Jupyter Notebook	27
Gambar 4. 2 Source Code Pemanggilan Python Library Proses Crawling	27
Gambar 4. 3 Source Code Proses Crawling.....	28
Gambar 4. 4 Excel Hasil Crawling	28
Gambar 4. 5 Excel Hasil Crawling dan Labelling	29
Gambar 4. 6 kode program proses cleaning.....	30
Gambar 4. 7 Deklarasi library nltk.....	31
Gambar 4. 8 Kode Program Proses Remove Stopword.....	31
Gambar 4. 9 Kode Program Proses Tokenization.....	32
Gambar 4. 10 Pendeklarasian library sastrawi.....	32
Gambar 4. 11 Kode Program Proses Steamming	33
Gambar 4. 12 Fungsi pendeklarasian library yang digunakan	39
Gambar 4. 13 Fungsi memanggil data set	39
Gambar 4. 14 Proses Pengimplementasian Class Pipeline	39
Gambar 4. 16 Proses Menghitung dari Nilai Presisi, Recall dan F-1 score	41
Gambar 4. 17 Hasil Uji Model	41

DAFTAR TABLE

Tabel 2. 1 Model Confusion Matrix	15
Tabel 3. 1 Data Hasil Labelling	20
Tabel 3. 2 Contoh Data Hasil Cleaning	21
Tabel 3. 3 Contoh Data Hasil Remove Stopword.....	22
Tabel 3. 4 Contoh Data Hasil Tokenization	23
Tabel 3. 5 Contoh Data Hasil Stemming.....	24
Tabel 4. 1 Pembuatan Word Vector	36
Tabel 4. 2 Proses Perhitungan TF (Term Frequency).....	37
Tabel 4. 3 Proses Perhitungan DF (Document Frequency)	37
Tabel 4. 4 Proses IDF (Inverse Document Frequency)	37
Tabel 4. 5 Contoh Proses Perhitungan TF-IDF	38
Tabel 4. 6 Contoh Word Vector yang sudah dibobotkan	38
Tabel 4. 7 Model Confusion Matrix	40
Tabel 4. 8 Hasil Confusion Matrix.....	40
Tabel 4. 9 Hasil dari Nilai Precision, Recall, dan F-1 score.....	42

DAFTAR LAMPIRAN

- 1. Form ferifikasi penjilidan skripsi**
- 2. Surat keterangan lulus**
- 3. SK pembimbing**
- 4. Formulir permohonan pengajuan judul skripsi**
- 5. Lembar konsultasi skripsi**
- 6. Formulir perbaikan komprehensif**
- 7. Hasil turnitin**