

**KLASIFIKASI SENTIMEN TERHADAP DATA TEXT JEJARING SOSIAL
DENGAN TOPIK VAKSIN COVID-19 MENGGUNAKAN
K-NEAREST NEIGHBOR (KNN)**

TUGAS AKHIR

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer**



Oleh

**M. Rizki Ronaldo
09011381621110**

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2021**

**KLASIFIKASI SENTIMEN TERHADAP DATA TEXT JEJARING SOSIAL
DENGAN TOPIK VAKSIN COVID-19 MENGGUNAKAN
K-NEAREST NEIGHBOR (KNN)**

TUGAS AKHIR

**Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer**



Oleh

**M. Rizki Ronaldo
09011381621110**

**JURUSAN SISTEM KOMPUTER
FAKULTAS ILMU KOMPUTER
UNIVERSITAS SRIWIJAYA
2021**

LEMBAR PENGESAHAN

**KLASIFIKASI SENTIMEN TERHADAP DATA TEXT JEJARING SOSIAL
DENGAN TOPIK VAKSIN COVID-19 MENGGUNAKAN
K-NEAREST NEIGHBOR (KNN)**

TUGAS AKHIR

**Program Studi Sistem Komputer
Jenjang S1**

Oleh

**M. Rizki Renaldi
09011381621110**

Indralaya, Juli 2021

**Mengetahui,
Ketua Jurusan Sistem Komputer**

Pembimbing Tugas Akhir



**Dr. Ir. H. Sukemi, M.T.
NIP. 196612032006041001**

**Huda Ubaya, S.T., M.T.
NIP. 198106162012121003**

HALAMAN PERSETUJUAN

Telah diuji dan lulus pada :

Hari Jum'at

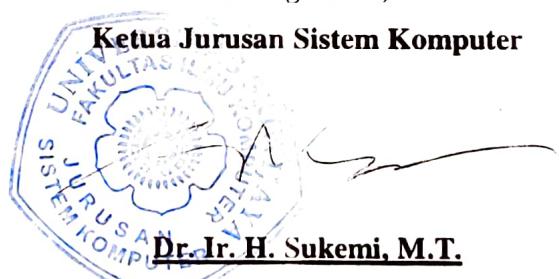
Tanggal : 2 Juli 2021

Tim Penguji :

1. Ketua : **Ahmad Zarkasih, S.T., M.T.**
2. Sekretaris : **Aditya Putra P Prasetyo, S.Kom., M.T.**
3. Pembimbing : **Huda Ubaya, S.T., M.T.**
4. Penguji : **Ahmad Fali Oklilas, S.T., M.T.**

Mengetahui, 68/1

Ketua Jurusan Sistem Komputer



Dr. Ir. H. Sukemi, M.T.

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : M. Rizki Ronaldo
NIM : 09011381621110
Judul : Klasifikasi Sentimen Terhadap *Data Text* Jejaring Sosial Dengan Topik Vaksin Covid-19 Menggunakan *K-Nearest Neighbor (KNN)*

Hasil Pengecekan Software iThenticate/Turnitin : 5 %

Menyatakan bahwa laporan tugas akhir saya merupakan hasil karya sendiri dan bukan hasil penjiplakan atau plagiat. Apabila ditemukan unsur penjiplakan atau plagiat dalam laporan tugas akhir ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya.

Demikian, pernyataan ini saya buat dalam keadaan sadar dan tanpa paksaan dari siapapun.



Indralaya, Juli 2021



M. Rizki Ronaldo
NIM. 09011381621110

HALAMAN PERSEMBAHAN

***“Salah satu hal terbaik dalam hidup adalah melihat senyum
di wajah orang tuamu dan menyadari bahwa kamu lahir alasannya”***

Sebagai tanda bakti, hormat dan rasa terima kasih yang tak terhingga, saya persembahkan karya ini untuk orang yang sangat saya cintai dan saya sayangi Ibunda (Mama) Tutiana, S.Pd. dan Ayahanda (Papa) Bakhtiar yang selalu memberikan dukungan moril maupun materil

Yang selalu memanjatkan doa dimanapun, kapanpun dan tiada henti, serta cinta kasih untuk putramu ini yang tiada mungkin dapat saya balas hanya dengan selembar kertas persembahan ini.

Semoga ini menjadi langkah awal putramu ini untuk membuat ibunda dan ayahanda selalu tersenyum bahagia.

Atas Ridho Allah SWT, Skripsi ini saya persembahkan untuk:

1. Kedua orang tua tercinta, Ibunda Tutiana dan Ayahanda Bakhtiar.
2. Saudara-saudaraku tercinta Anggun Febi Parizka, S.Pd., M.Pd. dan M. Jimmy Alvindo.
3. Dosen dan pegawai Fasilkom Unsri
4. Teman-teman seperjuangan Jurusan Sistem Komputer 2016
5. Almamater Universitas Sriwijaya

KATA PENGANTAR

Assalamualaikum Warahmatullahi Wabarakatuh,

Puji dan syukur penulis panjatkan kepada Allah Subhanahu Wa Ta’ala, atas segala karunia, berkat, dan rahmat-Nya sehingga penulis dapat menyelesaikan penyusunan Penelitian Tugas Akhir ini dengan judul “Klasifikasi Sentimen Terhadap *Data Text* Jejaring Sosial Dengan Topik Vaksin Covid-19 Menggunakan *K-Nearest Neighbor (KNN)*”.

Dalam laporan ini penulis menjelaskan mengenai bagaimana cara mengklasifikasi *Sentimen* dengan menggunakan *K-Nearest Neighbor (KNN)* disertai data-data yang diperoleh penulis saat melakukan pengujian. Penulis berharap tulisan ini dapat bermanfaat bagi orang banyak, dan menjadi tambahan bahan bacaan dan referensi untuk para akademisi dan peneliti lain yang juga berada atau sedang menekuni bidang *data text*, khususnya vaksin covid.

Pada penyusunan laporan ini, penulis banyak mendapatkan ide dan saran serta bantuan dari berbagai pihak baik secara langsung maupun tidak langsung. Oleh karena itu, pada kesempatan ini penulis mengucapkan rasa syukur kepada Allah Subhanahu Wa Ta’ala dan terimakasih kepada yang terhormat:

1. Kedua orang tua beserta keluarga yang selalu mendoakan, memberikan dukungan baik moril maupun materil, serta motivasi dan semangat selama hidup penulis.
2. Bapak Jaidan Jauhari, S.Pd. M.T selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya
3. Bapak Dr. Ir. Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya
4. Bapak Huda Ubaya, M.T. selaku Pembimbing Tugas Akhir.
5. Bapak Dr. Ir. Sukemi, M.T. selaku Pembimbing Akademik Jurusan Sistem Komputer.

6. Jerry Pratama Saputra dan Ilham Solehan sebagai partner dan teman dari awal penggerjaan tugas akhir, baik dalam pencarian literatur, pemrograman, serta diskusi mengenai riset dan lainnya.
7. Teman - teman dari Jurusan Sistem Komputer yang tidak bisa disebutkan satu-persatu. Khususnya seluruh teman-teman dari kelas SK16B Indralaya.
8. Zakiah, S.Ap. yang telah menemani dan mendukung pembuatan penelitian ini.
9. Dan semua pihak yang telah membantu.

Penulis menyadari bahwa laporan ini masih jauh dari kesempurnaan, oleh karena itu penulis sangat mengharapkan kritik dan saran yang bersifat membangun agar lebih baik lagi dikemudian hari.

Akhir kata dengan segala keterbatasan, penulis berharap semoga laporan ini menghasilkan sesuatu yang bermanfaat bagi kita semua khususnya bagi mahasiswa Fakultas Ilmu Komputer Universitas Sriwijaya secara langsung ataupun tidak langsung sebagai sumbangan pemikiran dalam peningkatan mutu pembelajaran dan penelitian.

Indralaya, Juli 2021

Penulis

**KLASIFIKASI SENTIMEN TERHADAP DATA TEXT JEJARING SOSIAL
DENGAN TOPIK VAKSIN COVID-19 MENGGUNAKAN
K-NEAREST NEIGHBOR (KNN)**

M. Rizki Ronaldo (09011381621110)
Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya
Email : mrizkironaldo.rnl@gmail.com

ABSTRAK

Vaksin covid-19 adalah sebuah zat atau produk yang cara kerjanya dimasukkan kedalam tubuh manusia dan akan terjadi stimulasi sistem imun yang akhirnya bisa melakukan proteksi serta melakukan perlindungan terhadap virus covid-19. Pada bulan agustus 2020 pemerintah mengumumkan akan mendistribusikan vaksin tersebut, kebijakan ini memicu banyak reaksi publik. Masyarakat ingin menyampaikan segala pendapat, kritik dan saran. Dalam hal ini jejaring sosial merupakan tempat yang cocok untuk mengungkapkan hal tersebut, dan Twitter adalah salah satu platform yang tepat. Penelitian ini, melakukan klasifikasi sentimen dengan *Natural Language Processing* (NLP) terhadap data teks jejaring sosial. Metode *K-Nearest Neighbor* diusulkan karena mampu mengklasifikasikan sentimen terhadap data vaksin covid-19. Data yang digunakan diperoleh dari *Drone Emprit Academy*. Hasil dari klasifikasi menggunakan KNN di evaluasi menggunakan *confusion matrix*. Klasifikasi KNN dengan perbandingan *data training* 70% dan *data testing* 30% menghasilkan nilai tertinggi dengan akurasi 87,35%, presisi 88,65%, *recall* 85,92%, dan *f1-score* 86,73%, sedangkan *data training* 20% dan *data testing* menghasilkan nilai terendah dengan akurasi 81,89%, presisi 80,96%, *recall* 80,99%, dan *f1-score* 80,94%. Penelitian ini menunjukan bahwa algoritma yang dirancang mampu melakukan klasifikasi sentimen dengan baik, dan hasil perhitungan dapat dijadikan sebagai pertimbangan pemerintah dalam mengambil keputusan mengenai vaksin covid-19.

Kata Kunci : Vaksin Covid-19, *Natural Language Precessing*, *K-Nearest Neighbor*, sentimen, data text, *confusion matrix*.

Ketua Jurusan Sistem Komputer

 6/8/21



Dr. Ir. H. Sukemi, M.T.

NIP. 196612032006041001

Indralaya, Juli 2021

Mengetahui,

Pembimbing Tugas Akhir



Huda Ubaya, S.T., M.T.

NIP. 198106162012121003

**CLASSIFICATION OF SENTIMENTS ON SOCIAL NETWORK TEXT DATA
WITH THE TOPIC OF COVID-19 VACCINE
USING K-NEAREST NEIGHBOR (KNN)**

M. Rizki Ronaldo (09011381621110)

Dept. of Computer Engineering, Faculty of Computer Science, Sriwijaya

University

Email : mrizkironaldo.rnl@gmail.com

ABSTRACT

The covid-19 vaccine is a substance or product whose working method is inserted into the human body and there will be stimulation of the immune system which can finally protect and protect against the covid-19 virus. In August 2020 the government announced it would distribute the vaccine, this policy sparked a lot of public backlash. The public wants to express all opinions, criticisms and suggestions. In this case, social networks are a suitable place to express this, and Twitter is one of the right platforms. In this study, sentiment classification using Natural Language Processing (NLP) was carried out on social network text data. The K-Nearest Neighbor method was proposed because it was able to classify sentiment on the COVID-19 vaccine data. The data used was obtained from the Drone Emprit Academy. The results of the classification using KNN are evaluated using a confusion matrix. The KNN classification with a comparison of 70% training data and 30% testing data resulted in the highest score with 87.35% accuracy, 88.65% precision, 85.92% recall, and 86.73% f1-score, while the training data was 20% and data testing produces the lowest score with 81.89% accuracy, 80.96% precision, 80.99% recall, and 80.94% f1-score. This study shows that the algorithm designed is able to classify sentiment well, and the results of the calculations can be used as a consideration for the government in making decisions regarding the covid-19 vaccine.

Keyword : Covid-19 vaccine, Natural Language Precessing, K-Nearest Neighbor, sentiment, data text, confusion matrix.

Head of Department Computer Engineering



Dr. Ir. H. Sukemi, M.T.

NIP. 196612032006041001

6/8/21

Indralaya, July 2021

Supervisor



Huda Ubaya, S.T., M.T.

NIP. 198106162012121003

DAFTAR ISI

	Halaman
Halaman Judul	i
Lembar Pengesahan	ii
Halaman Persetujuan	iii
Halaman Pernyataan	iv
Halaman Persembahan.....	v
Kata Pengantar	vi
Abstrak.....	viii
Abstarc.....	ix
Daftar Isi	x
Daftar Gambar	xiv
Daftar Tabel	xvii
Daftar lampiran.....	xix
Bab I. Pendahuluan	1
1.1 Latar Belakang	1
1.2. Tujuan	2
1.3. Manfaat	2
1.4. Perumusan Masalah	3
1.5 Batasan Masalah	3
1.6 Metodologi Penelitian	3
1.7 Sistematika Penulisan	5

Bab II. Tinjauan Pustaka	6
2.1. Analisis Sentimen.....	6
2.2. Twitter.....	6
2.2.1. API Twitter	7
2.2.2. Vaksin Covid-19.....	7
2.3. <i>Natural Language Processing</i> (NLP).....	8
2.4. <i>Machine Learning</i>	9
2.5. <i>Text Mining</i>	9
2.6. <i>Preprocessing</i>	10
2.6.1. Cleaning	10
2.6.2. Case Folding.....	10
2.6.3. Tokenizing	10
2.6.4. Stemming.....	10
2.7. Ekstraksi fitur (<i>Feature Extraction</i>).....	11
2.8. Pemilihan fitur (<i>Feature Selection</i>).....	11
2.9. <i>Algoritma</i>	12
2.9.1. Mempersiapkan Algorima	12
2.10. <i>K-Nearrest Neighbor</i> (KNN).....	13
2.11. <i>K – Fold Cross Validation</i>	15
2.12. <i>Confusion Matrix</i>	15
Bab III. Metodologi Penelitian	19
3.1. Pendahuluan	19
3.2. Kerangka Kerja	19

3.3. Pengambilan Data	21
3.4. <i>Preprocessing</i>	23
3.4.1. <i>Cleaning</i>	24
3.4.2. <i>Case Folding</i>	25
3.4.3. <i>Tokenizing</i>	26
3.4.4. <i>Stemming</i>	26
3.5. Pemisahan Feature dan Label	29
3.6. Ekstraksi Fitur.....	29
3.7. Evaluasi dan validasi.....	30
3.8. <i>K – Fold Cross Validation</i>	31
3.9. <i>Confusion Matrix</i>	31
Bab IV. Pengujian dan Analisa.....	33
4.1. Pendahuluan.....	33
4.2. Data Text.....	33
4.3. Hasil Preprocessing.....	33
4.3.1. Input.....	34
4.3.2. Proses Cleaning.....	34
4.3.3. Proses Tokenizing.....	35
4.3.4. Proses Case Folding & Stemming.....	36
4.4. Pengujian Ekstraksi Fitur Data Text.....	36
4.5. Hasil dan Analisa Menggunakan Metode KNN.....	37
4.5.1. Pengujian dengan Nilai Cross Validaton (CV = 2)... ..	37
4.5.2. Pengujian dengan Nilai Cross Validaton (CV = 5)... ..	39

4.5.3. Pengujian dengan Nilai Cross Validaton (CV = 10)...	40
4.5.4. Pengujian dengan Nilai Cross Validaton (CV = 20)...	41
4.6. Perbandingan Nilai Hasil Cross Validation (CV)	43
4.7. Pengujian dengan Perbedaan Data Training dan Data Testing...45	
4.7.1. Pengujian dengan Perbandingan 80% dan 20%.....	45
4.7.2. Analisis dari Pengujian perbandingan 80% : 20%.....	46
4.7.3. Pengujian dengan Perbandingan 70% dan 30%.....	48
4.7.4. Analisis dari Pengujian perbandingan 70% : 30%.....	50
4.7.5. Pengujian dengan Perbandingan 60% dan 40%.....	52
4.7.6. Analisis dari Pengujian perbandingan 60% : 40%.....	54
4.7.7. Pengujian dengan Perbandingan 50% dan 50%.....	56
4.7.8. Analisis dari Pengujian perbandingan 50% : 50%.....	58
4.7.9. Pengujian dengan Perbandingan 40% dan 60%.....	60
4.7.10. Analisis dari Pengujian perbandingan 40% : 60%....	62
4.7.11. Pengujian dengan Perbandingan 30% dan 70%.....	64
4.7.12. Analisis dari Pengujian perbandingan 30% : 70%....	66
4.7.13. Pengujian dengan Perbandingan 20% dan 80%.....	68
4.7.14. Analisis dari Pengujian perbandingan 40% : 60%....	70
4.8. Analisa Hasil dari Perbedaan Data Training dan Data Testing..	72
Bab V. Kesimpulan	76
5.1 Kesimpulan	76
5.2 Saran.....	76
Daftar Pustaka	77

DAFTAR GAMBAR

	Halaman
Gambar 3.1 Kerangka Kerja Penelitian	20
Gambar 3.2 Kumpulan Data.....	21
Gambar 3.3 Wordcloud sentimen negatif.....	22
Gambar 3.4 Wordcloud sentimen netral.....	22
Gambar 3.5 Wordcloud sentimen positif	23
Gambar 3.6 Kerangka Kerja Cleanning	24
Gambar 3.7 Proses Cleanning	24
Gambar 3.8 Kerangka Kerja Case Folding	25
Gambar 3.9 Proses Case Folding	25
Gambar 3.10 Kerangka Kerja Tokenizing	26
Gambar 3.11 Proses Tokenizing	26
Gambar 3.12 Kerangka Kerja Stemming	27
Gambar 3.13 Kerangka Kerja Stemming lanjutan.....	28
Gambar 3.14 Proses Stemming	28
Gambar 3.15 Jumlah data keseluruhan.....	29
Gambar 3.16 Kata dalam proses ekstraksi fitur	30
Gambar 3.17 Flowchart evaluasi dan validasi.....	31
Gambar 4.1 Data dalam format .xlsx	34
Gambar 4.2 Hasil Proses Cleaning	35

Gambar 4.3 Hasil Proses Tokenizing	35
Gambar 4.4 Hasil Proses Case Folding & Stemming	36
Gambar 4.5 Sampel Data Ekstraksi Fitur 200 Sampai 215	37
Gambar 4.6 Sampel Data Ekstraksi Fitur 1000 Sampai 1015.....	37
Gambar 4.7 Hasil pengujian dengan Nilai Cross Validaton (CV = 2)	38
Gambar 4.8 Hasil pengujian dengan Nilai Cross Validaton (CV = 5)	39
Gambar 4.9 Hasil pengujian dengan Nilai Cross Validaton (CV = 10)	41
Gambar 4.10 Hasil pengujian dengan Nilai Cross Validaton (CV = 20)	42
Gambar 4.11 Grafik transisi nilai dari pengujian dengan parameter Accuracy	43
Gambar 4.12 Grafik dari pengujian dengan parameter presisi	44
Gambar 4.13 Grafik dari pengujian dengan parameter f1-score.....	44
Gambar 4.14 Split data training 80% dan data testing 20%.....	45
Gambar 4.15 Hasil dari pengujian dengan perbandingan 80:20.....	46
Gambar 4.16 Split data training 70% dan data testing 30%.....	49
Gambar 4.17 Hasil dari pengujian dengan perbandingan 70:30.....	50
Gambar 4.18 Split data training 60% dan data testing 40%.....	53
Gambar 4.19 Hasil dari pengujian dengan perbandingan 60:40.....	54
Gambar 4.20 Split data training 50% dan data testing 50%.....	57
Gambar 4.21 Hasil dari pengujian dengan perbandingan 50:50.....	58
Gambar 4.22 Split data training 40% dan data testing 60%.....	61
Gambar 4.23 Hasil dari pengujian dengan perbandingan 40:60.....	62
Gambar 4.24 Split data training 30% dan data testing 70%.....	65

Gambar 4.25 Hasil dari pengujian dengan perbandingan 30:70.....	66
Gambar 4.26 Split data training 20% dan data testing 80%.....	69
Gambar 4.27 Hasil dari pengujian dengan perbandingan 20:80.....	70
Gambar 4.28 Grafik perbandingan hasil dengan parameter <i>Accuracy</i>	73
Gambar 4.29 Grafik perbandingan hasil dengan parameter <i>precision</i>	73
Gambar 4.30 Grafik perbandingan hasil dengan parameter <i>recall</i>	74
Gambar 4.31 Grafik perbandingan hasil dengan parameter <i>f1-score</i>	74

DAFTAR TABEL

	Halaman
Tabel 2.1 Confusion matrix 2x2	15
Tabel 2.2 Confusion matrix 3x3	17
Tabel 3.1 Confusion matrix 3x3	32
Tabel 3.2 Persamaan matematis nilai <i>precision</i>	32
Tabel 3.3 Persamaan matematis nilai <i>recall</i>	32
Tabel 4.1 Perbandingan pengujian pada nilai <i>Cross Validation</i> yang di uji.....	43
Tabel 4.2 Hasil perhitungan <i>precision</i> dengan perbandingan 80% dan 20%	47
Tabel 4.3 Hasil perhitungan <i>recall</i> dengan perbandingan 80% dan 20%	48
Tabel 4.4 Hasil perhitungan <i>f1-score</i> dengan perbandingan 80% dan 20%	48
Tabel 4.5 Hasil perhitungan <i>precision</i> dengan perbandingan 70% dan 30%	51
Tabel 4.6 Hasil perhitungan <i>recall</i> dengan perbandingan 70% dan 30%	51
Tabel 4.7 Hasil perhitungan <i>f1-score</i> dengan perbandingan 70% dan 30%	52
Tabel 4.8 Hasil perhitungan <i>precision</i> dengan perbandingan 60% dan 40%	55
Tabel 4.9 Hasil perhitungan <i>recall</i> dengan perbandingan 60% dan 40%	55
Tabel 4.10 Hasil perhitungan <i>f1-score</i> dengan perbandingan 60% dan 40%	56
Tabel 4.11 Hasil perhitungan <i>precision</i> dengan perbandingan 50% dan 50%.....	59
Tabel 4.12 Hasil perhitungan <i>recall</i> dengan perbandingan 50% dan 50%	59
Tabel 4.13 Hasil perhitungan <i>f1-score</i> dengan perbandingan 50% dan 50%	60
Tabel 4.14 Hasil perhitungan <i>precision</i> dengan perbandingan 40% dan 60%....	63
Tabel 4.15 Hasil perhitungan <i>recall</i> dengan perbandingan 40% dan 60%	63

Tabel 4.16 Hasil perhitungan <i>f1-score</i> dengan perbandingan 40% dan 60%	64
Tabel 4.17 Hasil perhitungan <i>precision</i> dengan perbandingan 30% dan 70%.....	67
Tabel 4.18 Hasil perhitungan <i>recall</i> dengan perbandingan 30% dan 70%	67
Tabel 4.19 Hasil perhitungan <i>f1-score</i> dengan perbandingan 30% dan 70%	68
Tabel 4.20 Hasil perhitungan <i>precision</i> dengan perbandingan 20% dan 80%.....	71
Tabel 4.21 Hasil perhitungan <i>recall</i> dengan perbandingan 20% dan 80%	71
Tabel 4.22 Hasil perhitungan <i>f1-score</i> dengan perbandingan 20% dan 80%.....	72
Tabel 4.23 Perbandingan pengujian dengan jumlah data yang berbeda.....	72

DAFTAR LAMPIRAN

Lampiran 1. Form Revisi

Lampiran 2. Rekomendasi Sidang

BAB I

PENDAHULUAN

1.1 LATAR BELAKANG

Sejak munculnya wabah penyakit coronavirus (COVID-19) di China pada Desember 2019, informasi tentang COVID-19 menyebar luas di internet dan media sosial dengan cepat serta langsung menjadi fokus perhatian dunia. Coronavirus (COVID-19) adalah penyakit menular yang penularannya dapat melalui udara, tetesan dan kontak pernapasan. Virus ini umumnya menular ke manusia. Pada tanggal 30 Januari 2020, Organisasi Kesehatan Dunia (WHO) telah mendaftarkan virus corona sebagai darurat kesehatan masyarakat yang menjadi perhatian internasional [1]. Virus ini masuk Indonesia sejak awal maret 2020.

Pada bulan agustus 2020 pemerintah mengumumkan telah menemukan vaksin virus corona yang akan dibagikan kepada masyarakat dalam rangka menghentikan pandemi ini, kebijakan yang diambil pemerintah kali ini memicu banyak reaksi publik, baik dengan sentimen negatif, netral, dan positif terhadap vaksin yang disetujui pemerintah.

Masyarakat ingin menyampaikan segala pendapat, kritik dan saran, namun karena terbatasnya ruang dan waktu membuat saran tersebut tidak dapat disampaikan. Di sisi lain, adanya jejaring sosial yang merupakan tempat yang cocok untuk mengungkapkan kritik dan saran tersebut, dan Twitter adalah salah satu platform yang tepat untuk menyampaikan hal tersebut. Twitter merupakan salah satu jejaring sosial yang dapat memungkinkan para pengguna mengirimkan *tweet* dengan jumlah text maksimal 140 karakter [2], [3], dan [4].

Text mining adalah cara untuk menemukan pola yang bermakna dari teks dokumen yang tersedia [5] dan sumber data biasanya diperoleh dari dokumen dengan tujuan untuk menemukan kata-kata yang bisa menjadi wakil dari isi dokumen untuk dianalisis.

Natural Language Processing (NLP) adalah area penelitian dan aplikasi yang mengeksplorasi bagaimana komputer dapat digunakan untuk memahami dan memanipulasi teks bahasa alami. Peneliti NLP bertujuan untuk mengumpulkan pengetahuan tentang bagaimana manusia memahami dan menggunakan bahasa sehingga alat dan teknik pemasangan dapat dikembangkan untuk membuat sistem komputer memahami dan memanipulasi bahasa alami untuk melakukan tugas-tugas yang disuka [6], [7].

Dalam penelitian ini, data yang dikumpulkan dari jejaring sosial yang mengandung sejumlah besar sentimen publik akan digunakan untuk memperoleh hasil dari pengklasifikasian terutama pada topik vaksin Covid-19 menggunakan Natural Language Processing (NLP). K-Nearest Neighbor (KNN) [4] Sebagai dasar klasifikasi yang merupakan metode machine learning yang mengklasifikasikan objek berdasarkan data pembelajaran yang paling dekat dengan objek tersebut. Metode ini sangat sederhana, mudah direpresentasikan, tahan terhadap data latih dengan banyak noise, dan sangat efektif dalam proses pengelompokan data sentimen yang dalam penelitian ini merupakan data yang dibagi menjadi sentimen positif, netral dan negatif.

1.2 Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Menganalisa tingkat akurasi algoritma *K-Nearest Neighbor* (KNN) yang digunakan dalam analisis sentimen dengan topik vaksin Covid-19.
2. Melakukan klasifikasi pada sentimen dari *data text* jejaring sosial menggunakan *K-Nearest Neighbor* (KNN).
3. Membuat algoritma untuk melakukan klasifikasi sentimen.

1.3 Manfaat

Adapun manfaat dari penelitian ini adalah:

1. Menambah pengetahuan mengenai diterapkannya metode *K-Nearest Neighbor* (KNN) terhadap klasifikasi sentimen masyarakat dengan topik vaksin Covid-19.
2. Mengetahui persentase sentimen masyarakat mengenai vaksin Covid-19.

1.4 Perumusan Masalah

Adapun perumusan masalah dalam tugas akhir ini adalah :

1. Bagaimana hasil dan akurasi dari algoritma *K-Nearest Neighbor* (KNN) untuk melakukan klasifikasi pada *sentiment* dari data text jejaring sosial dengan topik vaksin covid-19 ?
2. Bagaimana peran algoritma dalam mengklasifikasikan *sentiment* negatif, netral, dan positif ?

1.5 Batasan Masalah

Dari rumusan masalah diatas, akan dibuat pembatasan masalah sebagai berikut :

1. Penelitian ini menggunakan dataset Drone Emprit Academy yang berupa dataset sentimen masyarakat terhadap vaksin covid-19 dalam bentuk API Twitter mulai dari tanggal 24 Agustus 2020 sampai dengan tanggal 24 November 2020.
2. Penelitian ini menggunakan *K-Nearest Neighbor* (KNN) untuk melakukan klasifikasi sentimen.
3. Penelitian ini menggunakan data teks menggunakan bahasa Indonesia.

1.6 Metodologi Penelitian

Tahapan-tahapan yang terdapat pada tugas akhir ini menggunakan metodologi sebagai berikut :

1. Perumusan Masalah

Pada tahap ini dilakukan sebuah cara tentang bagaimana menentukan masalah dalam melakukan *preprocessing data* dalam

bentuk data teks dengan topik Vaksin Covid-19, serta *feature extraction*, dan penyelesaiannya.

2. Studi Pustaka / Literatur

Pada tahap kedua akan dilakukan pencarian literatur mengenai penelitian tentang klasifikasi sentiment masyarakat, menggunakan metode K-Nearest Neighbor (KNN).

3. Perancangan

Tahap ketiga ini merupakan tahap menyusun sistem yang dilakukan berlandaskan perumusan masalah yang terdapat dalam penelitian. Pembahasan dalam tahap ini yaitu tentang bagaimana proses program dalam melakukan penormalan dan klasifikasi data teks, serta *feature segmentation*.

4. Pengambilan Data

Pada tahap ini akan diambilnya data yang berupa kumpulan sentimen masyarakat terhadap vaksin Covid-19.

5. Pengujian

Pada tahap kelima ini data yang telah didapat kemudian akan diuji dan dibandingkan dengan k-fold cross validation sehingga pengujian akan didapatkan hasil akhir.

6. Analisis

Tahap ini akan dilakukan pengambilan data dan menganalisa data training dengan data testing menggunakan algoritma dan program yang telah dibuat. Analisa akan didapatkan perbandingan hasil dari jumlah nilai cross validation yang dimasukkan dan juga jumlah data split sehingga dapat diketahui nilai mana yang lebih baik.

7. Kesimpulan dan Saran

Tahap terakhir merupakan kesimpulan dan saran, dimana tahap ini akan dilakukan pengambilan kesimpulan yang berupa hasil penelitian dan analisa serta studi literatur. Kemudian akan disertakan saran yang ditujukan untuk penelitian berikutnya.

1.7 Sistematika Penulisan

Sistematika penulisan yang terdapat pada tugas akhir ini memiliki beberapa tahapan sebagai berikut :

BAB I PENDAHULUAN

Pada bab yang pertama ini berisi tentang jabaran dari dilakukannya penelitian ini yang berupa latar belakang, tujuan dan manfaat, rumusan masalah, batasan masalah, metodologi dan sistematika penulisan tugas akhir.

BAB II TINJAUAN PUSTAKA

Pada bab ini berisi penjelasan mengenai teori apa saja yang memiliki kaitan dengan penelitian yang akan dilakukan. Pembahasan penelitian ini meliputi analisis sentimen, jejaring sosial, text mining, NLP dan K-Nearest Neighbor.

BAB III METODOLOGI PENELITIAN

Dalam bab ketiga berisi penjelasan tentang bagaimana metodologi ini akan dilakukan, diantaranya yaitu pre-processing, ekstraksi fitur, serta klasifikasi dan validasi.

BAB IV HASIL DAN ANALISIS

Bab keempat yaitu berisi tentang bagaimana cara mendapatkan hasil, kemudian akan dilakukan analisis dari penelitian yang telah dilakukan pada tugas akhir ini.

BAB V KESIMPULAN

Bab terakhir yaitu berisi tentang kesimpulan dari hasil yang diperoleh dan analisa yang telah dilakukan pada bab sebelumnya terhadap penelitian yang telah dilakukan.

DAFTAR PUSTAKA

- [1] Y. Zhao, S. Cheng, X. Yu, and H. Xu, “Chinese Public ’ s Attention to the COVID-19 Epidemic on Social Media : Observational Descriptive Study Corresponding Author :,” vol. 22, pp. 1–13, 2020.
- [2] M. Anjaria, R. Mahana, and R. Guddeti, “Influence Factor Based Opinion Mining of Twitter Data Using Supervised Learning,” 2014.
- [3] A. Go, R. Bhayani, and L. Huang, “Twitter Sentiment Classification using Distant Supervision.”
- [4] J. Samuel, G. G. N. Ali, M. Rahman, E. Esawi, and Y. Samuel, “COVID-19 Public Sentiment Insights and Machine Learning for Tweets Classification,” pp. 1–22.
- [5] R. Patel and G. Sharma, “A survey on text mining techniques,” vol. 3, no. 5, pp. 5621–5625, 2014.
- [6] S. Vijayarani, M. J. Ilamathi, and M. Nithya, “Preprocessing Techniques for Text Mining - An Overview,” vol. 5, no. 1, pp. 7–16.
- [7] V. Gupta, L. C. Science, and G. S. Lehal, “A Survey of Text Mining Techniques and Applications,” vol. 1, no. 1, pp. 60–76, 2009.
- [8] A. Ali, “Sentiment Analysis on Twitter Data using KNN and SVM,” vol. 8, no. 6, pp. 19–25, 2017.
- [9] R. Prabowo and M. Thelwall, “Sentiment Analysis : A Combined Approach.”
- [10] R. Ferdiana *et al.*, “Dataset Indonesia untuk Analisis Sentimen,” vol. 8, no. 4, pp. 334–339, 2019.
- [11] T. T. Le *et al.*, “The COVID-19 vaccine development landscape,” *Nat. Rev. Drug Discov.*, vol. 19, no. May, pp. 305–306, 2020.

- [12] I. Science, “Chowdhury , G . Natural language processing,” pp. 51–89.
- [13] S. R. Joseph, “Natural Language Processing : A Review,” vol. 6, no. 3, 2016.
- [14] A. Dey and A. S. Learning, “Machine Learning Algorithms : A Review,” vol. 7, no. 3, pp. 1174–1179, 2016.
- [15] V. Bijalwan, V. Kumar, P. Kumari, and J. Pascual, “KNN based Machine Learning Approach for Text and Document Mining,” vol. 7, no. 1, pp. 61–70, 2014.
- [16] W. Zhang, T. Yoshida, and X. Tang, “Expert Systems with Applications A comparative study of TF Ï IDF , LSI and multi-words for text classification,” *Expert Syst. Appl.*, vol. 38, no. 3, pp. 2758–2765, 2011.
- [17] F. Aloul, “Journal of Advances in Information Technology,” 2009.
- [18] R. Munir, *Algoritma dan pemrograman : Dalam bahasa pascal dan C*. 2011.
- [19] A. Hamed, A. Sobhy, and H. Nassar, “Accurate Classification of COVID-19 Based on Incomplete Heterogeneous Data using a K NN Variant Algorithm.”
- [20] K. Shi, L. Li, H. Liu, J. He, N. Zhang, and W. Song, “AN IMPROVED KNN TEXT CLASSIFICATION ALGORITHM BASED ON DENSITY,” no. 1.
- [21] M. Bramer and T. Edition, *Principles of Data Mining*. .
- [22] J. A. Septian, T. M. Fahrudin, and A. Nugroho, “Analisis Sentimen Pengguna Twitter Terhadap Polemik Persepakbolaan Indonesia Menggunakan Pembobotan TF - IDF dan K - Nearest Neighbor,” pp. 43–49.
- [23] H. Kong, H. Kong, H. Kong, H. Kong, E. Management, and H. Kong, “Cross-validation based,” pp. 1–58.

- [24] V. M. Patro and M. R. Patra, “Augmenting Weighted Average with Confusion Matrix to Enhance Classification Accuracy,” no. August 2014.