

**KLASIFIKASI SENTIMEN TERHADAP *DATA TEXT*
JEJARING SOSIAL DENGAN TOPIK VAKSIN
COVID-19 MENGGUNAKAN
*NAÏVE BAYES CLASSIFIER***



OLEH:

JERRY PRATAMA SAPUTRA

09011281621059

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

**KLASIFIKASI SENTIMEN TERHADAP *DATA TEXT*
JEJARING SOSIAL DENGAN TOPIK VAKSIN
COVID-19 MENGGUNAKAN
*NAÏVE BAYES CLASSIFIER***

TUGAS AKHIR

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



Oleh:

JERRY PRATAMA SAPUTRA

09011281621059

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

LEMBAR PENGESAHAN
KLASIFIKASI SENTIMEN TERHADAP *DATA TEXT*
JEJARING SOSIAL DENGAN TOPIK VAKSIN COVID-19
MENGGUNAKAN *NAÏVE BAYES CLASSIFIER*

TUGAS AKHIR

Diajukan Untuk Melengkapi Salah Satu Syarat
Memperoleh Gelar Sarjana Komputer

Oleh:

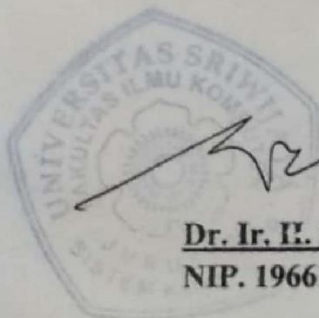
JERRY PRATAMA SAPUTRA
09011281621059

Indralaya, Juli 2021

Mengetahui,

Ketua Jurusan Sistem Komputer

Pembimbing Skripsi



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

6/8/21

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

HALAMAN PERSETUJUAN

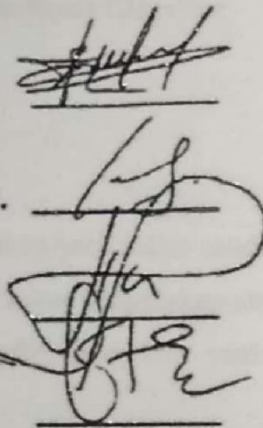
Telah diuji dan lulus pada :

Hari : Jum'at

Tanggal : 2 Juli 2021

Tim Penguji :

1. Ketua : Sarmayanta Sembiring, S.Si., M.T.
2. Sekretaris : Iman Saladin B. Azhar, S.Kom., M.MSI.
3. Pembimbing : Huda Ubaya, M.T.
4. Penguji : Dr. Ir. Bambang Tutuko, M.T.



Mengetahui, 6/8/21

Ketua Jurusan Sistem Komputer



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

NIP. 196612032006041001

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini :

Nama : Jerry Pratama Saputra

NIM : 09011281621059

Judul : Klasifikasi Sentimen Terhadap *Data Text* Jejaring Sosial Dengan Topik Vaksin Covid-19 Menggunakan *Naïve Bayes Classifier*

Hasil Penyecekan *Software iThenticate/Turnitin* : 10 %

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



Jerry Pratama Saputra
NIM. 09011281621059

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 Proposal Tugas Akhir ini dengan judul “Klasifikasi Sentimen Terhadap *Data Text* Jejaring Sosial Dengan Topik Vaksin Covid-19 Menggunakan *Naïve Bayes Classifier*”.

Dalam laporan ini penulis menjelaskan mengenai bagaimana cara mengklasifikasi *Sentimen* dengan menggunakan *Naïve Bayes Classifier* 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 textmining, khususnya pengusaha dan pemerintah.

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. Orang tua beserta keluarga yang selalu mendoakan, memberikan dukungan baik moral 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. H.Sukemi, M.T. selaku Ketua Jurusan Sistem Komputer Fakultas Ilmu Komputer Universitas Sriwijaya.
4. Bapak Huda Ubaya, M.T selaku Pembimbing 1 Tugas Akhir.
5. Bapak Dr. Ir. H. Sukemi, M.T. selaku Pembimbing Akademik Jurusan Sistem Komputer.
6. Ilham Solehan dan Rizki Ronaldo sebagai partner dan teman dari awal pengerjaan 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. Para sahabat saya, Hanif Habibie, Aditya Jovandi, Nauval Faris Fadhlullah , Hisyam Sanusi, Ichwanul Hakim, Sarianto, Rap Nur Muhammad, Madyus Randika & Muhammad Sultan Alif yang selalu memotivasi saya untuk cepat lulus.
9. Dan semua pihak yang telah membantu.

Penulis menyadari bahwa Proposal 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

Jerry Pratama Saputra
NIM. 09011281621059

KLASIFIKASI SENTIMEN TERHADAP DATA TEXT JEJARING SOSIAL DENGAN TOPIK VAKSIN COVID-19 MENGGUNAKAN NAÏVE BAYES CLASSIFIER

Jerry Pratama Saputra (09011281621059)

Jurusan Sistem Komputer, Fakultas Ilmu Komputer, Universitas Sriwijaya

Email : jerrypratamasa84@gmail.com

ABSTRAK

Vaksin *Coronavirus* yang akan di bagikan ke warga negara sebagai rangka menghentikan pandemi ini telah diumumkan oleh pemerintah pada bulan Agustus tahun 2020, kebijakan pemerintah kali ini banyak memicu reaksi masyarakat baik itu sentimen negatif ataupun positif terhadap vaksin yang baru saja akan di sahkan pemerintah. Dengan beragamnya sentimen masyarakat terhadap vaksin Covid-19 ini, klasifikasi sentimen masyarakat terhadap vaksin ini dianggap perlu sebab tingginya antusiasme masyarakat dalam menyambut vaksin yang akan diluncurkan pemerintah. Klasifikasi sentimen kali ini dimungkinkan dengan adanya data yang telah di kumpulkan pada suatu jejaring sosial yang memuat banyak sentimen masyarakat khususnya mengenai topik Vaksin Covid-19, dengan menggunakan *Natural Language Processing* (NLP) klasifikasi sentimen yang berupa *data text* dapat dimungkinkan untuk dilaksanakan, dan dengan menggunakan *Naïve Bayes Classifier* sebagai dasar dari klasifikasi yang digunakan maka data sentimen masyarakat akan terklasifikasi menjadi beberapa bagian Berdasarkan percobaan yang telah dilakukan, Data uji 30% memiliki hasil performa yang tertinggi dimana nilai presisinya 87.46%, recall 85.83%, dan F1-scorenya 86.51%. Sedangkan Data uji 80% memiliki performa terendah dimana nilai presisinya 84.25%, recall 80.94%, dan F1-score 82.12%. Metode *Naïve Bayes* yang digunakan pada penelitian ini mampu melakukan klasifikasi data text dengan baik, sehingga sentimen dengan topik Vaksin Covid-19 mampu diklasifikasi oleh sistem. Hasil prediksi sentimen yang dilakukan dengan algoritma *Naïve Bayes* dapat digunakan sebagai bahan acuan untuk pemerintah dalam pengendalian kontribusi vaksin ke masyarakat.

Kata Kunci : *Naïve Bayes*, *Natural Language Processing*, Vaksin Covid-19, jejaring sosial, data text.

CLASSIFICATION OF SENTIMENTS ON SOCIAL NETWORK TEXT DATA WITH COVID-19 VACCINE TOPIC USING NAÏVE BAYES CLASSIFIER

Jerry Pratama Saputra (09011281621059)

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

Email : jerrypratamasa84@gmail.com

ABSTRACT

The coronavirus vaccine which will be distributed to citizens in order to stop this pandemic that has been announced by the government in August 2020, in this time the government's policy has triggered many public reactions, both negative and positive sentiments towards the vaccine that the government has just approved. With the variety of public sentiments towards the Covid-19 vaccine, the classification of public sentiment towards the vaccine is considered necessary because of the high enthusiasm of the community in welcoming the vaccine that will be launched by the government. This sentiment classification is made possible by the data that has been collected on a social network that contains a lot of public sentiment, especially regarding the topic of the Covid-19 Vaccine, by using Natural Language Processing (NLP), it is possible to classify sentiments in the form of text data, and with using the Naïve Bayes Classifier as the basis for the classifier used, the community sentiment data will be classified into several parts. While the 80% test data has the lowest performance where the precision value is 84.25%, recall is 80.94%, and F1-score is 82.12%. The Naïve Bayes method used in this study is able to classify text data well, so that the sentiment on the topic of the Covid-19 vaccine can be clarified by the system. The results of sentiment predictions made with the Naïve Bayes algorithm can be used as reference material for the government in controlling the contribution of vaccines to the community.

Keywords: *Naïve Bayes, Natural Language Processing, Covid-19 Vaccine, social networking, text data.*

DAFTAR ISI

	Halaman
Halaman Judul	i
Lembar Pengesahan	ii
Halaman Persetujuan	iii
Halaman Pernyataan	iv
Kata Pengantar	v
Abstrak	vii
Abstract	viii
Daftar Isi	ix
Daftar Gambar	xiii
Daftar Tabel	xvi
Bab I. Pendahuluan	1
1.1 Latar Belakang	1
1.2. Tujuan	2
1.3. Perumusan Masalah	2
1.4. Batasan Masalah	2
1.5. Metodologi Penelitian	3
1.6. Sistematika Penulisan	4
Bab II. Tinjauan Pustaka	5
2.1. Analisis Sentimen	5
2.2. Jejaring Sosial	5

2.3. Vaksin COVID-19.....	6
2.4. API Twitter Dataset	6
2.5. <i>Natural Language Processing (NLP)</i>	7
2.5.1. <i>Syntax</i>	7
2.5.2. <i>Semantics</i>	8
2.6. <i>Machine Learning</i>	9
2.7. <i>Text Mining</i>	10
2.7.1. <i>Text Mining vs Data Mining</i>	10
2.8. <i>Naïve Bayes Classifier</i>	11
2.8.1. Struktur sistem learning	12
2.9. K-Fold Cross Validation	12
2.10. Confusion Matrix	13
Bab III. Metodologi Penelitian	18
3.1. Pendahuluan	18
3.2. Kerangka Kerja	18
3.3. Pengambilan Data	20
3.4. <i>Preprocessing</i>	22
3.4.1. <i>Cleaning</i>	23
3.4.2. <i>Stemming</i>	24
3.4.3. <i>Word Tokenizing</i>	26
3.4.4. Pemisahaan Feature dan label.....	26
3.5. Ekstraksi fitur	27
3.6. Cross Validation	28

3.7. Confusion Matrix	28
Bab IV. Hasil dan Pembahasan	30
4.1. Pendahuluan	30
4.2. <i>Dataset Text</i>	30
4.3. Proses <i>Preprocessing</i>	30
4.4. Ekstraksi Fitur Data Text	33
4.5. Percobaan dan Pengujian data training untuk menentukan nilai CV	33
4.5.1. Pengujian data training dengan nilai <i>Cross</i> <i>Validation 2</i>	35
4.5.2. Pengujian data training dengan nilai <i>Cross</i> <i>Validation 5</i>	37
4.5.3. Pengujian data training dengan nilai <i>Cross</i> <i>Validation 10</i>	39
4.5.4. Pengujian data training dengan nilai <i>Cross</i> <i>Validation 20</i>	42
4.6. Perbandingan hasil pengujian dengan nilai cross validation yang bervariasi	44
4.7. Klasifikasi dengan Naïve Bayes Classifier	46
4.7.1. Percobaan dengan Data Training 80% dan Data Test 20%	47
4.7.2. Percobaan dengan Data Training 70% dan Data Test 30%	49
4.7.3. Percobaan dengan Data Training 60% dan Data Test 40%	51
4.7.4. Percobaan dengan Data Training 50% dan Data Test 50%	53
4.7.5. Percobaan dengan Data Training 40% dan Data Test 60%	55
4.7.6. Percobaan dengan Data Training 30% dan Data	

Test 70%	57
4.7.7. Percobaan dengan Data Training 20% dan Data	
Test 80%	59
4.8. Perbandingan hasil pengujian Data Test	61
4.9. Analisis	62
Bab V. Kesimpulan	64
5.1 Kesimpulan	64
5.2 Saran.....	64
Daftar Pustaka	65

DAFTAR GAMBAR

	Halaman
Gambar 3.1. Diagram alur penelitian	19
Gambar 3.2. Bentuk Dataset yang diambil dari Drone Emprit Academy.....	20
Gambar 3.3. Wordcloud sentimen negatif	21
Gambar 3.4. Wordcloud sentimen netral	21
Gambar 3.5. Wordcloud sentimen positif	22
Gambar 3.6. Diagram tahap preprocessing	22
Gambar 3.7. Diagram proses cleaning	23
Gambar 3.8. Sentimen yang diproses cleaning	24
Gambar 3.9. Diagram proses stemming	25
Gambar 3.10. Sentimen yang diproses stemming	25
Gambar 3.11. Diagram proses tokenizing	26
Gambar 3.12. Sentimen yang diproses tokenisasi	26
Gambar 3.13. Total data dan label masing-masing tweet	27
Gambar 3.14. Kata-kata yang menjadi skalar	27
Gambar 4.1. Diagram proses <i>Preprocessing</i>	31
Gambar 4.2. Bentuk data text setelah melalui tahap <i>Cleaning</i>	32
Gambar 4.3. Bentuk data text setelah melalui tahap <i>Stemming</i>	32
Gambar 4.4. Bentuk data text setelah melalui tahap <i>Tokenization</i>	32
Gambar 4.5. Sampel data ekstraksi fitur ke 200 sampai 215	33

Gambar 4.6. Sampel data ekstraksi fitur ke 1000 sampai 1015.....	33
Gambar 4.7. Hasil prediksi dan confusion matrix pada training set CV = 2	36
Gambar 4.8. Hasil prediksi dan confusion matrix pada training set CV = 5	38
Gambar 4.9. Hasil prediksi dan confusion matrix pada training set CV = 10	40
Gambar 4.10. Hasil prediksi dan confusion matrix pada training set CV = 20	43
Gambar 4.11. Perbandingan presisi pada nilai Cross Validation yang berbeda – beda	45
Gambar 4.12. Perbandingan recall pada nilai Cross Validation yang berbeda – beda	45
Gambar 4.13. Perbandingan F1-score pada nilai Cross Validation yang berbeda – beda	46
Gambar 4.14. Hasil percobaan dengan Data Test 20%	47
Gambar 4.15. Hasil percobaan dengan Data Test 30%	49
Gambar 4.16. Hasil percobaan dengan Data Test 40%	51
Gambar 4.17. Hasil percobaan dengan Data Test 50%	53
Gambar 4.18. Hasil percobaan dengan Data Test 60%	55
Gambar 4.19. Hasil percobaan dengan Data Test 70%	57
Gambar 4.20. Hasil percobaan dengan Data Test 80%	59
Gambar 4.21. Perbandingan nilai Presisi pada tiap Data Test	61

Gambar 4.22. Perbandingan nilai Recall pada tiap Data Test	62
Gambar 4.23. Perbandingan nilai F1-Score pada tiap Data Test	62

DAFTAR TABEL

	Halaman
Tabel 2.1. Confusion Matrix 2x2	14
Tabel 2.2. Confusion Matrix 3x3	15
Tabel 3.1. Confusion Matrix 3x3	28
Tabel 4.1. Confusion Matrix 3x3 untuk pengujian	34
Tabel 4.2. Hasil perbandingan pengujian antara nilai Cross Validation yang berbeda – beda	45
Tabel 4.3. Split Data Training dan Data Test	46
Tabel 4.4. Hasil perbandingan pengujian antara Data Test yang jumlahnya berbeda – beda	61

BAB I

PENDAHULUAN

1.1 Latar Belakang

Upaya global untuk mengembangkan dan menyebarkan vaksin COVID-19 berkembang pesat. Untuk mencapai kumpulan imunisasi yang baik diperlukan vaksinasi yang meluas, banyak masyarakat yang diperlukan untuk bekerja sama. Maka, pemerintah dan lembaga kesehatan masyarakat harus memahami pandangan masyarakat tentang vaksin, yang dapat membantu didalam kegiatan pendidikan dan menargetkan intervensi kebijakan yang lainnya.

Vaksin coronavirus yang akan di bagikan ke warga negara sebagai rangka menghentikan pandemi ini telah diumumkan oleh pemerintah pada bulan Agustus tahun 2020, kebijakan pemerintah kali ini banyak memicu reaksi masyarakat baik itu sentimen negatif ataupun positif terhadap vaksin yang baru saja akan di sahkan pemerintah.

Dengan beragamnya sentimen masyarakat terhadap vaksin Covid-19 ini, klasifikasi sentimen masyarakat terhadap vaksin ini dianggap perlu sebab tingginya antusiasme masyarakat dalam menyambut vaksin yang akan diluncurkan pemerintah.

Klasifikasi sentimen kali ini dimungkinkan dengan adanya data yang telah di kumpulkan pada suatu jejaring sosial yang memuat banyak sentimen masyarakat khususnya mengenai topik Vaksin Covid-19, dengan menggunakan *Natural Language Processing* (NLP) klasifikasi sentimen yang berupa data text dapat dimungkinkan untuk dilaksanakan, dan dengan menggunakan *Naïve Bayes Classifier* sebagai dasar dari Klasifer yang digunakan maka data sentimen masyarakat akan terklasifikasi menjadi beberapa bagian.

Berdasarkan pada penelitian [1]. Kelebihan algoritma *Naïve Bayes* yaitu algoritmanya hanya memerlukan data training yang berukuran kecil untuk mengestimasi parameter dalam pengklasifikasiannya. Meskipun kesederhanaan

dan asumsi yang kuat, klasifikasi *Naïve Bayes* telah terbukti bekerja dengan baik di banyak domain. Klasifikasi *Bayesian* menyediakan algoritma belajar yang praktis dan data sebelumnya dapat digabungkan dengan data yang telah diobservasi. Dalam teknik *Naïve Bayes*, ide dasar untuk menemukan probabilitas pada tiap kategori yang telah diberikan pada dokumen teks menggunakan probabilitas gabungan antara kata dan kategori. Hal ini didasarkan oleh asumsi independensi pada setiap kata.

1.2 Tujuan

Adapun tujuan dilakukannya penelitian ini yaitu :

1. Melakukan klasifikasi pada sentiment dari data text jejaring sosial menggunakan *Naïve Bayes Classifier*.
2. Membuat algoritma untuk mengklasifikasi sentimen-sentimen menjadi sentimen negatif, sentimen netral, dan sentimen positif.

1.3 Perumusan Masalah

Permasalahan dalam penelitian ini adalah belum tersedianya informasi tentang klasifikasi terhadap sentimen-sentimen masyarakat tentang vaksin Covid-19 yang ada di Indonesia.

1.4 Batasan Masalah

1. Jejaring sosial yang dianalisa sentimennya adalah tweet-tweet masyarakat dari jejaring sosial Twitter.
2. Data teks yang digunakan dalam proses klasifikasi hanya data teks berbahasa Indonesia.
3. Waktu pengambilan data yang diambil yaitu dari 24 Agustus 2020 hingga 24 November 2020.

1.5 Metodologi Penelitian

Metodologi yang digunakan dalam tugas akhir ini akan melewati beberapa tahapan sebagai berikut:

1. Perumusan masalah

Tahap ini ialah tahap yang menentukan permasalahan yang ada tentang bagaimana melakukan *preprocessing* data text Vaksin Covid-19, ekstraksi fitur, dan cara menyelesaikan permasalahan tersebut.

2. Studi Pustaka/literatur

Tahap ini adalah proses pencarian literatur yang menunjang dalam penelitian yang akan dilakukan, baik menggunakan kata kunci tertentu atau spesifik dalam bidang yang diteliti.

3. Perancangan

Tahap ini ialah tahap perancangan sistem yang dibuat berdasarkan perumusan masalah yang dicari dalam penelitian. Pada tahap ini akan dibahas tentang proses pembuatan program untuk *Preprocessing* data text, segmentasi fitur, *Klasifikasi* Data text.

4. Pengujian

Tahap ini dilakukan dengan pengujian terhadap apa yang dibuat sesuai dengan rancangan.

5. Analisis

Tahap ini dilakukan dengan mengambil data dan menganalisa data yang sudah diolah dengan algoritma dan program yang dibuat. Analisa akan divisualisasikan untuk mengetahui seberapa baik program yang telah dibuat dapat menjalankan fungsi dan tujuannya.

6. Kesimpulan dan Saran

Tahap ini dilakukan dengan menarik kesimpulan dari analisa dan studi literatur serta saran untuk penulis selanjutnya jika akan melanjutkan penelitian ini.

1.6 Sistematika Penulisan

Sistematika yang akan digunakan dalam penulisan tugas akhir ini akan melewati beberapa tahap sebagai berikut :

BAB I PENDAHULUAN

Pada bab pertama ini berisi tentang penjabaran secara sistematis topik yang diambil meliputi latar belakang, tujuan dan manfaat, perumusan dan batasan masalah, metodologi dan sistematika penulisan

BAB II TINJAUAN PUSTAKA

Pada bab kedua ini menjelaskan dasar teori yang menunjang pembahasan dari penelitian ini. Dasar teori ini berisi tentang Analisis Sentimen, Sosial media, *Machine Learning*, *Natural Language Processing*, dan *Naïve Bayes*

BAB III METODOLOGI PENELITIAN

Pada bab ketiga ini menjelaskan bagaimana penelitian ini dilakukan yang dimulai dari pra-pemrosesan data, ekstraksi fitur pada data, klasifikasi, serta validasi.

BAB IV HASIL DAN ANALISIS

Pada bab keempat ini menjelaskan hasil dan analisa terhadap penelitian yang telah dilakukan.

BAB V KESIMPULAN

Pada bab kelima ini menarik kesimpulan berdasarkan hasil dan analisa terhadap penelitian yang telah dilakukan.

DAFTAR PUSTAKA

- [1] L. Dey, S. Chakraborty, A. Biswas, B. Bose, and S. Tiwari, "Sentiment Analysis of Review Datasets Using Naïve Bayes' and K-NN Classifier," *Int. J. Inf. Eng. Electron. Bus.*, vol. 8, no. 4, pp. 54–62, 2016, doi: 10.5815/ijieeb.2016.04.07.
- [2] X. Chen, H. Xie, G. Cheng, L. K. M. Poon, M. Leng, and F. L. Wang, "Trends and features of the applications of natural language processing techniques for clinical trials text analysis," *Appl. Sci.*, vol. 10, no. 6, pp. 1–36, 2020, doi: 10.3390/app10062157.
- [3] E. Martínez-Cámara, M. T. Martín-Valdivia, L. A. Ureña-López, and A. R. Montejo-Ráez, "Sentiment analysis in Twitter," *Nat. Lang. Eng.*, vol. 20, no. 1, pp. 1–28, 2014, doi: 10.1017/S1351324912000332.
- [4] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "A survey of text mining in social media: Facebook and Twitter perspectives," *Adv. Sci. Technol. Eng. Syst.*, vol. 2, no. 1, pp. 127–133, 2017, doi: 10.25046/aj020115.
- [5] T. Thanh Le *et al.*, "The COVID-19 vaccine development landscape," *Nat. Rev. Drug Discov.*, vol. 19, no. 5, pp. 305–306, 2020, doi: 10.1038/d41573-020-00073-5.
- [6] Z. Jianqiang and G. Xiaolin, "Comparison research on text pre-processing methods on twitter sentiment analysis," *IEEE Access*, vol. 5, pp. 2870–2879, 2017, doi: 10.1109/ACCESS.2017.2672677.
- [7] J. Samuel, G. G. M. N. Ali, M. M. Rahman, E. Esawi, and Y. Samuel, "COVID-19 public sentiment insights and machine learning for tweets classification," *Inf.*, vol. 11, no. 6, pp. 1–22, 2020, doi: 10.3390/info11060314.

- [8] B. J. Grosz, "Natural language processing," *Artif. Intell.*, vol. 19, no. 2, pp. 131–136, 1982, doi: 10.1016/0004-3702(82)90032-7.
- [9] J. T. Wolohan, M. Hiraga, A. Mukherjee, and Z. A. Sayyed, "Detecting Linguistic Traces of Depression in Topic-Restricted Text: Attending to Self-Stigmatized Depression with NLP," *Workshop*, pp. 11–21, 2018, [Online]. Available: <https://doi.org/10.18653/v1/P17>.
- [10] M. Rahul, P. #1, and M. Gaurav Sharma, "Page No. 5621-5625 Mr," *Int. J. Eng. Comput. Sci.*, vol. 3, no. 5, pp. 5621–5625, 2014, [Online]. Available: www.ijecs.in.
- [11] F. C. Permana, Y. Rosmansyah, and A. S. Abdullah, "Naive Bayes as opinion classifier to evaluate students satisfaction based on student sentiment in Twitter Social Media," *J. Phys. Conf. Ser.*, vol. 893, no. 1, 2017, doi: 10.1088/1742-6596/893/1/012051.
- [12] S. Ruuska, W. Hämäläinen, S. Kajava, M. Mughal, P. Matilainen, and J. Mononen, "Evaluation of the confusion matrix method in the validation of an automated system for measuring feeding behaviour of cattle," *Behav. Processes*, vol. 148, pp. 56–62, 2018, doi: 10.1016/j.beproc.2018.01.004.