



**ANALISIS SENTIMEN *TWEET* VAKSIN COVID-19
MENGUNAKAN *LONG SHORT-TERM MEMORY*
DENGAN PEMBOBOTAN TF-IDF**

Skripsi

**Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer
Oleh**

**JUAIRIAH
NIM. 1611016220010**

**PROGRAM STUDI S-1 ILMU KOMPUTER
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS LAMBUNG MANGKURAT
BANJARBARU**

JUNI 2023

SKRIPSI

**ANALISIS SENTIMEN *TWEET* VAKSIN COVID-19
MENGUNAKAN *LONG SHORT-TERM MEMORY*
DENGAN PEMBOBOTAN TF-IDF**

Oleh:

**JUAIRIAH
NIM. 1611016220010**

Telah dipertahankan di depan Dosen Penguji pada 23 Juni 2023.

Susunan Dosen Penguji:

Pembimbing I

Mohammad Reza Faisal, S.T., M.T., Ph.D.
NIP.197612202008121001

Dosen Penguji I

Irwan Budiman, S.T., M.Kom.
NIP.197703252008121001

Pembimbing II

**Triando Hamonangan Saragih, S.Kom.,
M.Kom.**
NIP.199308242019031012

Dosen Penguji II

**Muhammad Itqan Mazdadi, S.Kom.,
M.Kom.**
NIP. 199006122019031013



PERNYATAAN

Dengan ini saya menyatakan bahwa dalam skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diberikan orang lain, kecuali yang secara tertulis diacu didalam naskah ini dan disebutkan dalam Daftar Pustaka

Banjarbaru, Juni 2023



Juairiah
NIM. 1611016220010

ABSTRAK

ANALISIS SENTIMEN TWEET VAKSIN COVID-19 MENGGUNAKAN METODE LONG SHORT-TERM MEMORY DENGAN PEMBOBOTAN TF-IDF (Oleh : Juairiah; Pembimbing : Mohammad Reza Faisal, S.T., M.T., Ph.D dan Triando Hamonangan Saragih, S.Kom., M.Kom; 2023; 92 halaman)

Satgas Penanganan Pandemi Covid-19 di Indonesia, menunjukkan bahwa tercatat jumlah kasus terkonfirmasi positif di Indonesia mencapai angka 1,22 juta kasus per 16 Februari 2020. Salah satu upaya pemerintah Indonesia untuk meminimalisir persebaran virus ini yaitu dengan mendistribusikan vaksin covid-19 kepada masyarakat. Dataset yang digunakan untuk penelitian ini adalah data tweet vaksin covid-19 yang merupakan hasil *crawling* data menggunakan package *r*, menggunakan 1800 data dengan 3 kelas menjadi 600 data berlabel Netral, 600 data berlabel Positif, dan 600 data berlabel Negatif. Hasil *crawling* data tidak berbentuk angka namun berbentuk teks, kondisi ini akan menyulitkan metode klasifikasi dalam melakukan proses data mining. Ekstraksi fitur *Term Frequency-Inverse Document Frequency* (TF-IDF) yang bisa digunakan untuk menyelesaikan permasalahan tersebut. Klasifikasi yang digunakan *Long Short-Term Memory* (LSTM) adalah versi modifikasi dari jaringan saraf berulang, yang membuatnya lebih mudah untuk mengingat data masa lalu dalam memori. Metode LSTM digunakan dalam melakukan analisis sentimen yang memiliki hasil yang lebih baik dibandingkan dengan metode konvensional. Hasil penelitian pada data tweet vaksin covid-19 menggunakan *Long Short-Term Memory* dengan Pembobotan TF-IDF pada pengujian menggunakan 1 layer dengan *epoch* 50, *batch size* 128 diperoleh akurasi sebesar 42,8%, presisi 58,9%, dan recall 42,8% dan pada pengujian menggunakan 2 layer dengan *epoch* 200, *batch size* 512 diperoleh akurasi sebesar 50%, presisi 58,8%, dan recall 50%.

Kata Kunci : Tweet Vaksin Covid-19, TF-IDF, *Long Short-Term Memory*, Klasifikasi Teks, Twitter

ABSTRACT

ANALYSIS OF COVID-19 VACCINE TWEET SENTIMENT USING THE LONG SHORT-TERM MEMORY METHOD WITH TF-IDF WEIGHTING (By : Juairiah; Supervisors : Mohammad Reza Faisal, S.T, M.T, Ph.D and Triando Hamonangan Saragih, S.Kom., M.Kom; 2023; 92 pages)

The Task Force for Handling the Covid-19 Pandemic in Indonesia shows that the recorded number of positive confirmed cases in Indonesia has reached 1.22 million cases as of February 16, 2020. One of the efforts by the Indonesian government to minimize the spread of this virus is by distributing the Covid-19 vaccine to the public. The dataset used for this research is the Covid-19 vaccine tweet data which is the result of data crawling using an R package containing 1800 data divided into 3 classes which include 600 data labeled as Neutral, 600 data labeled as Positive, and 600 data labeled as Negative. The results of the data crawling are not in the form of numbers but in the form of text, this condition will complicate the classification method in carrying out the data mining process. Term Frequency-Inverse Document Frequency (TF-IDF) feature extraction can be used to solve this problem. The classification method used in this research is Long Short-Term Memory (LSTM) which is a modified version of a recurrent neural network, making it easier to recall past data in memory. The LSTM method is used in sentiment analysis which has better results than the conventional method. The research on the Covid-19 vaccine tweet data using Long Short-Term Memory with TF-IDF Weighting in the test using 1 layer with 50 epochs and batch size 128 resulted in accuracy of 42.8%, precision 58.9%, and recall 42,8%, and in the test using 2 layers with 200 epochs and batch size of 152 resulted in accuracy of 50%, precision 58.8%, and recall 50%.

Keywords : Covid-19 Vaccine Tweet, TF-IDF, Long Short-Term Memory, Text classification, Twitter

PRAKATA

Puji syukur kepada Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **Analisis Sentimen *Tweet* Vaksin Covid-19 Menggunakan *Long Short-Term Memory* dengan Pembobotan TF-IDF** untuk memenuhi syarat dalam menyelesaikan pendidikan program Strata-1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tidak lupa juga penulis panjatkan sholawat serta salam kepada Rasulullah SWT beserta para kerabat, sahabat, dan pengikut beliau hingga hari kiamat.

Dalam penyusunan skripsi ini penulis banyak mengalami hambatan, namun berkat dukungan, bantuan, bimbingan, dan kerjasama dari berbagai pihak sehingga dapat terselesaikan dengan baik. Sehingga penulis mengucapkan terimakasih kepada:

1. Allah SWT, karena atas limpahan karunia-Nya penulis dapat menyelesaikan skripsi ini.
2. Diri sendiri yang senantiasa mempertahankan semangat dan memberikan energi positif dalam jiwa raga sehingga dapat menyelesaikan skripsi dalam kondisi apapun.
3. Keluarga yang senantiasa memberikan dukungan serta memberikan doa sehingga terselesaikannya skripsi ini.
4. Bapak Mohammad Reza Faisal, S.T., M.T., Ph.D selaku dosen pembimbing utama yang turut serta membantu, meluangkan waktu, serta memberi arahan selama proses penyelesaian skripsi.
5. Bapak Triando Hamonangan Saragih, S.Kom., M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi.
6. Bapak Irwan Budiman S.T., M.Kom dan bapak Muhammad Itqan Mazdadi, S.Kom., M.Kom selaku dosen penguji yang memberi saran dan meluangkan waktu untuk penyelesaian skripsi.
7. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
8. Teman-teman serta Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna. Oleh karena itu kritik dan saran yang bersifat membangun sangat diharapkan agar dapat melengkapi segala kekurangan dan keterbatasan dalam penyusunan skripsi ini. Akhir kata penulis semoga tulisan ini dapat bermanfaat bagi pengembangan ilmu pengetahuan serta senantiasa mendapatkan lindungan dan ridha Allah SWT.

Banjarbaru, Juni 2023

Juairiah

DAFTAR ISI

LEMBAR JUDUL	i
LEMBAR PENGESAHAN	ii
PERNYATAAN.....	iii
ABSTRAK	iv
ABSTRACT	v
PRAKATA	vi
DAFTAR ISI.....	viii
DAFTAR TABEL.....	x
DAFTAR GAMBAR	xii
DAFTAR LAMPIRAN.....	xiv
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Rumusan Masalah.....	3
1.3. Batasan Masalah	3
1.4. Tujuan.....	3
1.5. Manfaat Penelitian.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1. Literatur Terdahulu.....	5
2.2. Keaslian Penelitian	6
2.3. Twitter	9
2.4. Vaksin COVID-19.....	9
2.5. Analisis Sentimen	10
2.6. Pelabelan.....	11
2.7. Text Mining	13
2.8. N-Gram	16
2.9. TF-IDF.....	18
2.10. Recurrent Neural Network.....	20
2.11. Long Short-Term Memory	21
2.12. Confusion Matrix.....	25

BAB III METODE PENELITIAN.....	28
3.1. Bahan dan Alat Penelitian	28
3.2. Bahan Penelitian	28
3.3. Variabel Penelitian.....	28
3.4. Prosedur Penelitian	29
BAB IV HASIL DAN PEMBAHASAN	31
4.1 Hasil.....	31
4.1.1 Pengumpulan Data.....	31
4.1.2 Preprocessing Data.....	33
4.1.3 Data Mining.....	43
4.1.4 Evaluasi	82
4.2 Pembahasan	86
BAB V PENUTUP.....	90
5.1. Kesimpulan.....	90
5.2. Saran	90
DAFTAR PUSTAKA	91
LAMPIRAN.....	93
RIWAYAT HIDUP.....	111

DAFTAR TABEL

Tabel	Halaman
Tabel 1. Keaslian Penelitian.....	6
Tabel 2. Tabel Rancangan Penelitian.....	8
Tabel 3. Acuan Pelabelan.....	12
Tabel 4. Pendekatan Text mining.....	14
Tabel 5. Contoh Case Folding.....	15
Tabel 6. Contoh Cleaning	15
Tabel 7. Contoh Tokenizing.....	16
Tabel 8. Contoh Filtering	16
Tabel 9. Contoh Stemming	16
Tabel 10. Contoh N-Gram.....	17
Tabel 11. N-gram berdasarkan perkata	18
Tabel 12. Confusion Matrix multiclass.....	25
Tabel 13. Contoh data tweet.....	32
Tabel 14. Label data	34
Tabel 15. Pembagian data kelas	35
Tabel 16. Data Cleansing	35
Tabel 17. Data case folding.....	36
Tabel 18. Kamus Slang	37
Tabel 19. Penerapan formalisasi	37
Tabel 20. Penerapan stemming	38
Tabel 21. Kamus Stopword.....	39
Tabel 22. Penerapan Stopword Removal	39
Tabel 23. Contoh dokumen sebelum proses Uni-gram.....	40
Tabel 24. Contoh hasil Uni-gram.....	40
Tabel 25. Perhitungan TF-IDF.....	41
Tabel 26. Contoh hasil ekstraksi fitur	42
Tabel 27. Detail Pembagian Dataset.....	43
Tabel 28. Data training.....	43
Tabel 29. Data testing	44
Tabel 30. Hasil Pengujian layer 1 LSTM dengan epoch 25	47
Tabel 31. Hasil Pengujian layer 1 LSTM dengan epoch 50	50
Tabel 32. Hasil Pengujian layer 1 LSTM dengan epoch 100	54
Tabel 33. Hasil Pengujian layer 1 LSTM dengan epoch 200	57
Tabel 34. Hasil Pengujian layer 1 LSTM dengan epoch 300	61
Tabel 35. Hasil Pengujian layer 2 LSTM dengan epoch 25	64
Tabel 36. Hasil Pengujian layer 2 LSTM dengan epoch 50	68
Tabel 37. Hasil Pengujian layer 2 LSTM dengan epoch 100	71
Tabel 38. Hasil Pengujian layer 2 LSTM dengan epoch 200	75

Tabel 39. Hasil Pengujian layer 2 LSTM dengan epoch 300	78
Tabel 40. Confusion Matrix Layer 1 Long Short-Term Memory	82
Tabel 41. Perhitungan akurasi klasifikasi Layer 1 Long Short-Term Memory	82
Tabel 42. Perhitungan presisi klasifikasi Layer 1 Long Short-Term Memory	83
Tabel 43. Perhitungan Recall klasifikasi Layer 1 Long Short-Term Memory	83
Tabel 44. Confusion Matrix Layer 2 Long Short-Term Memory	84
Tabel 45. Perhitungan akurasi klasifikasi Layer 2 Long Short-Term Memory	84
Tabel 46. Perhitungan presisi klasifikasi Layer 2 Long Short-Term Memory	84
Tabel 47. Perhitungan Recall klasifikasi Layer 2 Long Short-Term Memory	85
Tabel 48. Hasil kinerja Long Short-Term Memory	85
Tabel 49. Pembagian Variabel Klasifikasi Long Short-Term memory	88
Tabel 50. Kinerja Klasifikasi Long Short-Term Memory pada layer 1 dan layer 2 ..	89

DAFTAR GAMBAR

Gambar	Halaman
Gambar 1. Kerangka umum Text Mining	14
Gambar 2. Hasil TF-IDF	20
Gambar 3. Recurrent Neural Network	20
Gambar 4. Arsitektur RNN	21
Gambar 5. Memori pada RNN	22
Gambar 6. Arsitektur memory cells LSTM	22
Gambar 7. Arsitektur memory cells LSTM	23
Gambar 8. Prinsip Kerja Akurasi, Presisi dan Recall	26
Gambar 9. Alur penelitian.....	29
Gambar 10. Tahap Preprocessing	33
Gambar 11. Model Summary 1 layer.....	45
Gambar 12. Model Summary 2 Layer.....	45
Gambar 13. Model Plot 1 Layer.....	46
Gambar 14. Model Plot 2 Layer.....	46
Gambar 15. Grafik Pembelajaran Model 1 layer, epoch 25, dan Batch size 16	47
Gambar 16. Grafik Pembelajaran Model 1 layer, epoch 25, dan Batch size 32	48
Gambar 17. Grafik Pembelajaran Model 1 layer, epoch 25, dan Batch size 64	48
Gambar 18. Grafik Pembelajaran Model 1 layer, epoch 25, dan Batch size 128	49
Gambar 19. Grafik Pembelajaran Model 1 layer, epoch 25, dan Batch size 256	49
Gambar 20. Grafik Pembelajaran Model 1 layer, epoch 25, dan Batch size 512	50
Gambar 21. Grafik Pembelajaran Model 1 layer, epoch 50, dan Batch size 16	51
Gambar 22. Grafik Pembelajaran Model 1 layer, epoch 50, dan Batch size 32	51
Gambar 23. Grafik Pembelajaran Model 1 layer, epoch 50, dan Batch size 64	52
Gambar 24. Grafik Pembelajaran Model 1 layer, epoch 50, dan Batch size 128	52
Gambar 25. Grafik Pembelajaran Model 1 layer, epoch 50, dan Batch size 256	53
Gambar 26. Grafik Pembelajaran Model 1 layer, epoch 50, dan Batch size 512	53
Gambar 27. Grafik Pembelajaran Model 1 layer, epoch 100, dan Batch size 16	54
Gambar 28. Grafik Pembelajaran Model 1 layer, epoch 100, dan Batch size 32	55
Gambar 29. Grafik Pembelajaran Model 1 layer, epoch 100, dan Batch size 64	55
Gambar 30. Grafik Pembelajaran Model 1 layer, epoch 100, dan Batch size 128	56
Gambar 31. Grafik Pembelajaran Model 1 layer, epoch 100, dan Batch size 256	56
Gambar 32. Grafik Pembelajaran Model 1 layer, epoch 100, dan Batch size 512	57
Gambar 33. Grafik Pembelajaran Model 1 layer, epoch 200, dan Batch size 16	58
Gambar 34. Grafik Pembelajaran Model 1 layer, epoch 200, dan Batch size 32	58
Gambar 35. Grafik Pembelajaran Model 1 layer, epoch 200, dan Batch size 64	59
Gambar 36. Grafik Pembelajaran Model 1 layer, epoch 200, dan Batch size 128	59
Gambar 37. Grafik Pembelajaran Model 1 layer, epoch 200, dan Batch size 256	60
Gambar 38. Grafik Pembelajaran Model 1 layer, epoch 200, dan Batch size 512	60

Gambar 39. Grafik Pembelajaran Model 1 layer, epoch 300, dan Batch size 16	61
Gambar 40. Grafik Pembelajaran Model 1 layer, epoch 300, dan Batch size 32	62
Gambar 41. Grafik Pembelajaran Model 1 layer, epoch 300, dan Batch size 64	62
Gambar 42. Grafik Pembelajaran Model 1 layer, epoch 300, dan Batch size 128	63
Gambar 43. Grafik Pembelajaran Model 1 layer, epoch 300, dan Batch size 256	63
Gambar 44. Grafik Pembelajaran Model 1 layer, epoch 300, dan Batch size 512	64
Gambar 45. Grafik Pembelajaran Model 2 layer, epoch 25, dan Batch size 16	65
Gambar 46. Grafik Pembelajaran Model 2 layer, epoch 25, dan Batch size 32	65
Gambar 47. Grafik Pembelajaran Model 2 layer, epoch 25, dan Batch size 64	66
Gambar 48. Grafik Pembelajaran Model 2 layer, epoch 25, dan Batch size 128	66
Gambar 49. Grafik Pembelajaran Model 2 layer, epoch 25, dan Batch size 256	67
Gambar 50. Grafik Pembelajaran Model 2 layer, epoch 25, dan Batch size 512	67
Gambar 51. Grafik Pembelajaran Model 2 layer, epoch 50, dan Batch size 16	68
Gambar 52. Grafik Pembelajaran Model 2 layer, epoch 50, dan Batch size 32	69
Gambar 53. Grafik Pembelajaran Model 2 layer, epoch 50, dan Batch size 64	69
Gambar 54. Grafik Pembelajaran Model 2 layer, epoch 50, dan Batch size 128	70
Gambar 55. Grafik Pembelajaran Model 2 layer, epoch 50, dan Batch size 256	70
Gambar 56. Grafik Pembelajaran Model 2 layer, epoch 50, dan Batch size 512	71
Gambar 57. Grafik Pembelajaran Model 2 layer, epoch 100, dan Batch size 16	72
Gambar 58. Grafik Pembelajaran Model 2 layer, epoch 100, dan Batch size 32	72
Gambar 59. Grafik Pembelajaran Model 2 layer, epoch 100, dan Batch size 64	73
Gambar 60. Grafik Pembelajaran Model 2 layer, epoch 100, dan Batch size 128	73
Gambar 61. Grafik Pembelajaran Model 2 layer, epoch 100, dan Batch size 256	74
Gambar 62. Grafik Pembelajaran Model 2 layer, epoch 100, dan Batch size 512	74
Gambar 63. Grafik Pembelajaran Model 2 layer, epoch 200, dan Batch size 16	75
Gambar 64. Grafik Pembelajaran Model 2 layer, epoch 200, dan Batch size 32	76
Gambar 65. Grafik Pembelajaran Model 2 layer, epoch 200, dan Batch size 64	76
Gambar 66. Grafik Pembelajaran Model 2 layer, epoch 200, dan Batch size 128	77
Gambar 67. Grafik Pembelajaran Model 2 layer, epoch 200, dan Batch size 256	77
Gambar 68. Grafik Pembelajaran Model 2 layer, epoch 200, dan Batch size 512	78
Gambar 69. Grafik Pembelajaran Model 2 layer, epoch 300, dan Batch size 16	79
Gambar 70. Grafik Pembelajaran Model 2 layer, epoch 300, dan Batch size 32	79
Gambar 71. Grafik Pembelajaran Model 2 layer, epoch 300, dan Batch size 64	80
Gambar 72. Grafik Pembelajaran Model 2 layer, epoch 300, dan Batch size 128	80
Gambar 73. Grafik Pembelajaran Model 2 layer, epoch 300, dan Batch size 256	81
Gambar 74. Grafik Pembelajaran Model 2 layer, epoch 300, dan Batch size 512	81
Gambar 75. Grafik kinerja klasifikasi Long Short-Term Memory	86

DAFTAR LAMPIRAN

Lampiran 1. Kamus Slang.....	94
Lampiran 2. Kamus Stopword	99
Lampiran 3. Source Code Preprocessing	103
Lampiran 4. Source Code Klasifikasi Long Short-Term Memory	105
Lampiran 5. Source Code Evaluasi.....	109