ANALISIS SENTIMEN TIKTOK *SHOP* PADA TWITTER MENGGUNAKAN METODE *MULTINOMIAL NAÏVE BAYES* DENGAN PEMBOBOTAN FITUR BM25

SKRIPSI

Diajukan untuk memenuhi salah satu syarat kelulusan
di Program Studi Sains Data

Disusun Oleh:

M. ANDREW ARJUNANDA YASIN

20083010014

PROGRAM STUDI SAINS DATA

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL "VETERAN" JAWA TIMUR

SURABAYA

2024

LEMBAR PENGESAHAN

ANALISIS SENTIMEN TIKTOK *SHOP* PADA TWITTER MENGGUNAKAN METODE *MULTINOMIAL NAÏVE BAYES* DENGAN PEMBOBOTAN FITUR *BM25*

SKRIPSI

Diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Sains Data
pada : Selasa, 14 Mei 2024

Program Studi S-1 Sains Data
Fakultas Ilmu Komputer
Universitas Pembangunan Nasional Veteran Jawa Timur
Surabaya

Oleh :

M. ANDREW ARJUNANDA YASIN

NPM. 20083010014

Disetujui oleh Tim Penguji Skripsi :

Penguji 1	Penguji 2
Trimono, S.Si., M.Si.	Amri Muhaimin, S.Stat., M.Stat., M.S.
NIP. 199509082022031003	NIP. 21119950723270
Pembimbing 1	Pembimbing 2
Dr.Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU	Tresna Maulana Fahrudin, S.ST., M.T.
NIP. 198012052005011002	NIP. 199305012022031007

Fakultas Ilmu Komputer
Dekan,

Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 196811261994032001

Mengetahui,
Program Studi Sains Data
Fakultas Ilmu Komputer
Koordinator,

Dr.Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU
NIP. 198012052005011002

Surabaya, Mei, 2024

SURAT PERNYATAAN

Saya yang bertanda tangan di bawah ini:

Nama	: M. Andrew Arjunanda Yasin
NPM	: 20083010014
Program Studi	: Sains Data

Menyatakan bahwa judul Skripsi / Tugas Akhir sebagai berikut:

ANALISIS SENTIMEN TIKTOK *SHOP* PADA TWITTER MENGGUNAKAN METODE *MULTINOMIAL NAÏVE BAYES* DENGAN PEMBOBOTAN FITUR BM25

Bukan merupakan plagiat dari Skripsi/ Tugas Akhir/ Penelitian orang lain dan juga bukan merupakan produk/ *software*/ hasil karya yang saya beli dari orang lain

Saya juga menyatakan bahwa Skripsi/ Tugas Akhir ini adalah pekerjaan saya sendiri, kecuali yang dinyatakan dalam Daftar Pustaka, dan tidak pernah diajukan untuk syarat memperoleh gelar di Universitas Pembangunan Nasional "Veteran" Jawa Timur maupun di institusi pendidikan lain.
Jika ternyata dikemudian hari pernyataan ini terbukti tidak benar, maka Saya bertanggung jawab penuh dan siap menerima segala konsekuensi, termasuk pembatalan ijazah dikemudian hari

Surabaya, 29 Mei 2024

Hormat Saya

M. Andrew Arjunanda Yasin
NPM. 20083010014

ABSTRAK

ANALISIS SENTIMEN TIKTOK *SHOP* PADA TWITTER MENGGUNAKAN METODE *MULTINOMIAL NAÏVE BAYES* DENGAN PEMBOBOTAN FITUR BM25

Nama Mahasiswa / NPM	**:**	**M. Andrew Arjunanda Yasin / 20083010014**
Program Studi	**:**	**Sains Data, FASILKOM, UPN Veteran Jatim**
Dosen Pembimbing 1	**:**	**Dr.Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU**
Dosen Pembimbing 2	**:**	**Tresna Maulana Fahrudin, S.ST., M.T.**

Abstrak

Kemajuan teknologi internet telah mengubah banyak hal dalam cara pengguna di Indonesia berinteraksi terutama dalam perdagangan dan komunikasi sosial. Salah satu platform yang memanfaatkan kemajuan ini adalah TikTok dengan fitur TikTok *Shop* yang memungkinkan pengguna untuk berbelanja tanpa meninggalkan aplikasi. Namun, TikTok *Shop* sempat ditutup pada 4 Oktober 2023 karena kewajiban mematuhi aturan perdagangan online sebelum kemudian dibuka kembali pada 12 Desember 2023. Kondisi ini menimbulkan berbagai tanggapan di Twitter karena khawatir akan terjadinya monopoli dagang sehingga analisis sentimen diperlukan. Salah satu metode analisis sentimen adalah *Multinomial Naïve Bayes* yang menghitung probabilitas. Proses penelitian ini meliputi pengumpulan data dari Twitter dengan menggunakan *library* python "*tweet harvest*" sebanyak 1413 data, data *preprocessing*, pelabelan data, *term weighting* dengan BM25 dan TF-IDF, seleksi fitur, *validation model*, model klasifikasi menggunakan metode *Multinomial*, *Gaussian*, dan *Bernoulli Naïve Bayes*, serta visualisasi *wordcloud*. Tujuan penelitian untuk membantu pemerintah dalam menentukan kebijakan yang lebih tepat dan memberikan wawasan bagi masyarakat dalam merespons situasi tersebut secara bijaksana. Hasil penelitian menunjukkan bahwa *Multinomial Naïve Bayes* dengan BM25 mencapai akurasi tertinggi sebesar 0.75, dengan mayoritas respons menunjukkan sentimen negatif.

Kata kunci: ***TikTok Shop, Analisis Sentimen, BM25, Multinomial Naïve Bayes.***

ABSTRACT

SENTIMENT ANALYSIS OF TIKTOK SHOP ON TWITTER USING *MULTINOMIAL NAÏVE BAYES* METHOD WITH BM25 FEATURE WEIGHTING

Student Name / NPM	**:**	**M. Andrew Arjunanda Yasin / 20083010014**
Study Program	**:**	**Sains Data, FASILKOM, UPN Veteran Jatim**
Advisor 1	**:**	**Dr.Eng. Ir. Dwi Arman Prasetya, S.T., M.T., IPU**
Advisor 2	**:**	**Tresna Maulana Fahrudin, S.ST., M.T.**

Abstract

Advancements in internet technology have significantly altered how users in Indonesia interact, particularly in commerce and social communication. One platform leveraging this advancement is TikTok with its TikTok *Shop* feature, which allows users to shop without leaving the app. However, TikTok Shop was temporarily closed on October 4, 2023, to comply with online trading regulations, before reopening on December 12, 2023. This situation sparked various responses on Twitter due to concerns about potential trade monopolies, necessitating sentiment analysis to understand public opinion. One sentiment analysis method is *Multinomial Naïve Bayes*, which calculates probabilities. This research process includes data collection from Twitter using the Python *library* "*tweet harvest*", totaling 1413 data points, *data preprocessing*, data labeling, term weighting with BM25 and TF-IDF, feature selection, model validation, classification using *Multinomial*, *Gaussian*, and *Bernoulli Naïve Bayes* methods, and *word cloud visualization*. The research aims to help the government make more informed policy decisions and provide insights for the public to respond wisely to the situation. The results indicate that *Multinomial Naïve Bayes* with BM25 achieves the highest accuracy of 0.75, with the majority of responses showing negative sentiment.

Keywords: *TikTok Shop, Sentiment Analysis,* BM25, *Multinomial Naïve Bayes.*

KATA PENGANTAR

Puji dan syukur kehadirat ALLAH SWT, atas limpahan Rahmat serta Kasih Sayang-Nya sehingga penulis dapat menyelesaikan laporan Skripsi yang merupakan persyaratan dalam menyelesaikan mata kuliah Skripsi pada Program Studi S1 Sains Data di Universitas Pembangunan Nasional "Veteran" Jawa Timur. Dalam penyusunan Skripsi ini tidak terlepas dari bantuan berbagai pihak dan dalam kesempatan ini penulis ingin mengucapkan terima kasih kepada:

1. Orang tua dan keluarga selalu memberikan dukungan dan doa.
2. Prof. Dr. Ir. Akhmad Fauzi, M.MT., IPU selaku Rektor Universitas Pembangunan Nasional "Veteran" Jawa Timur.
3. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional "Veteran" Jawa Timur
4. Bapak Dr. Eng. Ir. Dwi Arman Prasetya, ST., MT., IPU selaku Koordinator Program Studi Sains Data Universitas Pembangunan Nasional "Veteran" Jawa Timur dan Dosen Wali serta Dosen Pembimbing 1.
5. Bapak Tresna Maulana Fahrudin, S.ST., MT selaku Dosen Pembimbing 2.
6. Bapak dan Ibu dosen Program Studi Sains Data UPN "Veteran" Jawa Timur yang sudah berkenan untuk memberikan waktu untuk berkontribusi pada penelitian ini.
7. Teman-teman Sains Data angkatan 2020 dan teman-teman lainnya yang senantiasa memberikan dukungan dalam menyelesaikan penyelesaian skripsi.
8. Cenditya Ayu Aurelia yang senantiasa mendukung untuk menyelesaikan skripsi ini.

Penulis menyadari bahwa masih terdapat banyak kekurangan Skripsi ini, namun penulis berharap semoga laporan Skripsi ini dapat memberikan kontribusi terhadap perkembangan ilmu pengetahuan, khususnya dalam bidang ilmu sains data.

Surabaya, 29 Mei 2024

M. Andrew Arjunanda Yasin

DAFTAR ISI

HALAMAN SAMPUL ii
LEMBAR PENGESAHAN ii
SURAT PERNYATAAN iii
ABSTRAK iv
ABSTRACT v
KATA PENGANTAR vi
DAFTAR ISI vii
DAFTAR GAMBAR x
DAFTAR TABEL xi
DAFTAR LAMPIRAN xiii
BAB I PENDAHULUAN 1
1.1. Latar Belakang 1
1.2. Rumusan Masalah 4
1.3. Batasan Masalah 4
1.4. Tujuan Penelitian 5
1.5. Manfaat Penelitian 5
BAB II TINJAUAN PUSTAKA 6
2.1. Dasar Teori 6
2.1.1. TikTok Shop 6
2.1.2. Twitter 7
2.1.3. Analisis Sentimen 7
2.1.4. *Text Preprocessing* 7
2.1.5. BM25 8
2.1.6. TF-IDF 9
2.1.7. Seleksi Fitur *Chi-Square* 10
2.1.8. *Hold-out Validation* 11
2.1.9. Distribusi *Multinomial* 11
2.1.10. Teorema *Bayes* 12
2.1.11. *Naïve Bayes* 13
2.1.12. *Multinomial Naïve Bayes* 15
2.1.13. *Confusion Matrix* 16

2.1.14. *Wordcloud* 18
2.2. Penelitian Terdahulu 18
BAB III METODOLOGI PENELITIAN 24
3.1. Variabel Penelitian dan Sumber Data 24
3.2. Langkah Analisis 25
3.1.1. Pengumpulan Data 25
3.1.2. *Data Preprocessing* 25
3.1.3. Pelabelan Data 28
3.1.4. *Term Weighting* 28
3.1.5. Seleksi Fitur 31
3.1.6. *Validation Mode*l 31
3.1.7. Model Klasifikasi 31
3.1.8. Visualisasi *Wordcloud* 35
3.3. Diagram Alir Penelitian 35
3.4. Jadwal Penelitian 35
BAB IV HASIL DAN PEMBAHASAN 36
4.1. Pengumpulan Data 36
4.2. *Data preprocessing* 38
4.2.1. *Text Preprocessing* 38
4.2.2. *Handling Missing Value* 50
4.2.3. *Handling Duplicate Data* 51
4.2.4. Penyaringan Data 51
4.3. Pelabelan Data 51
4.4. *Term Weighting* 54
4.6.1. BM25 54
4.6.2. TF-IDF 57
4.5. Seleksi Fitur 60
4.6. *Validation Model* 62
4.7. Model Klasifikasi 63
4.6.1. Fase *Train* 64
4.6.2. Fase *Test* 71
4.6.3. Analisa Hasil 78

4.6.4. Fase Data Baru 82
4.8. Visualisasi *Wordcloud* 84
BAB V PENUTUP 88
DAFTAR PUSTAKA 90
LAMPIRAN 97
BIODATA PENULIS 104

DAFTAR GAMBAR

Gambar 2. 1 Kurva Distribusi *Multinomial* 12
Gambar 3. 1 Diagram *data preprocessing* 26
Gambar 3. 2 *Flowchart Multinomial Naïve Bayes* 32
Gambar 3. 3 Diagram alir penelitian 35
Gambar 4. 1 Jumlah data duplikat 51
Gambar 4. 2 Jumlah data yang kurang dari 3 kata 51
Gambar 4. 3 Distribusi sentimen 54
Gambar 4. 4 Nilai K tertinggi TF-IDF a. K terpilih BM25, b. K terpilih TF-IDF 61
Gambar 4. 5 *Wordcloud* kata positif 85
Gambar 4. 6 *Wordcloud* kata negatif 86
Gambar 4. 7 *Wordcloud* kata negatif 87

DAFTAR TABEL

Tabel 2.1. *Confusion matrix* 16
Tabel 2.2. Tabel penelitian terdahulu 18
Tabel 3.1. Contoh variabel penelitian 24
Tabel 3.2. Manualisasi perhitungan nilai TF 29
Tabel 3.3. Manualisasi perhitungan nilai DF dan IDF 29
Tabel 3.4. Manualisasi *score* BM25 30
Tabel 3.5. Sampel data untuk perhitungan manual 33
Tabel 3.6. Jadwal kegiatan 35
Tabel 4.1. Algoritma *tweet harves*t 36
Tabel 4.2. Data yang diperoleh 37
Tabel 4.3. Algoritma *case folding* 38
Tabel 4.4. Contoh hasil *case folding* 39
Tabel 4.5. Algoritma *cleaning* 40
Tabel 4.6. Contoh hasil *cleaning* 41
Tabel 4.7. Tabel *normalization_dict* 42
Tabel 4.8. Algoritma *normalization* 43
Tabel 4.9. Contoh hasil *normalization* 44
Tabel 4.9. Algoritma *tokenization* 45
Tabel 4.10. Contoh hasil *tokenization* 45
Tabel 4.11. Algoritma *stemming* 46
Tabel 4.12. Contoh hasil *stemming* 47
Tabel 4.13. Algoritma *stopword removal* 48
Tabel 4.14. Contoh hasil *stopword removal* 49
Tabel 4.15. *Missing value* pada setiap kolom 50
Tabel 4.16. Algoritma pelabelan data 51
Tabel 4.17. Contoh hasil pelabelan data 53
Tabel 4.18. Algoritma BM25 55
Tabel 4.19. Hasil pembobotan BM25 56
Tabel 4.20. Kata dengan bobot tertinggi dan terendah BM25 56
Tabel 4.21. Algoritma TF-IDF 58

Tabel 4.22. Hasil pembobotan TF-IDF 58
Tabel 4.23. Kata dengan bobot tertinggi dan terendah TF-IDF 58
Tabel 4.24. Algoritma seleksi fitur *Chi-Square* 60
Tabel 4.25. Kata dengan *Chi-Square* tertunggi tertinggi dan terendah 62
Tabel 4.26. Kata dengan *Chi-Square* tertunggi tertinggi dan terendah 62
Tabel 4.27. Jumlah data *validation model* 63
Tabel 4.28. Algoritma seleksi model klasifikasi 63
Tabel 4.29. *Confusion matrix* dengan BM25 *Train* 65
Tabel 4.30. *Classification report* dengan BM25 *train* 66
Tabel 4.31. *Confusion matrix* dengan BM25 dan *Chi-Square Train* 67
Tabel 4.32. *Classification report* dengan BM25 dan *Chi-Square train* 67
Tabel 4.33. *Confusion matrix* dengan TF-IDF *train* 68
Tabel 4.34. *Classification report* dengan TF-IDF *train* 69
Tabel 4.35. *Confusion matrix* dengan TF-IDF dan *Chi-Square train* 70
Tabel 4.36. *Classification report* dengan TF-IDF dan *Chi-Square train* 71
Tabel 4.37. *Confusion matrix* dengan BM25 *test* 72
Tabel 4.38. *Classification report* dengan BM25 *test* 73
Tabel 4.39. *Confusion matrix* dengan BM25 dan *Chi-Square test* 73
Tabel 4.40. *Classification report* dengan BM25 dan *Chi-Square test* 74
Tabel 4.41. *Confusion matrix* dengan TF-IDF *test* 75
Tabel 4.42. *Classification report* dengan TF-IDF *test* 76
Tabel 4.43. *Confusion matrix* dengan TF-IDF dan *Chi-Square test* 77
Tabel 4.44. *Classification report* dengan TF-IDF dan *Chi-Square test* 77
Tabel 4.45. Akurasi setiap skenario 79
Tabel 4.46. Nilai *prio*r 80
Tabel 4.47. Nilai *likelihood* 80
Tabel 4.48. Nilai *posterior* 81
Tabel 4.49. Algoritma fase data baru 82
Tabel 4.50. Prediksi dengan data baru 83

DAFTAR LAMPIRAN

Lampiran 1. Hasil uji plagiasi 97
Lampiran 2. Data penelitian 102
Lampiran 3. Source code yang digunakan untuk analisis 103