IMPLEMENTASI EKSTRAKSI FITUR *BAG OF WORDS* MENGGUNAKAN SELEKSI FITUR *CHI-SQUARE* DENGAN KLASIFIKASI *CATBOOST* PADA ANALISIS SENTIMEN MEDIA SOSIAL : COVID – 19

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

Muhammad Zamzam

NIM 1711016210022

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

JUNI 2023

IMPLEMENTASI EKSTRAKSI FITUR *BAG OF WORDS* MENGGUNAKAN SELEKSI FITUR *CHI-SQUARE* DENGAN KLASIFIKASI *CATBOOST* PADA ANALISIS SENTIMEN MEDIA SOSIAL : COVID - 19

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Strata-1 Ilmu Komputer

Oleh

Muhammad Zamzam

NIM 1711016210022

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

JUNI 2024

SKRIPSI

IMPLEMENTASI EKSTRAKSI FITUR *BAG OF WORDS* MENGGUNAKAN SELEKSI FITUR *CHI-SQUARE* DENGAN KLASIFIKASI *CATBOOST* PADA ANALISIS SENTIMEN MEDIA SOSIAL : COVID - 19

Oleh :

Muhammad Zamzam

1711016210022

Telah dipertahankan di depan Dosen Penguji pada tanggal 28 Mei 2024

Susunan Dosen Penguji :

Pembimbing I

Muliadi, S.Kom., M.Cs.
NIP. 197804222010121002

Dosen Penguji I

Dodon Turianto Nugrahadi, S.Kom., M.Eng.
NIP. 198001122009121002

Pembimbing II

Friska Abadi, S.Kom M.Kom
NIP. 198809132023211010

Dosen Penguji II

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

Banjarbaru, 28 Mei 2024

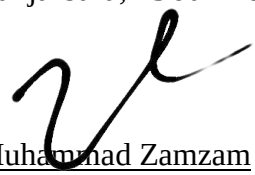
Ketua Program Studi Ilmu Komputer

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

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 diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Banjarbaru, 13 Juli 2023



Muhammad Zamzam
NIM. 1711016110007

ABSTRAK

IMPLEMENTASI EKSTRAKSI FITUR *BAG OF WORDS* MENGGUNAKAN SELEKSI FITUR *CHI-SQUARE* DENGAN KLASIFIKASI *CATBOOST* PADA ANALISIS SENTIMEN MEDIA SOSIAL : *COVID - 19*
(Oleh: Muhammad Zamzam; Pembimbing: Muliadi, S.Kom., M.Cs. dan Friska Abadi, S.Kom M.Kom; 2023; 60 halaman)

Sejak awal tahun 2020, virus *SARS-CoV-2* yang menyebabkan *COVID-19* telah menjadi perhatian global dengan dampak signifikan terhadap masyarakat dan berbagai sektor kehidupan. Media sosial, khususnya Twitter, menjadi platform penting untuk mengumpulkan opini publik terkait *COVID-19* dan vaksinasi. Kebutuhan untuk memahami sikap dan pandangan masyarakat terhadap vaksinasi *COVID-19* di Twitter menuntut adanya analisis sentimen yang efektif. Berbagai metode analisis sentimen telah dilakukan sebelumnya, termasuk menggunakan *Bag of Words (BoW)* dan berbagai metode klasifikasi seperti *Naïve Bayes, Decision Tree, Random Forest*, dan *Catboost*. Penelitian ini berfokus pada implementasi ekstraksi fitur *Bag of Words* dengan seleksi fitur *Chi-Square* dan metode klasifikasi *Catboost* untuk menghasilkan model yang dapat mengklasifikasikan komentar *Twitter* terkait *COVID-19* secara akurat. Hasil penelitian menunjukkan bahwa penggunaan seleksi fitur *Chi-Square* meningkatkan akurasi klasifikasi. Rata-rata akurasi tertinggi dengan seleksi fitur *Chi-Square* adalah 75,2%, dibandingkan dengan 71,2% tanpa seleksi fitur, menunjukkan peningkatan akurasi sebesar 4%. Kesimpulannya, menggabungkan seleksi fitur *Chi-Square* dengan ekstraksi fitur *Bag of Words* dapat meningkatkan akurasi klasifikasi *Catboost* pada analisis sentimen *COVID-19.*

Kata kunci: *Sentiment analysis, Twitter, COVID-19, Bag of Words (BoW), Chi-Square, Catboost*

ABSTRACT

(*By: Muhammad Zamzam; Supervisior: Muliadi, S.Kom., M.Cs. and Friska Abadi, S.Kom M.Kom*; 2023; *60 pages)*

Since early 2020, the SARS-CoV-2 virus, which causes COVID-19, has garnered global attention with significant impacts on society and various sectors of life. Social media, particularly Twitter, has become an important platform for gathering public opinion regarding COVID-19 and vaccination. The need to understand public attitudes and views on COVID-19 vaccination on Twitter necessitates effective sentiment analysis. Various sentiment analysis methods have been conducted previously, including using Bag of Words (BoW) and classification methods such as Naïve Bayes, Decision Tree, Random Forest, and Catboost. This study focuses on implementing Bag of Words feature extraction with Chi-Square feature selection and the Catboost classification method to produce a model that can accurately classify Twitter comments related to COVID-19. The results show that using Chi-Square feature selection increases classification accuracy. The highest average accuracy with Chi-Square feature selection is 75.2%, compared to 71.2% without feature selection, showing an accuracy increase of 4%. In conclusion, combining Chi-Square feature selection with Bag of Words feature extraction can improve the accuracy of Catboost classification in COVID-19 sentiment analysis.

Keywords*: *Sentiment analysis, Twitter, COVID-19, Bag of Words (BoW), Chi-Square, Catboost

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan ke hadirat Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul "*Implementasi Ekstraksi Fitur Bag of Words Menggunakan Seleksi Fitur Chi-Square Dengan Klasifikasi Catboost Pada Analisis Sentimen Media Sosial : Covid - 19*" untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan shalawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga *yaumul qiyamah.*

Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, adapun yang dimaksud adalah sebagai berikut :

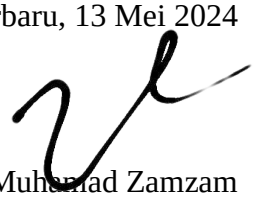
1. Keluarga, terutama kedua orang tua tercinta untuk mama dan abah yang dari awal proses perkuliahan selalu memberikan bantuan, semangat, doa dan dukungan hingga sampai pada tahap proses penyelesaian skripsi ini.
2. Bapak Muliadi, S.Kom., M.Cs selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran penyelesaian skripsi ini.
3. Bapak Friska Abadi, S.Kom M.Kom selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Irwan Budiman, S.T., M.Kom selaku Koordinator Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau skripsi ini dapat diselesaikan.
5. Seluruh Dosen dan Staff Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.
6. Teman-teman yang telah bersedia membantu memecahkan kebingungan-kebingungan dan memberikan saran, dukungan, semangat, serta motivasi, selama proses penyelesaian skripsi.

7. Teman-teman keluarga Ilmu Komputer angkatan 2017 yang memberikan dukungan dan bantuan semasa kuliah dan selama proses penyelesaian skripsi.
8. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian skripsi ini.

Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari sempurna, namun penulis mengharapkan bantuan berupa saran dan kritik yang membangun dari semua pihak demi kesempurnaan dan mutu penulisan skripsi ini.

Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 13 Mei 2024



Muhammad Zamzam

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN