PERBANDINGAN EKSTRAKSI FITUR BERBASIS *N-GRAM* DENGAN PEMBOBOTAN TF-IDF DAN ALGORITMA KLASIFIKASI *RANDOM FOREST* DAN SVM UNTUK DETEKSI PELAPORAN GEJALA COVID-19 DARI TWITTER

Skripsi

Untuk Memenuhi Persyaratan
Dalam Menyelesaikan Sarjana Strata-1 Ilmu Komputer

Oleh
ALMADEA RUSELLAWATI
NIM. 1611016220002

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

JUNI 2023

PERBANDINGAN EKSTRAKSI FITUR BERBASIS *N-GRAM* DENGAN PEMBOBOTAN TF-IDF DAN ALGORITMA KLASIFIKASI *RANDOM FOREST* DAN SVM UNTUK DETEKSI PELAPORAN GEJALA COVID-19 DARI TWITTER

Skripsi

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

ALMADEA RUSELLAWATI
NIM. 1611016220002

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

JUNI 2023

SKRIPSI

PERBANDINGAN EKSTRAKSI FITUR BERBASIS *N-GRAM* DENGAN PEMBOBOTAN TF-IDF DAN ALGORITMA KLASIFIKASI *RANDOM FOREST* DAN SVM UNTUK DETEKSI PELAPORAN GEJALA COVID-19 DARI TWITTER

Oleh:

ALMADEA RUSELLAWATI

NIM.1611016220002

Telah dipertahankan di depan Dosen Penguji pada tanggal 23 Juni 2023

Susunan Dosen Penguji:

Pembimbing I

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

Dosen Penguji I

Dwi Kartini, S.Kom, M.Kom
NIP.198704212012122003

Pembimbing II

Muhammad Itqan Mazdadi, M.Kom.
NIP. 199006122019031013

Dosen Penguji II

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

Banjarbaru, Juni 2023

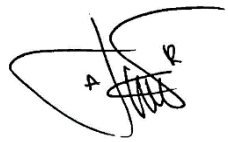
Ketua Program Studi Ilmu Komputer

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

PERYATAAN

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, 23 Juni 2023



Almadea Rusellawati
NIM.1611016220002

ABSTRAK

PERBANDINGAN EKSTRAKSI FITUR BERBASIS *N-GRAM* DENGAN PEMBOBOTAN TF-IDF DAN ALGORITMA KLASIFIKASI *RANDOM FOREST* DAN SVM UNTUK DETEKSI PELAPORAN GEJALA COVID-19 DARI TWITTER (Oleh : Almadea Rusellawati; Pembimbing : Mohammad Reza Faisal, S.T, M.T, Ph.D dan Muhammad Itqan Mazdadi, M.Kom.; 2023; 63 halaman)

Virus Covid-19 hingga saat ini masih sangat banyak menimbulkan gejala-gejala kepada masyarakat. Meski tidak separah dulu, tetapi Covid-19 juga belum kunjung hilang. Saat ini masyarakat sudah dengan sangat terbuka menceritakan tentang diri mereka yang sedang terkena atau memiliki gejala Covid-19 di berbagai media sosial mereka, salah satunya adalah Twitter. Dataset yang digunakan untuk penelitian ini adalah data gejala Covid-19 dari penelitian sebelumnya oleh Sari (2022). Pembobotan TF-IDF digunakan untuk menghitung jumlah kemunculan term pada data, dilakukan pengujian kembali menggunakan *N-Gram* sebagai fitur pemisah kata. Dimana *N-Gram* tersebut dibagi menjadi 3 yaitu *unigram, bigram & trigram*. Penelitian ini menggunakan klasifikasi *Random Forest* dan *Support Vector Machine* yang melakukan pengujian pada kernel *Linear, Radial Basic Function,* dan *Polymonial*. Performa hasil akurasi tertinggi diperoleh dari pengujian *Unigram* dengan Support *Vector Machine* kernel *Linear* sebesar 88,8%.

Kata Kunci : Ekstraksi Fitur, Gejala Covid-19, *N-Gram,* TF-IDF, Klasifikasi, *Random Forest, Support Vector Machine*

ABSTRACT

COMPARISON OF N-GRAM BASED FEATURE EXTRACTION WITH TF-IDF WEIGHTING AND RANDOM FOREST AND SVM CLASSIFICATION ALGORITHMS FOR DETECTION OF COVID-19 SYMPTOM REPORTING FROM TWITTER (By : Almadea Rusellawati; Pembimbing : Mohammad Reza Faisal, S.T, M.T, Ph.D dan Muhammad Itqan Mazdadi, M.Kom.; 2023; 63 pages)

The Covid-19 virus is still very much causing symptoms to the community. Although not as severe as before, Covid-19 has not disappeared. Currently, people are very openly telling about themselves who are affected or have symptoms of Covid-19 on their various social media, one of which is Twitter. The dataset used for this research is data on Covid-19 symptoms from previous research by Sari (2022). TF-IDF weighting is used to calculate the number of occurrences of terms in the data, testing again using *N-Gram* as a word separation feature. Where the *N-Gram* is divided into 3 namely *unigram, bigram* and *trigram*. This research uses *Random Forest* and *Support Vector Machine* classification which tests the *Linear*, *Radial Basic Function*, and *Polymonial* kernels. The highest accuracy performance is obtained from testing *Unigram* with *Support Vector Machine Linear* kernel of 88.8%.

Keywords : Feature Extraction, Covid-19 Symptoms, N-Gram, TF-IDF, Classification, Random Forest, Support Vector Machine

PRAKATA

Assalamualaikum Warahmatullahi Wabarakatuh.

Puji syukur penulis panjatkan kepada Allah SWT karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan skripsi yang berjudul **Perbandingan Ekstrasi Fitur Berbasis *N-Gram* dengan Pembobotan TF-IDF pada Algoritma Klasifikasi *Random Forest* dan SVM untuk Deteksi Pelaporan Gejala Covid-19 dari Twitter** untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat. Tak lupa pula penulis panjatkan sholawat dan salam ke hadirat Rasulullah Muhammad SAW beserta para sahabat, keluarga, dan pengikut beliau hingga yaumul qiama.

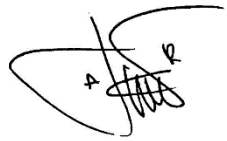
Pada lembar ini penulis ingin menyampaikan ucapan terimakasih kepada pihak-pihak yang sangat mendukung penulis dalam pembuatan dan penyusunan skripsi ini, diantaranya:

1. Diri saya sendiri yang tidak pernah patah semangat walaupun banyak menemui kesulitan baik disebabkan oleh diri sendiri maupun dari hal lain.
2. Keluarga terutama kedua orang tua saya bapak Rusmawardi dan Ibu Suswati, dan juga adik-adik saya Almaulida Azzahra dan M. Raffa Almadani yang selalu memberikan dukungan, doa, dan bantuan dalam proses penyelesaian skripsi ini.
3. Bapak Mohammad Reza Faisal, S.T., M.T., Ph.D selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
4. Bapak Muhammad Itqan Mazdadi, M.Kom. selaku dosen pembimbing pendamping dan Koordinator Program Studi Ilmu Komputer FMIPA ULM yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian skripsi ini.
5. Ibu Dwi Kartini, S.Kom, M.Kom dan bapak Triando Hamonangan Saragih, S.Kom., M.Kom.selaku dosen penguji yang banyak memberi saran dan meluangkan waktu untuk penyelesaian skripsi ini.
6. Seluruh Dosen dan staf Program Studi Ilmu Komputer FMIPA ULM atas ilmu dan bantuan yang diberikan selama ini yang sangat bermanfaat.

7. Teman-teman saya yang telah bersedia membantu memecahkan kebingungan-kebingungan dan memberikan saran, dukungan, semangat, serta motivasi, selama proses penyelesaian skripsi.
8. Teman-teman keluarga Ilmu Komputer angkatan 2016 yang memberikan dukungan dan bantuan semasa kuliah dan selama proses penyelesaian skripsi.
9. 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, 23 Juni 2023



Almadea Rusellawati

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN