IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

SKRIPSI SARJANA

Oleh

Muhammad Rizkiansyah
183112706450091

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL

2022

IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

SKRIPSI SARJANA

Karya ilmiah sebagai salah satu syarat untuk memperoleh gelar
Sarjana Teknik Teknologi Informatika dari Fakultas Teknologi Komunikasi dan Informatika

Oleh

Muhammad Rizkiansyah
183112706450091

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNOLOGI KOMUNIKASI DAN INFORMATIKA
UNIVERSITAS NASIONAL
2022

HALAMAN PENGESAHAN
TUGAS SARJANA

IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

Oleh

Muhammad Rizkiansyah
183112706450091

Program Studi Teknik Informatika
Fakultas Teknologi Komunikasi dan Informatika

Disetujui pada tanggal: 03 November 2022

Pembimbing I	Pembimbing II
(Ratih Titi Komala Sari, ST, MM, MMSI)	(Albaar Rubhasy, S.Si, MTI)
NID. 0103150850	NID. 050020069

KATA PENGANTAR

Puji dan syukur penulis panjatkan kepada Allah SWT Tuhan Yang Maha Esa yang telah memberikan rahmat dan karunia sehingga penulis dapat menyelesaikan skripsi dengan judul "**Implementasi Metode Fitur Ekstraksi dan Algoritma Supervised Learning dalam Mengidentifikasi Jenis Kelamin berdasarkan Nama**" sebagai salah satu syarat kelulusan Program Studi Sarjana Informatika Fakultas Teknologi Komunikasi dan Informatika.

Penelitian dan penulisan skripsi ini tidak terlepas dari bantuan berbagai pihak, oleh karena itu penulis menyampaikan banyak terima kasih terutama kepada dosen pembimbing Tugas Akhir, ibu Ratih Titi Komala Sari, ST, MM, MMSI dan bapak Albaar Rubhasy, S.Si, MTI yang telah meluangkan banyak waktu, tenaga, pikiran, bimbingan, arahan, motivasi serta memaklumi segala kekurangan penulis selama penelitian tugas akhir dan penyusunan skripsi. Penulis juga mengucapkan banyak terima kasih kepada:

1. Bapak dan Mamah selaku orangtua penulis yang telah banyak memberi dukungan dan doa yang tak terhitung.
2. Seluruh dosen pengajar di program studi Informatika FTKI maupun dosen di program studi lain, yang memberikan banyak pengetahuan dan ilmu.
3. Teman-teman seangkatan dan sehimpunan yang telah membantu dan memberikan dukungan.
4. Kaka dan sahabat-sahabat yang telah memberikan banyak dukungan secara mental maupun fisik.

Akhir kata, semoga Allah SWT Tuhan Yang Maha Esa membalas kebaikan dan bantuan yang telah diberikan dengan hal yang lebih baik. Penulis mengharapkan kritik dan saran yang bersifat membangun dan semoga skripsi ini dapat memberikan manfaat dibidang teknologi informatika.

Tangerang Selatan, 17 Desember 2022

Muhammad Rizkiansyah

HALAMAN PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademik Program Studi Teknik Informatika, Fakultas Teknologi Komunikasi dan Informatika, saya yang bertanda tangan di bawah ini:

Nama : Muhammad Rizkiansyah

NIM : 183112706450091

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Fakultas Teknologi Komunikasi dan Informatika, Hak Bebas Royalti Noneksklusif (*Non-exclusive Royalti Free Right*) atas karya ilmiah saya yang berjudul:

IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

Beserta perangkat yang ada (jika diperlukan). Dengan Hak ini Fakultas Teknologi Komunikasi dan Informatika berhak menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Dibuat di : Tangerang Selatan

Pada tanggal : 17 Desember 2022

Yang menyatakan

(Muhammad Rizkiansyah)

HALAMAN PENGESAHAN

TUGAS AKHIR

IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

Muhammad Rizkiansyah

183112706450091

Dosen Pembimbing 1	Dosen Pembimbing 2
(Ratih Titi Komalasari, S. T, MM., MMSI)	(Albaar Rubhasy, S.Si, MTI)

PERNYATAAN KEASLIAN TUGAS AKHIR

Saya menyatakan dengan sesungguhnya bahwa Tugas Akhir dengan judul :

IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

Yang dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional, sebagaimana yang saya ketahui adalah bukan merupakan tiruan atau publikasi dari Tugas Akhir yang pernah diajukan atau dipakai untuk mendapatkan gelar di lingkungan Universitas Nasional maupun perguruan tinggi atau instansi lainnya, kecuali pada bagian – bagian tertentu yang menjadi sumber informasi atau acuan yang dicantumkan sebagaimana mestinya.

Jakarta, 08 Maret 2023

Muhammad Rizkiansyah

183122706450091

LEMBAR PERSETUJUAN TUGAS AKHIR

Tugas Akhir dengan judul :

IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

Dibuat untuk melengkapi salah satu persyaratan menjadi Sarjana Komputer pada Program Studi Informatika, Fakultas Teknologi Komunikasi dan Informatika Universitas Nasional. Tugas Akhir ini diujikan pada Sidang Akhir Semester Ganjil 2022-2023 pada tanggal 22 Februari Tahun 2023

Dosen Pembimbing 1

Ratih Titi Komalasari, S. T, MM., MMSI

NID. 0103150850

Ketua Program Studi

Ratih Titi Komalasari, S. T, MM., MMSI

NID. 0103150850

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Muhammad Rizkiansyah

NPM : 183112706450091

Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika

Program Studi : Informatika

Tanggal Sidang : 22 Februari 2023

JUDUL DALAM BAHASA INDONESIA :

IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

JUDUL DALAM BAHASA INGGRIS :

IMPLEMENTATION OF FEATURE EXTRACTION METHOD AND SUPERVISED LEARNING ALGORITHM IN IDENTIFYING GENDER BASED ON NAME

TANDA TANGAN DAN TANGGAL		
Pembimbing 1	**Ka. Prodi**	**Mahasiswa**
TGL :	TGL :	TGL : 13 - 03 - 2023

LEMBAR PERSETUJUAN JUDUL YANG TIDAK ATAU YANG DIREVISI

Nama : Muhammad Rizkiansyah

NPM : 183112706450091

Fakultas/Akademi : Fakultas Teknologi Komunikasi dan Informatika

Program Studi : Informatika

Tanggal Sidang : 22 Februari 2023

JUDUL DALAM BAHASA INDONESIA :

IMPLEMENTASI METODE FITUR EKSTRAKSI DAN ALGORITMA SUPERVISED LEARNING DALAM MENGIDENTIFIKASI JENIS KELAMIN BERDASARKAN NAMA

JUDUL DALAM BAHASA INGGRIS :

IMPLEMENTATION OF FEATURE EXTRACTION METHOD AND SUPERVISED LEARNING ALGORITHM IN IDENTIFYING GENDER BASED ON NAME

TANDA TANGAN DAN TANGGAL		
Pembimbing 2	**Ka. Prodi**	**Mahasiswa**
TGL : 10-03-2023	TGL :	TGL : 10-03-2023

ABSTRAK

Pemegang data memiliki dokumen terkait data personal akan tetapi tidak adanya variabel gender dalam dokumen tersebut. Dalam hal ini, memungkinkan pemilik data mengklasifikasikan variabel gender berdasarkan nama yang ada secara manual. Mengklasifikasikan dokumen secara manual dinilai tidak seefisien seperti dulu, karena jumlah data yang semakin meningkat. Algoritma supervised learning yang ada pada machine learning dapat berperan, sebagai cara alternatif dalam pengklasifikasian dokumen. Penelitian ini dilakukan untuk mencari model yang terbaik dalam mengidentifikasi jenis kelamin berdasarkan nama, yang diimplementasikan ke dalam aplikasi prediksi berbasis web, dengan menggunakan fitur ekstraksi countvectorizer dan pemanfaatan n-gram serta membandingkan kedua algoritma supervised learning yaitu logistic regression dan multinomial naive bayes. Hasil dari penelitian ini didapatkan model terbaik yaitu model logistic regression pada rentan 2-12 gram dengan split data 80:20 yang memiliki tingkat akurasi 94.76%, dan berdasarkan uji validasi model menggunakan confusion matrix logistic regression memperoleh 0.95 f1-score pada semua labeling, beserta dari hasil uji prediksi yang mendapatkan kesalahan prediksi tersedikit yaitu 28. Maka dari itu model logistic regression dengan split data 80:20 yang akan diterapkan ke dalam aplikasi prediksi berbasis web.

Kata Kunci: Identifikasi, Fitur Ekstraksi, Model Klasifikasi, Confusion Matrix, Aplikasi Prediksi.

ABSTRACT

The data owner has documents containing personal data, but there is no mention of gender in those records. In this instance, it enables the data owner to manually categorize gender variables based on names that already exist. The growing volume of data makes manual document classification less effective than it formerly was. As a substitute method of classifying documents, the supervised learning technique used in machine learning may be useful. By comparing the two supervised learning algorithms, logistic regression and multinomial naive Bayes, and using the countvectorizer extraction feature and n-grams, this research sought to identify the most effective model for classifying gender based on name. This model was then implemented into a web-based prediction application. The results of this study obtained the best model, namely the logistic regression model at a susceptibility of 2-12 grams with a data split of 80:20 which had an accuracy rate of 94.76%, and based on the model validation test using a confusion matrix logistic regression obtained 0.95 f1-score on all labeling, along with from the results of the prediction test that gets the least prediction error, namely 28. Therefore, the logistic regression model with an 80:20 data split will be applied to a web-based prediction application.

Keywords: Identification, Feature Extraction, Classification Model, Confusion Matrix, Prediction Application.

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR