



**KLASIFIKASI GENDER DARI PESAN MEDIA SOSIAL TWITTER
DENGAN MENGGUNAKAN EKSTRAKSI FITUR BERBASIS FASTTEXT
DAN ALGORITMA LSTM**

Skripsi

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

Oleh

HALIMATUS SA'DIAH

NIM 1811016320010

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

MEI 2024



**KLASIFIKASI GENDER DARI PESAN MEDIA SOSIAL TWITTER
DENGAN MENGGUNAKAN EKSTRAKSI FITUR BERBASIS FASTTEXT
DAN ALGORITMA LSTM**

Skripsi

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

Oleh

HALIMATUS SA'DIAH

NIM 1811016320010

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

MEI 2024

SKRIPSI

KLASIFIKASI GENDER DARI PESAN MEDIA SOSIAL TWITTER DENGAN MENGUNAKAN EKSTRAKSI FITUR BERBASIS FASTTEXT DAN ALGORITMA LSTM

Oleh:


HALIMATUS SA'DIAH

NIM. 1811016320010

Telah dipertahankan di depan Dosen Penguji pada tanggal 31 Mei 2024.

Susunan Dosen Penguji:

Pembimbing I



M. Reza Faisal, S.T, M.T, Ph.D.

NIP. 197612202008121001


Dosen Penguji I



Friska Abadi, S.Kom, M.Kom

NIP. 198809132023211010

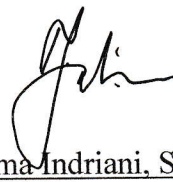
Pembimbing II



Andi Farmadi, S.Si, M.T

NIP. 197307252008011006

Dosen Penguji II



Fatma Indriani, S.T, M.I.T, Ph.D

NIP. 198404202008122004

Banjarmasin, 31 Mei 2024

Koordinator Program Studi Ilmu Komputer



Juwana Budiman, S. T., M. Kom.

NIP. 197703252008121001

SURAT PERNYATAAN

Dengan ini saya menyatakan bahwa dalam jurnal 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, 31 Mei 2024
Yang Menyatakan,



Halimatus Sa'diah
NIM. 1811016320010

ABSTRACT

KLASIFIKASI GENDER PADA PESAN MEDIA SOSIAL MENGGUNAKAN EKSTRAKSI FITUR BERBASIS FASTTEXT DAN LONG SHORT-TERM MEMORY

(Oleh : Halimatus Sa'diah; Pembimbing; Mohammad Reza Faisal, S.Si S.T M.T PhD dan Andi Farmadi, S.Si., M.T.; 2024; halaman)

Saat ini, media sosial digunakan sebagai platform untuk berinteraksi dengan banyak orang dan juga telah menjadi sumber informasi bagi peneliti atau analis media sosial. Twitter adalah salah satu platform yang sering digunakan untuk tujuan penelitian, terutama untuk data dari tweet yang ditulis oleh individu. Namun, di Twitter, informasi pengguna seperti jenis kelamin tidak ditampilkan secara eksplisit dalam profil akun, meskipun terdapat banyak informasi tidak terstruktur yang mengandung data tersebut, sering kali tidak disadari. Penelitian ini bertujuan untuk mengklasifikasikan jenis kelamin berdasarkan data tweet dan data deskripsi akun serta menentukan akurasi klasifikasi jenis kelamin menggunakan metode pembelajaran mesin. Metode yang digunakan melibatkan FastText sebagai metode ekstraksi fitur dan LSTM sebagai metode klasifikasi berdasarkan data yang diekstraksi, sementara untuk mencapai hasil yang paling akurat, klasifikasi dilakukan pada data tweet, data deskripsi akun, dan kombinasi keduanya. Penelitian ini menunjukkan bahwa klasifikasi LSTM pada data deskripsi akun dan data kombinasi mencapai akurasi sebesar 70%, sedangkan klasifikasi data tweet mencapai 69%. Penelitian ini menyimpulkan bahwa ekstraksi fitur FastText dengan klasifikasi LSTM dapat diimplementasikan untuk klasifikasi jenis kelamin. Namun, tidak ada perbedaan signifikan dalam hasil akurasi untuk setiap dataset. Meski demikian, penelitian ini menunjukkan bahwa kedua metode tersebut dapat bekerja sama dengan baik dan menghasilkan hasil yang optimal.

Kata Kunci - ekstraksi fitur, klasifikasi gender, fastText, RNN, LSTM.

ABSTRACT

GENDER CLASSIFICATION ON SOCIAL MEDIA MESSAGES USING FASTTEXT-BASE FEATURE EXTRACTION AND LONG SHORT-TERM MEMORY

(By : Halimatus Sa'diah; Supervisor; Mohammad Reza Faisal, S.Si S.T M.T PhD and Andi Farmadi, S.Si., M.T.; 2024; page)

Currently, social media is used as a platform for interacting with many people and has also become a source of information for social media researchers or analysts. X is one of the platforms commonly used for research purposes, especially for data from tweets written by individuals. However, on X, user information such as gender is not explicitly displayed in the account profile, yet there is a plethora of unstructured information containing such data, often unnoticed. This research aims to classify gender based on tweet data and account description data and determine the accuracy of gender classification using machine learning methods. The method used involves FastText as a feature extraction method and LSTM as a classification method based on the extracted data, while to achieve the most accurate results, classification is performed on tweet data, account description data, and a combination of both. This research shows that LSTM classification on account description data and combined data obtained an accuracy of 70%, while tweet data classification achieved 69%. This research concludes that FastText feature extraction with LSTM classification can be implemented for gender classification. However, there is no significant difference in accuracy results for each dataset. However, this research demonstrates that both methods can work well together and yield optimal results.

Keywords - feature extraction, gender classification, fastText, RNN, LSTM.

KATA PENGANTAR

Puji syukur penulis panjatkan ke Tuhan kita Yang Maha Esa karena atas berkat rahmat dan karunia-Nya penulis dapat menyelesaikan jurnal yang berjudul “*Gender Classification on Social Media Messages Using fasttext-base Feature Extraction and Long Short-term Memory*” untuk memenuhi syarat dalam menyelesaikan pendidikan program S1 Ilmu Komputer, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Lambung Mangkurat.

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

1. Keluarga, terutama papa dan mama yang selalu memberikan bantuan, semangat, doa dan dukungan dalam proses penyelesaian jurnal ini.
2. Bapak M. Reza Faisal, S.T, M.T, Ph.D. selaku dosen pembimbing utama yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.
3. Bapak Andi Farmadi, S.Si., M.T. selaku dosen pembimbing pendamping yang turut serta membantu dan meluangkan waktu demi kelancaran dalam penyelesaian jurnal ini.
4. Bapak Irwan Budiman, S.T., M. Kom. selaku Ketua Program Studi Ilmu Komputer FMIPA ULM, atas bantuan dan izin beliau jurnal 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. Wahyu, Fenny dan klinik tongfang yang selalu mendukung dalam berbagai hal
7. Teman-teman dan sahabat-sahabat keluarga Ilmu Komputer yang memberikan dukungan dan selalu mengingatkan serta mendoakan dalam proses mengerjakan jurnal.
8. Orang-orang yang pernah menemani penulis, terima kasih atas segalanya sehingga penulis terus termotivasi untuk menyelesaikan jurnal ini.
9. Halimatus Sa’diah, yang merupakan diri saya sendiri. Apresiasi sebesar-besarnya kepada diri ini karena telah bertanggung jawab untuk menyelesaikan

apa yang telah dimulai. Terima kasih sudah bertahan, terima kasih karena terus berusaha dan tidak menyerah.

10. Semua pihak yang tidak dapat disebutkan satu persatu yang telah turut membantu dalam penyelesaian Karya Tulis Ilmiah ini.

Semoga Allah SWT memberikan balasan yang berlipat atas amalan dan bantuan yang telah diberikan kepada penulis dalam menyelesaikan jurnal ini. Akhir kata penulis menyadari sepenuhnya bahwa penulisan ini jauh dari kata sempurna. Semoga tulisan ini dapat bermanfaat bagi ilmu pengetahuan dan pembaca khususnya serta mendapat keridhaan Allah SWT.

Banjarbaru, 31 Mei 2024



Halimatus Sa'diah