SKRIPSI

PERBANDINGAN HASIL PENERAPAN ALGORITMA KLASIFIKASI DAN *NATURAL LANGUAGE PROCESSING* TERHADAP DATA KEPUASAN PENGGUNA LAYANAN TRANSPORTASI UMUM MRT JAKARTA

MUHAMMAD NABIL NUFAIL PRIBADI

NIM. 1910511106

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

JUNI 2024

SKRIPSI

PERBANDINGAN HASIL PENERAPAN ALGORITMA KLASIFIKASI DAN *NATURAL LANGUAGE PROCESSING* TERHADAP DATA KEPUASAN PENGGUNA LAYANAN TRANSPORTASI UMUM MRT JAKARTA

MUHAMMAD NABIL NUFAIL PRIBADI

NIM. 1910511106

FAKULTAS ILMU KOMPUTER

UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAKARTA

JUNI 2024

PERNYATAAN ORISINALITAS

Skripsi ini adalah hasil karya sendiri dan semua sumber yang dikutip maupun yang ditunjuk telah saya nyatakan dengang benar:

Nama	: Muhammad Nabil Nufail Pribadi
NIM.	: 1910511106
Program Studi	: S1 Informatika
Tanggal	: 1 Juli 2024
Judul Skripsi/TA.	: **Perbandingan Hasil Penerapan Algoritma Klasifikasi Dan *Natural Language Processing* Terhadap Data Kepuasan Pengguna Layanan Transportasi Umum MRT Jakarta**

Bilamana pada kemudian hari ditemukan ketidaksesuaian dengan pernyataan saya ini, maka saya bersedia dituntut dan diproses sesuai dengan ketentuan yang berlaku.

Jakarta 1 Juli 2024

Yang Menyatakan,

Muhammad Nabil Nufail Pribadi

PERNYATAAN PERSETUJUAN PUBLIKASI TUGAS AKHIR UNTUK KEPENTINGAN AKADEMIS

Sebagai civitas akademika Universitas Pembangunan Nasional Veteran Jakarta, saya yang bertanda tangan di bawah ini:

Nama	: Muhammad Nabil Nufail Pribadi
NIM.	: 1910511106
Fakultas	: Ilmu Komputer
Program Studi	: S1 Informatika

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Pembangunan Nasional Veteran Jakarta Hak Bebas Royalti Non eksklusif (*Non-exclusive Royalty Free Right*) atas karya tulis ilmiah saya yang dipublikasikan dengan judul:

Perbandingan Hasil Penerapan Algoritma Klasifikasi Dan *Natural Language Processing* Terhadap Data Kepuasan Pengguna Layanan Transportasi Umum MRT Jakarta

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti ini Universitas Pembangunan Nasional Veteran Jakarta berhak menyimpan, mengalih media atau memformatkan, mengelola dalam bentuk pangkalan data (*database*), merawat, dan mempublikasikan Skripsi/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 : Jakarta
Pada tanggal : 1 Juli 2024

Yang Menyatakan,

Muhammad Nabil Nufail Pribadi

LEMBAR PENGESAHAN

Skripsi/Tugas Akhir ini diajukan oleh:

Nama	: Muhammad Nabil Nufail Pribadi
NIM.	: 1910511106
Program Studi	: S1 Informatika
Judul Skripsi/TA.	: Perbandingan Hasil Penerapan Algoritma Klasifikasi Dan *Natural Language Processing* Terhadap Data Kepuasan Pengguna Layanan Transportasi Umum MRT Jakarta

Telah berhasil dipertahankan di hadapan Tim Penguji dan diterima sebagai bagian persyaratan yang diperlukan untuk memperoleh gelar Sarjana pada Program Studi S1 Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional Veteran Jakarta.

Dr. Widya Cholil, S.Kom., M.IT.
Penguji I

I Wayan Widi Pradnyana, S.Kom. M.TI.
Penguji II

Iin Ernawati, S.Kom., M.Si.
Pembimbing

Prof. Dr. Ir. Supriyanto, M.Sc., IPM.
Dekan

Dr. Widya Cholil, S.Kom., M.IT.
Kepala Program Studi

Ditetapkan di : Jakarta
Tanggal Ujian : 27 Juni 2024

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT atas berkat dan rahmatnya, sehingga penulis mampu menyelesaikan penulisan proposal skripsi ini dengan lancar. Penulisan proposal skripsi dengan judul "Perbandingan Hasil Penerapan Algoritma Klasifikasi Dan *Natural Language Processing* Terhadap Data Kepuasan Pengguna Layanan Transportasi Umum MRT Jakarta" telah disusun sejak bulan Agustus 2023.

Dalam pelaksanaan penelitian ini, serta dalam penulisan proposal ini, penulis menyadari bahwa terdapat pihak lain yang turut membantu melancarkan penyelesaiannya. Untuk itu, penulis akan mengucapkan terima kasih kepada:

1. Orang tua penulis, yang telah memfasilitasi penulis dengan segala kebutuhan, sehingga penulis mampu menyelesaikan proposal ini dengan lancar.
2. Ibu Iin Ernawati, S. Kom., M.Si. selaku dosen pembimbing skripsi penulis yang telah membantu penulis selama penelitian berlangsung.
3. Bapak Prof. Dr. Ir. Supriyanto, M.Sc., IPM. selaku Dekan Fakultas Ilmu Komputer.
4. Ibu Dr. Widya Cholil, S. Kom., M. IT. selaku Kepala Program Studi Informatika.
5. Teman-teman penulis yang telah membantu dengan dukungan dan sedikit arahan yang diberikan kepada penulis.
6. Pihak lainnya yang turut membantu penulis dalam penyusunan skripsi ini yang tidak tercantum di atas, penulis sampaikan ucapan terima kasih.

Jakarta, 20 Mei 2024

Penulis

Abstrak

Seiring dengan bertambahnya jumlah penduduk dan terbatasnya lahan untuk tempat tinggal di DKI Jakarta, sudah tidak memungkinkan bagi setiap masyarakat untuk memiliki transportasi pribadi masing-masing. Hal ini dapat dibuktikan dengan kemacetan pada DKI Jakarta dengan penyebab terbesar berasal dari transportasi pribadi. MRT Jakarta merupakan salah satu solusi pemerintah DKI Jakarta dalam upaya mengatasi kemacetan. Namun tingkat kemacetan Jakarta kembali meningkat pada tahun 2023 dikarenakan bertambahnya kembali jumlah transportasi pribadi pada lalu lintas. Oleh karena itu, penulis memutuskan untuk melakukan klasifikasi terhadap kepuasan pengguna layanan transportasi umum MRT Jakarta untuk mengetahui alasan yang menyebabkan masyarakat bersedia atau enggan memilih untuk memanfaatkan sarana transportasi MRT Jakarta, yang diperoleh dari media sosial X, memanfaatkan *natural language processing* dan algoritma *Support Vector Machine*, algoritma *Random Forest Classifier*, serta algoritma *Logistic Regression* multinomial. Data tersebut berjumlah sebanyak 525 *post*, dengan kategori ‘positif’ sebanyak 222 data, kategori ‘negatif’ sebanyak 185 data, dan kategori ‘netral’ sebanyak 118 data. Dengan pembagian terhadap dataset berdasarkan perbandingan 80% data latih, dan 20% data uji, model klasifikasi dengan hasil paling akurat pada penelitian ini, yaitu model dengan algoritma *Random Forest Classifier*, menggunakan parameter terbaik yang diperoleh melalui teknik *hyperparameter tuning*, dengan nilai *class_weight='balanced'*, nilai *max_depth*=350, nilai *min_samples_split*=5, serta nilai *n_estimators*=200, menghasilkan nilai *accuracy* sebesar 81%, nilai *precision* sebesar 82.3%, nilai *recall* sebesar 81%, dan nilai *f1-score* sebesar 81.5%, serta berhasil secara akurat memprediksi 4 dari 4 sampel data baru berdasarkan target kelasnya.

Kata Kunci: *Natural Language Processing*, Klasifikasi, MRT Jakarta

Abstract

Along with the increasing population and limited amount of land for housing in DKI Jakarta, it is no longer possible for each person to privately own a car-based transportation. This can be proven by the occurrences of traffic jams in DKI Jakarta with the biggest cause coming from the usage of private transportation, namely cars. MRT Jakarta is one of the solutions proposed by the government of DKI Jakarta as one of their efforts to overcome traffic jams. However, Jakarta's congestion level increases again in 2023 due to the increased usage of private transportation. Therefore, the author decided to classify the opinion of users of MRT Jakarta to find out the reasons why people choose or are reluctant to utilize MRT Jakarta as their main method of transportation, obtained from the social media X, utilizing natural language processing and machine learning algorithm such as Support Vector Machine, Random Forest Classifier, and multinomial Logistic Regression. The data amounts to 525 posts, with 222 in the 'positive' category, 185 in the 'negative' category, and 118 in the 'neutral' category. By dividing the dataset based on a comparison of 80% training data and 20% test data, the classification model with the highest accuracy score in this research, namely the model with the Random Forest Classifier algorithm, using the best parameters obtained through the usage of hyperparameter tuning technique, with the value of class_weight='balanced', value max_depth=350, min_samples_split=5, and n_estimators=200, resulting in an accuracy value of 81%, a precision value of 82.3%, a recall value of 81%, and an f1-score value of 81.5%, and successfully predicted 4 of 4 new data samples based on their target classes.

Keywords: *Natural Language Processing*, *Classification*, MRT Jakarta

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR LAMPIRAN