

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*