

Perbandingan Akurasi *Word Embedding Word2Vec, GloVe, dan FastText* Menggunakan
Support Vector Machine untuk Analisis Sentimen Ulasan Aplikasi Spotify

Margaretha Anjani

Program Studi Informatika, Fakultas Ilmu Komputer,
Universitas Pembangunan Nasional Veteran Jakarta

Email : margarethanjani@upnvj.ac.id

ABSTRAK

Spotify merupakan salah satu aplikasi yang digunakan sebagai platform layanan streaming audio digital yang menyajikan berbagai musik dan podcast dan dapat diunduh dengan gratis pada Google Play Store. Ulasan merupakan sebuah fitur Google Play Store yang dapat dimanfaatkan oleh pengguna untuk memberikan penilaian terhadap sebuah aplikasi. Ulasan yang dapat diterima oleh aplikasi dapat mempengaruhi pengguna yang akan mengunduh aplikasi tersebut. Karakteristik teks ulasan yang tidak terstruktur akan menjadi sebuah tantangan dalam proses pemrosesan teks. Untuk menghasilkan sentimen analisis yang valid dibutuhkan adanya penerapan *word embedding*. Data set yang dimiliki dibagi dengan perbandingan 80:20 untuk data training dan data testing. Metode yang digunakan untuk ekspansi fitur *Word2Vec, GloVe, dan FastText* dan metode yang digunakan dalam klasifikasi adalah *Support Vector Machine (SVM)*. Ketiga metode *word embedding* tersebut dipilih karena dapat menangkap makna yang semantik, sintatik, serta konteks pada sekitar kata bila dibandingkan *feature engineering* tradisional seperti *Bag of Word*. Hasil evaluasi performa terbaik menunjukkan model GloVe menghasilkan kinerja terbaik dibandingkan dengan word embedding lainnya dengan nilai akurasi sebesar 85%, nilai presisi sebesar 90%, nilai *recall* 79%, dan *f1-score* 85%.

Kata kunci: word2vec, glove, fasttext, support vector machine, klasifikasi, analisis sentimen.

Comparison Accuracy of Word Embedding Word2Vec, GloVe, dan FastText Using Support Vector Machine for Sentiment Analysis Spotify App Reviews

Margaretha Anjani

Program Studi Informatika, Fakultas Ilmu Komputer,
Universitas Pembangunan Nasional Veteran Jakarta

Email : margarethanjani@upnvj.ac.id

ABSTRACT

Spotify is an application that is used as a digital audio streaming service platform that presents a variety of music and podcasts and can be downloaded for free on the Google Play Store. Reviews are a feature of the Google Play Store that can be used by users to rate an application. The reviews that an app can receive may affect the users who will download the app. Characteristics of unstructured review texts will be a challenge in the text processing process. To produce a valid sentiment analysis, it is necessary to apply word embedding. The data set that is owned is divided by a ratio of 80:20 for training data and testing data. The method used for the expansion of the Word2Vec, GloVe, and FastText features and the method used in the classification is the Support Vector Machine (SVM). The three word embedding methods were chosen because they can capture the semantic, syntactic, and contextual meanings around words when compared to traditional engineering features such as Bag of Word. The best performance evaluation results show that the GloVe model produces the best performance compared to other word embeddings with an accuracy value of 85%, a precision value of 90%, a recall value of 79%, and an f1-score of 85%.

Keywords: *word2vec, glove, fasttext, support vector machine, classification, sentiment analysis.*