

INTISARI

Penelitian ini melakukan perbandingan metode *Word2Vec-SVM* dan *TF-IDF-SVM*, untuk melakukan klasifikasi. Penelitian ini dibagi menjadi 5 tahap yaitu pengumpulan data, *labeling*, *preprocessing*, *testing* dan hasil. Penelitian ini menggunakan (*app script*) untuk *scraping* data yang didapat dari opini masyarakat di *Channel youtube* tentang *metaverse*. *Channel youtube* yang di gunakan adalah Deddy Corbuzier, Dr. Indrawan Nugroho, Sandiuno TV, dan Raymond Chin dengan kata kunci “*metaverse*”. Data yang dikumpulkan mulai dari bulan Januari 2021 hingga September 2022. Data yang digunakan sebanyak 5000 data dengan label positif atau negatif.

Hasil pengujian *SVM-TFIDF* dengan *Undersampling* menunjukkan bahwa nilai akurasi tertinggi di hasilkan dari kernel *RBF* dengan 74% *accuracy*, 74% *precision*, 73% *recall*, dan 73% *f1-score* pada pembagian data 75% data latih dan 25% data uji. Sedangkan dari hasil pengujian *SVM-WORD2VEC* tanpa *SMOTE* menghasilkan akurasi tertinggi pada kernel *polynomial* 61% *accuracy*, 60% *precision*, 63% *recall*, dan 61% *f1-score* pada pembagian data 75% data latih dan 25% data uji. Berdasarkan hasil pengujian yang dilakukan, dapat disimpulkan bahwa *SVM-TFIDF* memiliki performa yang lebih baik daripada *SVM-WORD2VEC* dalam proses klasifikasi data.

Kata Kunci: Klasifikasi Opini Masyarakat, *Metaverse*, *Word2Vec-SVM*, *TF-IDF-SVM*, *Smote*, *Undersampling*.

ABSTRACT

This research conducted a comparison between Word2Vec-SVM and TF-IDF-SVM methods for classification purposes. The study was divided into five stages, namely data collection, labeling, preprocessing, testing, and analysis of results. The research utilized data scraping using an app script from public opinions on the YouTube channels discussing the metaverse. The YouTube channels used in the study were Deddy Corbuzier, Dr. Indrawan Nugroho, Sandiuno TV, and Raymond Chin, with the keyword "metaverse". The data collection period spanned from January 2021 to September 2022. A total of 5000 data points were collected with positive or negative labels.

The testing results of SVM-TFIDF with Undersampling indicated that the highest accuracy was achieved using the RBF kernel with 74% accuracy, 74% precision, 73% recall, and 73% f1-score when the data was divided into 75% training data and 25% testing data. On the other hand, the testing results of SVM-WORD2VEC without SMOTE showed the highest accuracy with the polynomial kernel, achieving 61% accuracy, 60% precision, 63% recall, and 61% f1-score when the data was divided into 75% training data and 25% testing data. Based on the conducted testing, it can be concluded that SVM-TFIDF outperforms SVM-WORD2VEC in the data classification process.

Keywords: *Classification of Community Opinion, Metaverse, Word2Vec-SVM, TF-IDF-SVM, Smote, Undersampling.*