

ABSTRAK

Pajak merupakan sumber pendapatan penting bagi negara yang berkontribusi pada peningkatan kesejahteraan masyarakat. Indeks kepercayaan masyarakat terhadap otoritas perpajakan memainkan peran penting dalam peningkatan penerimaan pajak. Oleh karena itu, penting untuk mengukur indeks kepercayaan tersebut. Salah satu metode yang dapat digunakan adalah analisis sentimen, yang membantu memahami persepsi masyarakat terkait peraturan perpajakan, layanan perpajakan, kinerja instansi perpajakan, dan kebijakan perpajakan yang diterapkan oleh. Salah satu tujuan penelitian ini adalah mengukur tingkat sentimen masyarakat pengguna Twitter terhadap topik perpajakan di Indonesia. Analisis sentimen pada penelitian ini melibatkan enam tahap utama, yaitu proses pengumpulan data, pra-proses pengolahan data, pembagian *dataset*, ekstraksi fitur, klasifikasi, dan evaluasi. Proses Klasifikasi dilakukan menggunakan model *Multinomial Naive Bayes*, dengan rasio 80% data latih dan 20% data uji. Hasil penelitian menunjukkan bahwa sebanyak 89,65% *tweet* terkait topik perpajakan di Indonesia memiliki kecenderungan emosional yang negatif. Evaluasi model klasifikasi dilakukan dengan membandingkan dua skenario pengujian, yaitu pada data awal dan data hasil *random undersampling*. Klasifikasi pada data awal mencapai akurasi sebesar 89,97%, presisi sebesar 46,68%, dan sensitivitas sebesar 33,61%. Sedangkan pada data hasil *random undersampling*, akurasi mencapai 53,28%, presisi sebesar 52,66%, dan sensitivitas sebesar 52,52%. Hasil analisis menunjukkan adanya perbedaan yang signifikan antara klasifikasi menggunakan data awal dan data hasil *random undersampling*. Teknik *random undersampling* mampu mengubah jumlah data dalam setiap kategori sentimen menjadi seimbang. Namun, model masih menghadapi tantangan dalam mengklasifikasikan data sentimen yang pada kelas positif dan netral. Hal tersebut diakibatkan dominasi data *tweet* dengan kelas negatif.

Kata Kunci : Sentimen Pajak, *Indonesian Sentimen Lexicon*, *Colloquial Indonesian Lexicon* , TF-IDF, Multinomial Naive Bayes.



ABSTRACT

Taxes represent a crucial revenue source for countries, contributing to enhancing society's well-being. Public trust in taxation authorities plays a significant role in boosting tax revenue. Therefore, measuring this trust index becomes essential. One applicable method is sentiment analysis, aiding in comprehending public perceptions regarding tax regulations, tax services, tax agency performance, and government-implemented tax policies. The research aims to measure the sentiment level of Indonesian Twitter users towards taxation topics. Sentiment analysis in this study comprises six main stages: data collection, data preprocessing, dataset splitting, feature extraction, classification, and evaluation. Classification process employs the Multinomial Naive Bayes model, with an 80% training data and a 20% test data ratio. Research results indicate that approximately 89.65% of tweets related to taxation topics in Indonesia exhibit a negative emotional tendency. Classification model's evaluation is conducted by comparing two testing scenarios: initial data and data after random undersampling. Classification accuracy achieved on initial data is 89.97%, with precision at 46.68%, and sensitivity at 33.61%. Meanwhile, on data after random undersampling, accuracy reaches 53.28%, precision at 52.66%, and sensitivity at 52.52%. Analysis reveals a significant difference between classification using initial data and random undersampling data. Random undersampling technique successfully balances the number of data points in each sentiment category. However, model still faces challenges in classifying sentiment data in positive and neutral classes, mainly due to dominance of negative-class tweets.

Keywords: Taxes Sentiment Analysis, *Indonesian Sentimen Lexicon*, *Colloquial Indonesian Lexicon*, TF-IDF, Multinomial Naive Bayes.

