

## ABSTRAK

Analisis sentimen merupakan salah satu cabang dari pemrosesan bahasa alami (*Natural Language Processing*) yang bertujuan untuk mengidentifikasi opini atau emosi seseorang terhadap suatu topik, produk, atau kebijakan. Dalam penelitian ini, dilakukan analisis sentimen opini masyarakat terhadap kebijakan Tabungan Perumahan Rakyat (Tapera) yang menjadi isu publik di Indonesia. Model klasifikasi yang digunakan adalah *Naive Bayes*, yang kemudian menggunakan *Particle Swarm Optimization* (PSO) untuk mendapatkan parameter terbaik, serta diterapkan SMOTE (*Synthetic Minority Over-sampling Technique*) untuk menangani ketidakseimbangan kelas pada data latih.

Tahapan preprocessing mencakup pembersihan teks, penghapusan stopwords, tokenisasi, normalisasi kata, penanganan negasi, stemming, serta pembobotan menggunakan TF-IDF. Setelah dilakukan peningkatan performa dengan PSO terhadap parameter *alpha* dan *fit\_prior*, diperoleh konfigurasi terbaik yaitu  $\alpha = 0.01$  dan *fit\_prior* = False, yang menghasilkan akurasi *cross-validation* sebesar 92,97%. Model kemudian dilatih ulang dan diuji, dengan hasil performa pada data latih mencapai akurasi, *precision*, *recall*, dan *F1-score* sebesar 94%. Sementara pada data uji, evaluasi menggunakan *confusion matrix* menunjukkan bahwa model mampu mengklasifikasikan ketiga kelas sentimen (positif, netral, negatif) secara akurat dan seimbang.

Hasil penelitian ini menunjukkan bahwa kombinasi antara algoritma *Naive Bayes*, teknik SMOTE, dan peningkatan performa menggunakan PSO dapat meningkatkan performa klasifikasi dalam analisis sentimen kebijakan pemerintah secara signifikan.

**Kata Kunci:** Analisis Sentimen, Naive Bayes, PSO, SMOTE, Tapera, TF-IDF.

## **ABSTRACT**

*Sentiment analysis is a branch of Natural Language Processing (NLP) aimed at identifying a person's opinions or emotions toward a topic, product, or policy. In this study, sentiment analysis was conducted on public opinions regarding the Public Housing Savings Program (Tapera), which has become a public issue in Indonesia. The classification model used is Naive Bayes, which is then optimized using Particle Swarm Optimization (PSO) to obtain the best parameters, and applied with the Synthetic Minority Over-sampling Technique (SMOTE) to address class imbalance in the training data.*

*The preprocessing stages include text cleaning, stopword removal, tokenization, word normalization, negation handling, stemming, and term weighting using TF-IDF. After performance enhancement with PSO for the parameters  $\alpha$  and  $\text{fit\_prior}$ , the best configuration obtained was  $\alpha = 0.01$  and  $\text{fit\_prior} = \text{False}$ , yielding a cross-validation accuracy of 92.97%. The model was then retrained and tested, achieving performance metrics on the training data—accuracy, precision, recall, and F1-score—of 94%. Meanwhile, evaluation on the test data using a confusion matrix showed that the model was able to classify all three sentiment classes (positive, neutral, negative) accurately and evenly.*

*The results of this study indicate that the combination of the Naive Bayes algorithm, SMOTE technique, and performance optimization using PSO can significantly improve classification performance in government policy sentiment analysis.*

**Keywords:** *Sentiment Analysis, Naive Bayes, PSO, SMOTE, Tapera, TF-IDF.*