Abstract

The postponement of the appointment of Civil Servant Candidates (CPNS) and Government Employees with Work Agreements (PPPK) has generated various public responses, many of which were conveyed through social media X. This study aims to analyze public sentiment patterns towards this policy and compare the performance of the Naive Bayes and Random Forest algorithms in classifying public sentiment. The study uses a quantitative approach with an experimental method based on the Cross Industry Standard Process for Data Mining (CRISP-DM) framework. Data were collected through web scraping techniques on social media X with keywords related to CPNS and PPPK during the period of March 5–16, 2025, with an initial number of 1,423 comments, which after the duplication removal process became 1,180 data. The preprocessing stages include cleaning, case folding, normalization, tokenization, and stopword removal. Sentiment labeling is done using the InSet dictionary, while feature extraction uses the TF-IDF method. The model was built using the Naive Bayes and Random Forest algorithms with a data division of 80% training and 20% testing. Model performance evaluation was carried out using a confusion matrix and ROC curve. The results showed that public sentiment was dominated by negative sentiment, accounting for 61.27%. Random Forest achieved 78% accuracy with an AUC of 0.88, higher than Naive Bayes, which had 75% accuracy and an AUC of 0.86. These results indicate that Random Forest is more optimal in classifying sentiment in Indonesian-language social media text data.

Keywords: *Sentiment Analysis; CPNS; PPPK; Naive Bayes; Random Forest*

ABSTRAK

Penundaan pengangkatan Calon Pegawai Negeri Sipil (CPNS) dan Pegawai Pemerintah dengan Perjanjian Kerja (PPPK) menimbulkan beragam respons publik yang banyak disampaikan melalui media sosial X. Penelitian ini bertujuan untuk menganalisis pola sentimen publik terhadap kebijakan tersebut serta membandingkan kinerja algoritma Naive Bayes dan Random Forest dalam mengklasifikasikan sentimen masyarakat. Penelitian menggunakan pendekatan kuantitatif dengan metode eksperimen berbasis kerangka kerja *Cross Industry Standard Process for Data Mining* (CRISP-DM). Data dikumpulkan melalui teknik *web scraping* pada media sosial X dengan kata kunci terkait CPNS dan PPPK selama periode 5–16 Maret 2025 dengan jumlah awal 1.423 komentar, yang setelah proses penghapusan duplikasi menjadi 1.180 data. Tahapan preprocessing meliputi *cleaning*, *case folding*, *normalisasi*, *tokenisasi*, dan *stopword removal*. Pelabelan sentimen dilakukan menggunakan kamus InSet, sedangkan ekstraksi fitur menggunakan metode TF-IDF. Model dibangun menggunakan algoritma Naive Bayes dan Random Forest dengan pembagian data 80% training dan 20% testing. Evaluasi kinerja model dilakukan menggunakan *confusion matrix* dan kurva *Receiver Operating Characteristic (ROC)*. Hasil penelitian menunjukkan bahwa sentimen publik didominasi oleh sentimen negatif sebesar 61,27%. Random Forest memperoleh akurasi sebesar 78% dengan nilai AUC 0,88, lebih tinggi dibandingkan Naive Bayes dengan akurasi 75% dan AUC 0,86. Hasil ini menunjukkan bahwa Random Forest lebih optimal dalam mengklasifikasikan sentimen pada data teks media sosial berbahasa Indonesia.

Kata Kunci: Analisis Sentimen, CPNS, PPPK, Naive Bayes, Random Forest