

Analisis Sentimen dan Pemodelan Topik dengan Data Media Sosial Twitter: Studi Kasus Komisi Pemberantasan Korupsi = Sentiment Analysis and Topic Modelling using Twitter Social Media Data: A Case Study of the Corruption Eradication Commission

Dian Isnaeni Nurul Afra, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20523321&lokasi=lokal>

Abstrak

Komisi Pemberantasan Korupsi (KPK) memiliki kewenangan dalam melakukan pendaftaran dan pemeriksaan terhadap Laporan Harta Kekayaan Penyelenggara Negara (LHKPN). Pelaporan ini berfungsi untuk melakukan pengawasan kejujuran, integritas, dan deteksi kemungkinan adanya tindakan memperkaya diri secara melawan hukum oleh pejabat publik. Publikasi LHKPN sering menimbulkan prasangka negatif dan kecurigaan publik terhadap laporan harta kekayaan pejabat yang mengakibatkan kekhawatiran pejabat untuk melaporkan harta kekayaan secara lengkap dan benar. Persepsi ini menjadi kontraproduktif dengan upaya pencegahan korupsi yang dilakukan oleh KPK apabila tidak direspon dengan cepat. Penelitian ini bertujuan untuk membuat model analisis sentimen dan pemodelan topik yang dapat mengeksplorasi topik dari data media sosial Twitter. Indonesia memiliki jumlah pengguna aktif terbesar keenam di dunia dengan 15,7 juta pengguna yang didominasi kelompok usia 25-34 tahun. Dataset sejumlah 881 data diambil dari Twitter dengan kata kunci "lhkpn" dan "harta kekayaan pejabat" pada periode 1 Agustus sampai 5 November 2021. Penelitian ini mengeksplorasi beberapa algoritma klasifikasi, representasi fitur unigram, bigram, dan trigram dengan CountVectorizer dan TFIDF, serta metode oversampling SMOTE. Algoritma klasifikasi dengan performa paling baik pada penelitian ini adalah Multilayer Perceptron dengan fitur unigram CountVectorizer dan metode oversampling dengan accuracy 76,60%, precision 78,19%, recall 76,60%, dan F1 score 76,95%. Hasil pemodelan topik menggunakan Latent Dirichlet Allocation pada kategori 'negatif' didominasi ekspresi kekecewaan dan kemarahan masyarakat terhadap meningkatnya harta kekayaan pejabat selama masa pandemi Covid-19 yang berbanding terbalik dengan meningkatnya utang negara dan kesulitan yang dihadapi masyarakat selama pandemi. Topik yang dihasilkan pada kategori 'positif' cukup beragam mulai dari aturan untuk melakukan pembuktian terbalik, usulan mengenai kewajiban pelaporan dan sanksi, permintaan untuk membuka laporan kekayaan kepada publik, serta pembahasan mengenai kewajaran penambahan harta kekayaan yang disebabkan oleh meningkatnya nilai aset tidak bergerak.

.....The Corruption Eradication Commission (KPK) has the authority to register and examine Public Officials Wealth Reports (LHKPN). This report serves to monitor honesty, integrity, and detect the possibility of illegal enrichment by public officials. Publication of LHKPN often creates negative prejudice and public suspicion of official wealth reports, which causes officials to worry about reporting assets completely and correctly. This perception is counterproductive to the efforts to prevent corruption carried out by the KPK if it is not responded to quickly. This study aims to create a sentiment analysis model and topic modelling that can explore topics from Twitter social media data. Indonesia has the sixth-largest number of active users in the world with 15.7 million users, dominated by the 25-34 year age group. A dataset of 881 data was taken from Twitter with the keywords "lhkpn" and "official assets" in the period August 1 to November 5, 2021. This study explores several classification algorithms, representation of

unigram, bigram, and trigram features with CountVectorizer and TFIDF, as well as SMOTE oversampling methods. The classification algorithm with the best performance is the Multilayer Perceptron with the unigram CountVectorizer feature and the oversampling method with 76.60% accuracy, 78.19% precision, 76.60% recall, and 76.95% F1 score. The results of topic modelling using Latent Dirichlet Allocation in the 'negative' category are dominated by expressions of public disappointment and anger towards the increase in official wealth during the Covid-19 pandemic which is inversely proportional to the increase in state debt and the difficulties faced by the community during the pandemic. The topics generated in the 'positive' category are quite diverse, starting from the rules for conducting reverse verification, proposals on reporting obligations and sanctions, requests to disclose wealth reports to the public, as well as discussions on the reasonableness of adding to assets caused by the increase in the value of immovable assets.