

ABSTRAK

Penelitian ini bertujuan untuk memodelkan opini publik terhadap kebijakan Menteri Keuangan terkait Bea dan Cukai melalui analisis komentar pengguna Instagram. Data dikumpulkan dari berbagai akun pada periode Oktober 2025 hingga Februari 2026 sebanyak 18.859 komentar dan setelah melalui *preprocessing* menghasilkan 13.998 data bersih. Karakteristik komentar yang singkat, informal, mengandung slang, serta campuran bahasa Indonesia dan Inggris menjadi tantangan dalam analisis sehingga diperlukan metode yang mampu menangkap makna semantik. Penelitian ini menggunakan metode *Top2Vec* berbasis *unsupervised learning* yang mengintegrasikan *embedding* semantik, reduksi dimensi menggunakan *Uniform Manifold Approximation and Projection* (UMAP), serta clustering berbasis densitas menggunakan *Hierarchical Density-Based Spatial Clustering of Applications with Noise* (HDBSCAN). Selain itu, diterapkan strategi parameter tuning adaptif berdasarkan ukuran dataset, yaitu data kecil (5, 10, 3), sedang (15, 15, 5), dan besar (45, 25, 10), yang terbukti meningkatkan stabilitas klasterisasi dan pengendalian noise. Pengujian dilakukan melalui dua skenario, yaitu dengan UMAP dan tanpa UMAP, serta dievaluasi menggunakan *coherence score*, *c_v* dan *c_npmi*, *topic diversity*, *information gain*, dan proporsi *noise*. Hasil menunjukkan bahwa skenario dengan UMAP menghasilkan 46 topik dengan noise 2.883 komentar, *c_v* 0.3694, *c_npmi* -0.2794, *diversity* 0.4183, *IG* 0.0131, sedangkan tanpa UMAP menghasilkan 6 topik dengan noise 12.211 komentar, *c_v* 0.5353, *c_npmi* -0.4599, *diversity* 0.8750, *IG* 0,0418. Temuan ini menunjukkan adanya *trade-off* antara granularitas topik dan tingkat noise, di mana UMAP menghasilkan topik lebih rinci dengan noise lebih terkendali. Topik dominan meliputi pungutan liar, pajak dan cukai, rokok ilegal, serta respons publik terhadap kebijakan. Secara keseluruhan, metode *Top2Vec* dengan tuning adaptif efektif dalam memodelkan opini publik pada data komentar yang tidak terstruktur dan memberikan kontribusi dalam meningkatkan kualitas pemodelan topik.

Kata kunci: *Top2Vec*, UMAP, HDBSCAN, parameter tuning, opini publik, komentar Instagram, bea dan cukai.

ABSTRACT

This study aims to model public opinion regarding the Minister of Finance's policies on Customs and Excise through the analysis of Instagram user comments. The data were collected from various accounts during the period of October 2025 to February 2026, totaling 18,859 comments, and after preprocessing resulted in 13,998 clean data points. The characteristics of the comments—being short, informal, containing slang, and involving a mix of Indonesian and English—pose challenges for analysis, thus requiring a method capable of capturing semantic meaning. This study employs the Top2Vec method based on unsupervised learning, which integrates semantic embedding, dimensionality reduction using Uniform Manifold Approximation and Projection (UMAP), and density-based clustering using Hierarchical Density-Based Spatial Clustering of Applications with Noise (HDBSCAN). Additionally, an adaptive parameter tuning strategy is applied based on dataset size, namely small (5, 10, 3), medium (15, 15, 5), and large (45, 25, 10), which has proven to improve clustering stability and noise control. The evaluation was conducted through two scenarios: with UMAP and without UMAP, and assessed using coherence scores (c_v and c_{npmi}), topic diversity, information gain, and noise proportion. The results show that the scenario with UMAP produced 46 topics with 2,883 noisy comments, c_v of 0.3694, c_{npmi} of -0.2794, diversity of 0.4183, and IG of 0.0131. Meanwhile, the scenario without UMAP generated 6 topics with 12,211 noisy comments, c_v of 0.5353, c_{npmi} of -0.4599, diversity of 0.8750, and IG of 0.0418. These findings indicate a trade-off between topic granularity and noise level, where UMAP produces more detailed topics with better-controlled noise. The dominant topics include illegal levies, taxation and excise, illegal cigarettes, and public responses to policy. Overall, the Top2Vec method with adaptive tuning is effective in modeling public opinion from unstructured comment data and contributes to improving topic modeling quality.

Keywords: *Top2Vec, UMAP, HDBSCAN, parameter tuning, public opinion, Instagram comments, customs and excise.*