

ABSTRAK

ANUGRAH RESKY SAMUDRA R, Analisis Sentimen dan Clustering Komentar *Twitter* Menggunakan Metode *Lexicon-Based* dan Algoritma *K-Means*. (Dibimbing Oleh Chyquitha Danuputri, S.Kom., M.Kom. dan Titin Wahyuni, S.Pd., M.T.

Media sosial *Twitter* menjadi ruang publik yang aktif digunakan masyarakat untuk menyampaikan opini terkait isu kebijakan publik dan ekonomi nasional. Volume komentar yang besar dan bersifat tidak terstruktur menuntut adanya pendekatan komputasional untuk mengekstraksi informasi yang bermakna. Penelitian ini bertujuan untuk menganalisis sentimen publik serta mengelompokkan topik diskusi utama pada komentar *Twitter* menggunakan pendekatan *lexicon-based sentiment analysis* dan algoritma *K-Means* clustering.

Dataset penelitian terdiri dari 198 komentar *Twitter* yang dikumpulkan dari akun media nasional @kumparan dan @Metro_TV menggunakan ekstensi *TwCommentExport*. Data kemudian melalui tahapan *pre-processing* yang meliputi *case folding*, *cleaning*, *stopword removal*, tokenisasi, dan *stemming*. Analisis sentimen dilakukan dengan pendekatan berbasis *lexicon* untuk mengklasifikasikan komentar ke dalam kategori positif, negatif, dan netral. Selanjutnya, data direpresentasikan menggunakan TF-IDF dan dikelompokkan menggunakan algoritma *K-Means* dengan evaluasi *Silhouette Score*.

Hasil penelitian menunjukkan bahwa mayoritas komentar memiliki sentimen netral (87,88%), diikuti oleh sentimen positif dan negatif. Proses clustering menghasilkan konfigurasi optimal pada $K = 2$ dengan nilai *Silhouette Score* sebesar 0,39, yang membentuk dua topik utama diskusi, yaitu kebijakan anggaran dan efektivitas belanja publik, serta utang negara dan kebijakan fiskal. Penelitian ini membuktikan bahwa integrasi analisis sentimen dan *clustering* efektif dalam memetakan opini publik serta mengungkap pola diskursus ekonomi nasional di media sosial.

Kata Kunci: Analisis Sentimen, *Lexicon-Based*, *K-Means*, *Twitter*, *Text Mining*.

ABSTRACT

ANUGRAH RESKY SAMUDRA R, *Sentiment Analysis and Clustering of Twitter Comments Using Lexicon-Based Method and K-Means Algorithm. (Supervised by Chyquitha Danuputri, S.Kom., M.Kom. and Titin Wahyuni, S.Pd., M.T.)*

Twitter has become an active public sphere where people express opinions regarding public policy and national economic issues. The large volume of unstructured textual data requires a computational approach to extract meaningful insights. This study aims to analyze public sentiment and identify dominant discussion topics in Twitter comments using a lexicon-based sentiment analysis approach and the K-Means clustering algorithm.

The dataset consists of 198 Twitter comments collected from the national media accounts @kumparan and @Metro_TV using the TwCommentExport extension. The data were processed through several pre-processing stages, including case folding, cleaning, stopword removal, tokenization, and stemming. Sentiment analysis was conducted using a lexicon-based method to classify comments into positive, negative, and neutral categories. Furthermore, textual data were transformed using TF-IDF representation and clustered using the K-Means algorithm, with clustering quality evaluated through the Silhouette Score.

The results indicate that most comments exhibit a neutral sentiment (87.88%), followed by positive and negative sentiments. The optimal clustering configuration was achieved at $K = 2$, yielding a Silhouette Score of 0.39, which formed two main discussion topics: public budget policy and expenditure effectiveness, and national debt and fiscal policy. This study demonstrates that integrating sentiment analysis and clustering techniques is effective in mapping public opinion and uncovering economic discourse patterns on social media platforms.

Keywords: *Sentiment Analysis, Lexicon-Based, K-Means, Twitter, Text Mining.*