

ABSTRAK

Raif, M Irfan. 2024. Otomatisasi Pendeteksi Kata Baku dan Tidak baku Pada Data Twitter Berbasis KBBI Menggunakan Metode *Lexicon-Based*, Skripsi. Tanjungpinang: Jurusan Teknik Informatika, Fakultas Teknik dan Teknologi Kemaritiman, Univeristas Maritim Raja Ali Haji. Pembimbing I: Tekad Matulatan, S.Sos., S.Kom., M.Inf.Tech. Pembimbing II: Nuraisa Novia Hidayati, M.Kom

Berdasarkan Kamus Besar Bahasa Indonesia (KBBI), penelitian ini berkonsentrasi pada pembuatan sistem deteksi otomatis yang dapat membedakan kata baku dan tidak baku dalam data Twitter. Penelitian ini sangat penting untuk menjamin komunikasi yang efektif karena Twitter adalah *platform* media sosial yang sering menggunakan kata-kata yang tidak baku karena terdapat batasan penggunaan kata. Melalui normalisasi kata-kata tidak baku, penelitian ini membantu *preprocessing* dan analisis *tweet*, yang merupakan langkah penting dalam klasifikasi teks media sosial. Sistem otomatis yang dikembangkan oleh penelitian ini tidak hanya memudahkan peneliti untuk mengidentifikasi penggunaan kata-kata *slang* atau tidak baku, tetapi juga meningkatkan kualitas komunikasi dan pemahaman pesan dalam *tweet* yang mengikuti tren bahasa terbaru. Penelitian ini menggunakan metode *lexicon-Based* untuk kamus opini, mengumpulkan data, melakukan *preprocessing*, menemukan bahasa yang tidak baku, menemukan *slang*, dan menghapus kata berimbuhan. Metode ini membantu analisis sentimen dalam pengolahan teks dan memastikan bahwa hasil klasifikasi sentimen pada data Twitter lebih akurat. Hasil percobaan menunjukkan bahwa langkah *preprocessing* meningkatkan akurasi metode penentuan polarisasi secara efektif, Tingkat *accuracy* InSet adalah 66,66% dan *F1-score* adalah 61,40%.

Kata kunci: *Analisis sentimen, InSet, lexicon-based, normalisasi, pendeteksi kata tidak baku, preprocessing, Twitter, X.*

ABSTRACT

Raif, M Irfan. 2024. *Automation of Standard and Non-Standard Word Detection on Twitter Data Based on KBBI Using Lexicon-Based Method*, Thesis. Tanjungpinang: Department of Informatics Engineering, Faculty of Engineering and Maritime Technology, University of Maritim Raja Ali Haji. Advisor: Tekad Matulatan, S.Sos., S.Kom., M.Inf.Tech. Co-advisor: Nuraisa Novia Hidayati, M.Kom.

Based on the Indonesian Dictionary (KBBI), this research focuses on creating an automatic detection system that can distinguish between standard and non-standard words in Twitter data. This study is crucial for ensuring effective communication because Twitter is a social media platform that often uses non-standard words due to character limitations. By normalizing non-standard words, this research aids in the preprocessing and analysis of tweets, which is a critical step in the classification of social media text. The automated system developed by this research not only facilitates researchers in identifying the use of slang or non-standard words but also enhances the quality of communication and understanding of messages in tweets that follow the latest language trends. This study uses a lexicon-based method for the opinion dictionary, collecting data, preprocessing, identifying non-standard language, detecting slang, and removing affixed words. This method helps in sentiment analysis in text processing and ensures that the results of sentiment classification on Twitter data are more accurate. Experimental results show that the preprocessing steps effectively improve the accuracy of the sentiment determination method, with InSet achieving an accuracy rate of 66.66% and an F1-score of 61.40%.

Keywords: *Sentiment analysis, InSet, lexicon-based, normalization, non-standard word detection, preprocessing, Twitter, X.*