

INTISARI

PEMODELAN TOPIK DAN ANALISIS SENTIMEN TERHADAP ISU LINGKUNGAN KERJA MENGGUNAKAN *LATENT DIRICHLET ALLOCATION* (LDA) DAN *NAÏVE BAYES CLASSIFIER* (NBC)

Dinda Hana Zhafira
22/494038/PA/21240

Perkembangan digitalisasi telah meningkatkan jumlah data digital yang dihasilkan masyarakat, terutama melalui media sosial. Berdasarkan data Badan Pusat Statistik (2025), sebanyak 72,78% penduduk Indonesia telah mengakses internet pada tahun 2024. Media sosial, khususnya platform berbasis teks seperti X, menjadi wadah bagi masyarakat untuk mengekspresikan opini terkait berbagai topik, termasuk mengenai lingkungan kerja. Lingkungan kerja merupakan faktor penting yang memengaruhi kesejahteraan, produktivitas, dan kinerja pekerja (Dulloh *et al.*, 2024; Adhistry *et al.*, 2023). Banyaknya opini publik di media sosial menghasilkan data yang besar dan beragam, sehingga diperlukan metode untuk memahami isi dan arah sentimen dari data tersebut.

Pemodelan topik dalam penelitian ini dilakukan menggunakan metode *Latent Dirichlet Allocation* (LDA) dengan penentuan jumlah topik optimal berdasarkan nilai *coherence score*, sedangkan analisis sentimen dilakukan menggunakan algoritma *Naïve Bayes* dengan penerapan teknik *Synthetic Minority Oversampling Technique* (SMOTE) untuk mengatasi ketidakseimbangan data. Pelabelan data dilakukan secara manual dan menghasilkan 2540 *tweet* sentimen netral, 1218 *tweet* sentimen negatif, dan 321 *tweet* sentimen positif dari total 4079 *tweet*. Hasil penelitian menunjukkan bahwa model LDA menghasilkan *coherence score* sebesar 0,453 dengan 10 topik sebagai jumlah topik optimal. Model analisis sentimen *Naïve Bayes* dengan SMOTE menghasilkan performa dengan akurasi 72%, presisi 62%, recall 68%, dan F1-score 63%. Hasil kombinasi pemodelan topik dan analisis sentimen menunjukkan bahwa sebagian besar topik didominasi oleh sentimen netral, meskipun pada beberapa topik ditemukan dominasi sentimen negatif. Visualisasi wordcloud menunjukkan bahwa sentimen positif berkaitan dengan dukungan dan harapan terhadap peluang kerja, sentimen negatif berkaitan dengan tekanan mental serta ketidakpuasan terhadap lingkungan kerja, sedangkan sentimen netral didominasi oleh pembahasan informatif mengenai proses rekrutmen dan pengalaman pencarian kerja.

Kata Kunci – Lingkungan Kerja, Analisis Sentimen, Pemodelan Topik, *Naïve Bayes*, *Latent Dirichlet Allocation* (LDA)

ABSTRACT

TOPIC MODELING AND SENTIMENT ANALYSIS ON WORKPLACE ENVIRONMENT ISSUES USING LATENT DIRICHLET ALLOCATION (LDA) AND NAÏVE BAYES CLASSIFIER (NBC)

Dinda Hana Zhafira
22/494038/PA/21240

The development of digitalization has increased the amount of digital data generated by society, particularly through social media. According to data from the Central Statistics Agency (Badan Pusat Statistik, 2025), 72.78% of Indonesia's population accessed the internet in 2024. Social media platforms, especially text-based platforms such as X, have become a medium for people to express opinions on various topics, including work environment issues. The work environment is an important factor that influences employee well-being, productivity, and performance (Dulloh et al., 2024; Adhistry et al., 2023). The large volume and diversity of public opinions on social media generate massive textual data, requiring appropriate methods to understand the content and sentiment orientation of such data.

Topic modeling in this study was conducted using the Latent Dirichlet Allocation (LDA) method, with the optimal number of topics determined based on the coherence score. Meanwhile, sentiment analysis was performed using the Naïve Bayes algorithm with the application of the Synthetic Minority Oversampling Technique (SMOTE) to address data imbalance. Manual labeling was conducted, resulting in 2540 neutral tweets, 1218 negative tweets, and 321 positive tweets out of a total of 4079 tweets. The results show that the LDA model achieved a coherence score of 0.453 with 10 topics identified as the optimal number of topics. The Naïve Bayes sentiment analysis model with SMOTE achieved a performance of 72% accuracy, 62% precision, 68% recall, and 63% F1-score. The combination of topic modeling and sentiment analysis indicates that most topics are dominated by neutral sentiment, although several topics show a dominance of negative sentiment. Furthermore, the word cloud visualization shows that positive sentiment is related to support and hope regarding job opportunities, negative sentiment reflects mental pressure and dissatisfaction with the work environment, while neutral sentiment is dominated by informative discussions about recruitment processes and job-seeking experiences.

Keywords – Work Environment, Sentiment Analysis, Topic Modeling, Naïve Bayes, Latent Dirichlet Allocation (LDA)