Visualization of Sentiment Analysis Results of Public Opinion on Indonesian Public Figures in Electronic Media and Social Media

Mufadhil Hisyamul Ilmi[1] and Yan Puspitarani[2]

[1] Widyatama University, Bandung 40125, Indonesia
hisyamul.ilmi@widyatama.ac.id, yan.puspitarani@widyatama.ac.id

Abstract. This research aims to visualize public sentiment analysis towards Indonesian public figures on electronic media and social media platforms using Naive Bayes and Microsoft Power BI. The study involves several key steps, starting from data collection to visualization. Data from electronic media and social media are collected using web scraping techniques and converted into analyzable formats. Sentiment analysis is performed to categorize opinions into positive, neutral, and negative sentiments. The visualizations, created with Power BI, provide interactive dashboards that allow users to explore the sentiment distribution over time and across different media. This tool is designed to help understand the public perception of key Indonesian figures and their influence on media.

Keywords: Sentiment Analysis, Public Figures, Electronic Media, Social Media, Naive Bayes, Power BI, Data Visualization

1 Introduction

In the era of democracy, the press serves as a critical element in communication and public oversight of governmental systems, societal interactions, and national life, including the actions of prominent figures. It is undeniable that the advancement of human civilization in this modern age is inextricably linked to the role of media. Media, whether print, electronic, or internet, has become a crucial conduit for information flow in human life.

Prominent figures in Indonesian society, with their various policies, often receive a wide array of public opinions, both in favor and against, primarily due to perceived lack of transparency and decisions that may disadvantage the public. Public opinions on these figures are frequently expressed by Indonesians, particularly through online media and social media platforms. Therefore, analyzing opinions on online media is essential to determine whether these opinions are positive or negative. This information can provide insights into the sentiments and perspectives of the Indonesian public towards these figures. Consequently, there is a need for tools that can extract and process information from these opinion data to ultimately form patterns that generate knowledge and information, supporting business decision-making for public figures and the general populace

V. Mardiansyah and B. A. Prasetyo (eds.), *Proceedings of the Widyatama International Conference on Engineering 2024 (WICOENG 2024)*, Advances in Engineering Research 252,
https://doi.org/10.2991/978-94-6463-618-5_34

Numerous studies have previously explored sentiment analysis on social media, particularly Twitter. Tamora et al. conducted a sentiment analysis of public opinion on the Job Creation Law using the Naive Bayes Classifier and Term Frequency-Inverse Document Frequency (TF-IDF) methods, achieving an accuracy of 89.9% and visualizing commonly used words in tweets with a Word Cloud [1]. Similarly, I Komang Dharmendra et al. visualized public opinion data on Twitter regarding Nusantara as the capital city of Indonesia. The study involved creating visual representations such as graphs or other visual forms from the available data. The visualization highlighted significant aspects, patterns, and information from various dimensions, using elements like Word Cloud [2]. Building upon these previous studies, the current research emphasizes data visualization resulting from sentiment analysis. The research process includes data collection, preprocessing, data mining analysis, sentiment classification, and the development of interactive dashboards for dynamic and engaging data visualization.

In this research, the Naive Bayes method is employed to classify public sentiment from online media into positive, neutral, and negative categories, due to its reliability in handling textual data and efficiency in probability calculations. The classified sentiment is then visualized using Power BI, which offers interactive features like bar charts, line graphs, and word clouds to facilitate deeper analysis. Power BI's dashboard allows comprehensive exploration of sentiment patterns across time and platforms. This study addresses challenges such as the lack of transparency in public figures' policies, diverse public opinions that are difficult to categorize, and the need for effective tools to analyze and visualize these sentiments for strategic decision-making.

The goal of this research is to enable the community to build and make strategic decisions based on sentiment analysis outcomes. These outcomes aim to solve, reduce, and prevent negative issues while promoting innovation and positive developments through the directed actions of public figures.

2 Research Method

2.1 Research Target / Subject

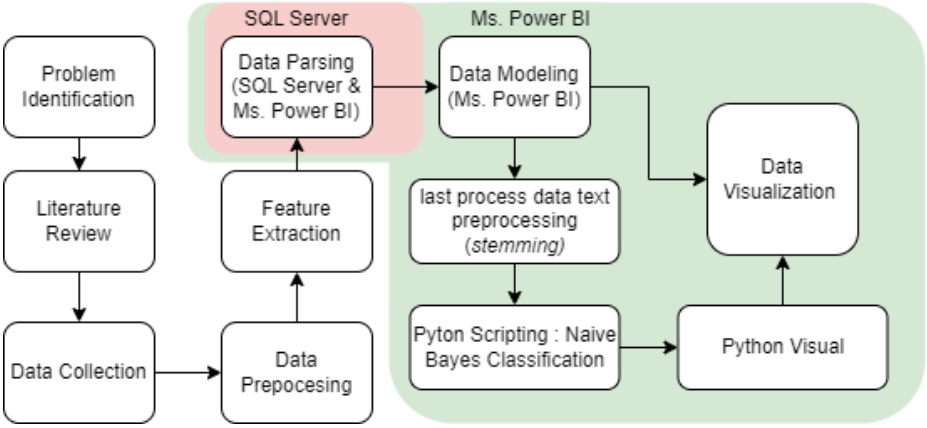

Fig. 1. Methodology for Creating Visualization of Sentiment Analysis Results of Public Opinion on Indonesian Public Figures in electronic media and social media.

The research methodology consists of eight main processes: problem identification, literature review, data collection, data preprocessing, feature extraction. data parsing, data modeling and data visualization.

Previous studies have explored sentiment analysis on social media, particularly Twitter. Tamora et al. conducted sentiment analysis on public opinion regarding the Job Creation Law using Naive Bayes Classifier and Term Frequency-Inverse Document Frequency (TF-IDF), achieving 89.9% accuracy and visualizing frequently used words in tweets through a word cloud [1]. I Komang Dharmendra et al. focused on visualizing public opinion on Twitter about Nusantara as Indonesia's new capital, using various visual elements, including a word cloud, to highlight key information and patterns [2]. These studies inform this research by emphasizing data visualization from sentiment analysis, covering data collection, preprocessing, sentiment classification, and building dynamic visualizations with an interactive dashboard.

The problem identification aims to support strategic business decisions through sentiment analysis of public opinion data, helping to reduce negative occurrences and promote positive actions via directed efforts from public figures. It seeks to effectively present social and electronic media data, analyze trends in public opinion and media coverage of prominent figures, and provide the public with insights for better judgment. Additionally, it aims to create engaging, interactive visualizations through an interactive dashboard. The literature review is a crucial phase in the research process that aids researchers in comprehending the conceptual framework, evaluating previous studies, and identifying areas of knowledge that still have gaps to be explored in the planned re-search. The sentiment analysis data collection process involves defining research objectives and scope, focusing on online articles from media sources such as Cnnindonesia.com, Kompas.com, and Detik.com. Web scraping and APIs are used to gather data, ensuring a comprehensive range of opinions and aspects relevant to public figures. Data preprocessing in sentiment analysis involves several key steps: cleaning text by removing irrelevant characters and HTML tags; tokenizing text into words or phrases; converting all text to lowercase; removing common stop words; performing stemming or lemmatization to reduce words to their base forms; eliminating irrelevant words; merging synonyms to reduce dimensionality; and conducting spell checking to correct typos [6]. Data parsing for sentiment analysis involves several stages: collecting text data for analysis, extracting relevant text using queries or extraction tools, and cleaning and preprocessing the data. The preprocessing includes removing punctuation, converting text to lowercase, eliminating stop words, and performing stemming or lemmatization. In this visualization context, preprocessing has already been done, so this stage is conducted in Power BI using Power Query to adjust data types. Data modeling for sentiment analysis includes feature extraction, where keywords and phrases are identified from the cleaned text, and data linking, where the data is associated with specific entities and connected to other relevant data for analysis.

These processes culminates in development of a Dashboard Management Reporting system published on Power BI Service include visual of sentiment analysis using the

Naive Bayes Classifier. The methodology is adapted from Ben Fry's "Visualizing Data" book, which consists of seven steps: acquire, parse, filter, mine, represent, refine, and interact [3]. These steps are tailored to fit the specific needs of this research. Output creation includes sentiment analysis using the Naive Bayes Classifier method, Dashboard Reporting, and the final project report.

The process aligns with the research objectives, starting with problem identification by analyzing the societal environment and public figures, and observing how electronic media convey public opinions about Indonesian public figures. A literature review is conducted to reference the Naive Bayes Classifier for sentiment analysis and Data-Driven Reporting using Microsoft Power BI, with data sourced from social media, specifically Twitter, via the Twitter API. Data parsing follows, mapping the information needs to the problem at hand, and designing visualizations such as graphs, bar charts, and word clouds. Visualization testing, implementation, and accuracy validation are performed. The final output includes sentiment analysis using Naive Bayes Classifier and Data-Driven Reporting. The scope of this research is limited to two data sources: online media articles from platforms such as CNNIndonesia.com, Kompas.com, and Detik.com, and social media data, specifically tweets from Twitter.

3 Result and Discussion

3.1 Data Collection

The data obtained from scraping Twitter and online media is detailed as follows.

Table 1. The details of data from Twitter

Number	Name of Figure	Total
1	Ahok	585
2	Anies	5000
3	Ganjar	5000
4	Prabowo	5000
5	Puan	1852
6	Ridwan	2380
	Total	19817

Table 2. The details of data from Online Media

Number	Name of Figure	Total
1	CNN Indonesia	1627
2	Kompas	1794
3	Liputan6	1794
	Total	5215

3.2 Text Preprocessing

In the preprocessing stage, data cleaning is performed to remove unnecessary data from the source [2]. This process aims to ensure that the data used is free from irrelevant elements or noise [2]. Python programming is employed for text preprocessing, utilizing the NLTK library for stopword removal and tokenization. For the stemming process, the StemmerFactory library from Sastrawi is used. For the collected tweet data, the preprocessing steps include tokenization, stemming, removal of slang words, and elimination of conjunctions using stopwords [2]. In the N-Gram modification section, the N-Gram code used is a trigram (3-gram) from the sentence "*saya makan nasi padang"* which results in ["*saya makan nasi*", "*makan nasi padang*"] [7]. Modifying N-Grams helps capture the context of words in text analysis.

3.3 Data Visualization in Microsoft Power BI

The previously preprocessed text data is then stored in a database using the data warehouse concept. A data warehouse is a centralized repository for storing large volumes of structured data from various sources [4]. The benefits of utilizing a data warehouse include providing a comprehensive view of data, enabling decision-makers to make

more informed choices. Additionally, data warehouses can store historical data, which helps in trend analysis and forecasting [5].

There are several types of data warehouse architectures, one of which is the two-layer architecture. The first layer is the source layer, where data exists as operational data. In this research, operational data refers to sentiment analysis results stored in databases. The second layer is the data staging layer, where operational data is extracted (ETL process) into the data warehouse. The third layer is the data warehouse layer, where information is stored in a centralized logical repository, which can be accessed directly. The final layer is the analysis layer, where special queries, such as Power Query in Power BI, are used for reporting and analysis.

The chosen topology is the Star Topology [5]. The star topology is a common schema used in data warehousing that organizes data into fact and dimension tables [5].

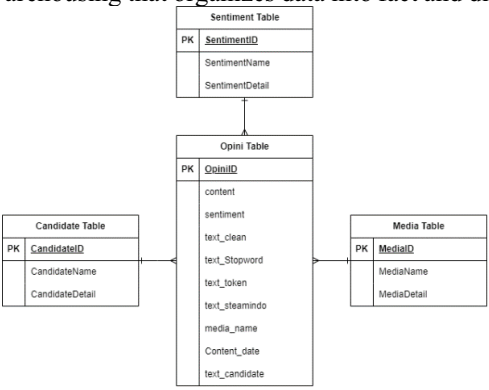

Fig. 2. Star Topology and relationship between data tables in each table retrieved in Microsoft Power BI for Creating Visualization of Sentiment Analysis Results of Public Opinion on Indonesian Public Figures in Electronic Media and Social Media.

In the Data Extraction phase, data from various sources is aligned with columns in fact and dimension tables, specifically by using an Excel file as the data source for the Opinion table (dimension table) processed with SSIS. Microsoft SQL Server is chosen as the DBMS, utilizing SSMS to create tables per the designed schema. After data cleaning, an effective data model is created, defining relationships and necessary measures for analysis. Visualization design follows, utilizing Power BI's bar charts, pie charts, geographic maps, and tables, each serving distinct purposes in the Sentiment Analysis Report of Key Figures.

Using the processed text data, Python scripting in Power BI Desktop is utilized by enabling the Python scripting feature in the Options menu. The final text-processing output column is used for visualization. For Naive Bayes classification, the confusion matrix and model evaluation bar chart are visualized using the sklearn library for classification and matplotlib for displaying the charts.

Fig. 3. Result of Report Sentiment Analysis Dashboard in Power BI.

4 Conclusions and Suggestions

4.1 Conclusion

Perceptions of figures through sentiment analysis, which categorizes sentiments as positive, neutral, or negative. This aids in identifying areas for improvement in image and communication strategies, aligning policies with public concerns, and predicting reactions to actions or statements. Audience segmentation through opinion data facilitates the development of targeted strategies. Effective presentation of this data is achieved using interactive dashboards, diverse visualizations, appropriate color palettes, and clean design, ensuring readability, simplicity, and user engagement.

4.2 Suggestion

Users are encouraged to utilize interactive dashboard features for deeper data exploration, specific insights, and regular sentiment monitoring for timely strategy adjustments. Policy makers should base policy formulation on sentiment analysis to align with public needs, enhance transparency, and consult the public for legitimacy. Future researchers should explore advanced sentiment analysis methods, study varied cases, and develop innovative visualization techniques to improve data presentation and user engagement.

References

1. T. N. Wijaya, R. Indriati, and M. N. Muzaki, “ANALISIS SENTIMEN OPINI PUBLIK TENTANG UNDANG-UNDANG CIPTA KERJA PADA TWITTER,” Jambura Journal of Electrical and Electronics Engineering, (2021)
2. K. Dharmendra, N. N. U. Januhari, R. A. N. Diaz, I. P. Ramayasa, and I. M. A. W. Putra, “Visualisasi Data Opini Publik pada Media Sosial Twitter (Studi Kasus : Nusantara Sebagai IKN Indonesia),” Jurnal Teknik Informatika UNIKA Santo Thomas (JTIUST), (2022)
3. B. Fry, “The Seven Stages of Visualizing Data,” in Visualizing Data, California, y O’Reilly Media, Inc.,pp. 5-6 (2008)
4. W. H. Inmon, Building the Data Warehouse, Indianapolis: Wiley, (2002)
5. M. Golfarelli, Data Warehouse Design: Modern Principles and Methodologies, Bologna: Tata McGraw Hill Education Private Limited, (2009)
6. S. Kannan and V. Gurusamy, “Preprocessing techniques for text mining,” International Journal of Computer Science & Communication Networks, vol. 5(1), pp. 7–16, (2014)
7. M. Irsa Nurodin, “Phrase Detection's Impact on Sentiment Analysis of Public Opinion and online Media Toward Political Figures,” ResearchGate, p. 75, (2024)