WEB-BASED ANALYSIS OF PUBLIC COMPLAINTS IN PERTUMBUKAN VILLAGE USING COSINE SIMILARITY AND TF-IDF TECHNIQUES

Randi Rian Putra[1]*, Nadya Andhika Putri[2], Agil Dwi Putra [3]
[1]Fakultas Sains Dan Teknologi, Teknologi Informasi, Universitas Pembangunan Panca Budi
[2]Fakultas Sains Dan Teknologi, Sistem Komputer, , Universitas Pembangunan Panca Budi
[3]Mahasiswa Teknologi Informasi, Universitas Pembangunan Panca Budi

Abstract: This research focuses on the development of a web-based system for analyzing public complaints in Pertumbuhan Village using Cosine Similarity and Term Frequency-Inverse Document Frequency (TF-IDF) techniques. The system is intended to assist local authorities in efficiently managing and categorizing public complaints by identifying patterns and trends in the data. TF-IDF is applied to assign weights to words based on their frequency within a complaint, while Cosine Similarity is used to determine the similarity between different complaint texts. By integrating these methods, the system groups complaints with similar themes, enabling the local government to prioritize and address the most pressing issues quickly. The web-based platform also offers real-time data analysis, enabling more informed decision-making processes for resolving public grievances. The evaluation results show that this system can significantly reduce manual sorting tasks, improve complaint classification accuracy, and provide scalable solutions for handling large datasets. This approach facilitates a more transparent, efficient, and responsive public service delivery.

INTRODUCTION

The management of public complaints is a critical aspect of local governance, as it directly impacts public satisfaction and the quality of services provided by government bodies. In local communities such as Pertumbuhan Village, the volume and diversity of complaints have increased significantly due to population growth and rising public awareness of service standards. These complaints can vary widely, from issues related to infrastructure, waste management, and public health to social services and safety concerns (Chen et al., 2016; Xiao, 2023). Traditionally, local authorities rely on manual methods to process and address these complaints, often leading to inefficiencies, delays in resolution, and even mismanagement due to the large volumes of data that need to be

377

handled (Kirana & Majid, 2022). As a result, there is a growing need for more efficient, accurate, and scalable solutions that can assist local governments in handling public complaints effectively (Maulana, 2023; Maulana & Dečman, 2023).

In the era of digital transformation, leveraging technology for better public service management has become increasingly feasible. One promising approach is to use text mining techniques to automate the analysis of complaint data. Text mining enables the extraction of valuable information from unstructured text data, allowing for more systematic analysis and categorization of public complaints. This study proposes the development of a web-based system that applies text mining techniques, specifically Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity, to facilitate the analysis and management of public complaints in Pertumbuhan Village (Maulana, 2023; Xiao, 2023).

TF-IDF is a statistical measure used to evaluate the importance of words within a document relative to a collection of documents, making it an effective tool for identifying key terms and themes in complaint data. By weighting words according to their frequency and relevance, TF-IDF helps highlight the most significant issues within each complaint(Putra et al., 2022). Meanwhile, Cosine Similarity is employed to compare the content of different complaints by measuring the similarity between the text vectors of the documents. This enables the system to group complaints with similar topics, helping local authorities to detect patterns and trends in the data, and to address recurring or high-priority issues more effectively(Supiyandi et al., 2022).

The integration of TF-IDF and Cosine Similarity within a web-based platform offers several advantages for local governments. First, it automates the complaint classification process, reducing the reliance on manual sorting and categorization. This leads to significant time savings and lowers the risk of human error (Perwitasari & Hendrawan, 2020; Ranti Eka Putri et al., 2023). Second, the system provides real-time analysis of complaints, enabling government officials to make faster and more informed decisions regarding public service improvements. Third, it supports scalability, allowing the system to handle increasing volumes of complaints without compromising performance or accuracy(Putri & Hartanto, 2022; Tasril et al., 2023).

This research focuses on the design, development, and evaluation of the proposed

system, emphasizing its role in improving the efficiency of public complaint management in Pertumbuhan Village. By implementing this system, local governments can not only streamline the processing of complaints but also improve the overall transparency and responsiveness of public service delivery. The results of this study will demonstrate how technology-driven solutions can be used to enhance governance and build stronger, more trustful relationships between the public and government institutions(Hafni Hafni, 2023).

RESEARCH METHODS

The development of the web-based public complaint analysis system in this study follows a structured methodology that integrates both technical design and implementation phases with data collection and evaluation processes. The primary goal of the system is to classify and analyze public complaints using text mining techniques, specifically Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity, to assist local authorities in Pertumbuhan Village in managing complaints more efficiently. The research methodology consists of the following steps:

1. Data Collection:

   The first step involves collecting public complaints from the local government's existing complaint system or manual logs. These complaints, written in natural language, form the dataset for analysis. The dataset is pre-processed to remove duplicates, correct spelling errors, and eliminate irrelevant data. The final dataset used for the system contains a representative sample of complaints covering various categories, including infrastructure, public services, and environmental issues.

2. Text Preprocessing:

   Text preprocessing is a crucial step in preparing the complaint data for analysis. This includes tokenization, which breaks down the text into individual words, and stemming, which reduces words to their base or root forms. Additionally, stop words (common words such as "the," "and," or "in") are removed to reduce noise in the data. This step ensures that the most relevant words are used in the subsequent analysis.

3. TF-IDF Application:

   Once the text is preprocessed, the Term Frequency-Inverse Document Frequency

(TF-IDF) technique is applied to weight the words in each complaint. TF-IDF quantifies the importance of each word by considering both its frequency within a single document and its rarity across the entire dataset. Words that appear frequently in one complaint but are rare across others are given higher importance. This weighting process allows the system to identify key issues within each complaint.

4. Cosine Similarity Calculation:

    After generating the TF-IDF scores, Cosine Similarity is used to measure the similarity between complaint texts. The complaints are represented as vectors based on their TF-IDF scores, and the cosine of the angle between these vectors is calculated to determine how similar the complaints are to each other. A value of 1 indicates identical complaints, while a value of 0 indicates no similarity. This method enables the grouping of similar complaints, which helps the system identify recurring issues and prioritize them for faster resolution.

5. System Design and Development:

    The web-based platform is designed to provide an intuitive user interface for local government officials. The system is built using modern web technologies, with a backend that processes and analyzes complaint data using the TF-IDF and Cosine Similarity algorithms. The platform also features data visualization tools, enabling users to view complaint categories, trends, and statistics in real-time. The system is hosted on a web server to ensure accessibility and scalability.

6. System Testing and Evaluation:

    After development, the system undergoes rigorous testing to evaluate its performance in terms of accuracy, efficiency, and scalability. The accuracy of complaint classification is measured by comparing the system's output with manually classified complaints. Efficiency is assessed by tracking the time taken to process and categorize complaints. Scalability is tested by gradually increasing the dataset size to ensure the system can handle large volumes of data without performance degradation. Feedback from local government officials is also gathered to assess the system's usability and effectiveness in improving complaint management.

RESULTS AND DISCUSSION

The web-based system for analyzing public complaints in Pertumbuhan Village was evaluated based on several key performance metrics: classification accuracy, processing speed, scalability, and user satisfaction. In addition, feedback from local government officials provided insights into the system's usability and its practical impact on complaint management. This section presents detailed results and discusses the significance of these findings, supported by quantitative data and tables to illustrate the system's performance.

1. **Classification Accuracy**

   The system's ability to categorize complaints accurately was measured by comparing its output to manually classified complaints. A total of 500 complaints were used for this evaluation, covering various categories such as infrastructure, waste management, water supply, and public health. The system achieved an overall classification accuracy of 87%. The results, as shown in Table 1, indicate that the accuracy varied depending on the complexity of the complaint. Complaints with more straightforward and technical descriptions (e.g., infrastructure or waste management) had higher accuracy, whereas more nuanced or subjective complaints (e.g., public health or social issues) were slightly less accurate due to the complexity of the language.

   Table 1. Classification Accuracy

Category	Number of Complaints	System Accuracy (%)
Infrastructure	150	91
Waste Management	100	89
Water Supply	80	85
Public Health	70	82
Social Services	100	86
Overall Average	500	**87**

   The table above shows that the system performed exceptionally well in handling complaints related to infrastructure and waste management, with accuracies of 91% and 89%, respectively. However, categories involving more subjective issues, such as public health (82%) and social services (86%), had slightly lower accuracy due to variations in complaint wording and tone. These

results highlight the effectiveness of the system but also point to areas where improvements could be made, such as enhancing the system's ability to understand more nuanced complaints using advanced natural language processing (NLP) techniques.

2. **Processing Speed and Scalability**

The efficiency of the system in processing large datasets was a critical factor in its evaluation. Table 2 presents the average time taken to process and classify complaints as the dataset size increased. The system maintained a high processing speed, with each complaint being processed in an average of 0.5 seconds. Even with a dataset size of 10,000 complaints, the system showed no significant drop in performance, demonstrating its scalability.

Table 2. Processing Speed and Scalability

Dataset Size	Average Processing Time per Complaint (seconds)
100 complaints	0.3
500 complaints	0.4
1,000 complaints	0.5
5,000 complaints	0.5
10,000 complaints	0.5

As shown in Table 2, the system's processing time remained consistent, even as the dataset size increased, indicating that it is capable of handling large volumes of complaints without significant delays. This scalability is crucial for local governments facing growing populations and rising volumes of public feedback.

3. **Identifying Recurring Issues**

One of the key benefits of using Cosine Similarity in the system was its ability to detect and cluster similar complaints. This feature allowed the system to identify recurring issues within the village. Table 3 illustrates the most frequent complaint themes identified by the system and the number of complaints in each cluster.

Table 3. Identifying Recurring Issues

Complaint Theme	Number of Complaints
Road Maintenance	120
Waste Collection Delays	90
Water Supply Disruptions	80

Complaint Theme	Number of Complaints
Public Health Concerns	70
Street Lighting Issues	60

As shown in Table 3, road maintenance and waste collection delays were the most commonly recurring issues in Pertumbuhan Village, with 120 and 90 complaints, respectively. By grouping these complaints together, the system enabled local authorities to prioritize addressing the most frequent and pressing problems. Additionally, the system provided real-time updates on emerging complaint trends, allowing for proactive management of community concerns.

4. **User Satisfaction and Usability**

Feedback from local government officials was gathered through surveys to assess user satisfaction with the system. The feedback focused on ease of use, system responsiveness, and the perceived impact on complaint management efficiency. The survey results, shown in Table 4, reflect high levels of satisfaction across all categories, particularly in the areas of ease of use and overall performance.

Table 4. User Satisfaction and Usability

Evaluation Criteria	Average Rating (out of 5)
Ease of Use	4.7
System Responsiveness	4.5
Complaint Classification	4.3
Data Visualization	4.6
Impact on Efficiency	4.8
Overall Satisfaction	**4.6**

As indicated in Table 4, the system scored particularly high in terms of ease of use (4.7) and its impact on operational efficiency (4.8). Local officials appreciated the system's ability to automate the complaint handling process, allowing them to allocate more time and resources to resolving issues rather than managing data. Additionally, the real-time data visualization tools were found to be highly useful in providing a clear overview of complaint trends.

5. **Limitations and Future Improvements**

Although the system performed well overall, certain limitations were identified. Complaints written in informal language or containing ambiguous or colloquial

terms were occasionally misclassified, particularly in categories such as public health and social services. This highlights the need for further refinement in the natural language processing capabilities of the system. In future developments, incorporating more advanced NLP techniques, such as synonym detection, context-aware models, and sentiment analysis, could help the system better handle complex or ambiguous complaints. Moreover, integrating sentiment analysis could provide additional insights into the urgency and emotional tone of the complaints, allowing local authorities to prioritize issues more effectively.

In conclusion, the web-based complaint analysis system using TF-IDF and Cosine Similarity demonstrated high accuracy, scalability, and usability. It effectively streamlined the complaint management process in Pertumbuhan Village, reducing manual effort and improving the speed of issue resolution. Future iterations of the system can enhance performance by addressing the identified limitations and incorporating additional text analysis features. This approach not only improves operational efficiency but also enhances the transparency and responsiveness of local governance.

CONCLUSION

This study successfully developed and implemented a web-based public complaint analysis system for Pertumbuhan Village, utilizing Term Frequency-Inverse Document Frequency (TF-IDF) and Cosine Similarity techniques. The system significantly improved the management of public complaints by automating the categorization process, identifying recurring issues, and providing real-time insights into community concerns. With an overall classification accuracy of 87%, the system demonstrated its capability to handle a wide range of complaint categories, such as infrastructure, waste management, and public health. The system's scalability and efficiency were also proven, maintaining fast processing times even with large datasets. Feedback from local government officials indicated high satisfaction with the system's ease of use, data visualization tools, and its positive impact on complaint handling efficiency.

Despite its success, the system faced challenges in accurately classifying complaints written in informal or ambiguous language. Future improvements could

focus on enhancing the system's natural language processing capabilities, incorporating sentiment analysis, and refining the classification algorithms to handle more complex or nuanced complaints.

In conclusion, this web-based system provides a scalable, efficient, and transparent solution for public complaint management, demonstrating the potential of text mining techniques to enhance local governance and public service delivery.

REFERENCE

Chen, K., Zhang, Z., Long, J., & Zhang, H. (2016). Turning from tf-idf to tf-igm for term weighting in text classification. Expert Systems With Applications, 66, 245-260. https://doi.org/10.1016/j.eswa.2016.09.009

Hafni Hafni, I. I. (2023). Perancangan sistem Pencatatan Kreatif Siswa Berbasis Desktop Pada SMK N 9 Medan. *OMIK (Konferensi Nasional Teknologi Informasi Dan Komputer)*, *6*(1), 843–846. https://doi.org/10.30865/komik.v6i1.5875

Kirana, N. and Majid, N. (2022). Challenges of digital transformation on good governance for improving public services quality. Nusantara Science and Technology Proceedings, 43-47. https://doi.org/10.11594/nstp.2022.2307

Maulana, R. (2023). Collaborative approach on digital government transformation in west java.. Jurnal Ilmu Administrasi Media Pengembangan Ilmu Dan Praktek Administrasi, 20(2), 141-155. https://doi.org/10.31113/jia.v20i2.962

Maulana, R. and Dečman, M. (2023). Collaborative governance in the digital transformation age: a systematic literature review with bibliometric mapping. Central European Public Administration Review, 21(1), 31-60. https://doi.org/10.17573/cepar.2023.1.02

Perwitasari, I. D., & Hendrawan, J. (2020). Rancang Bangun Sistem E-Posyandu Penjadwalan Dan Monitoring Perkembangan Bayi Berbasis Android. *INTECOMS: Journal of Information Technology and Computer Science*, *3*(1), 1–8. https://doi.org/10.31539/intecoms.v3i1.1331

Putra, R. R., Putri, N. A., & Wadisman, C. (2022). Village Fund Allocation Information System for Community Empowerment in Klambir Lima Kebun Village. *Journal of Applied ...*, *3*(2), 98–104. https://journal.yrpipku.com/index.php/jaets/article/view/681%0Ahttps://journal.yrpipku.com/index.php/jaets/article/download/681/467

Putri, N. A., & Hartanto, S. (2022). Sistem Informasi Manajemen Aset Online Dengan Penelusuran Data Menggunakan Konsep String Matching. *Device : Journal of Information System, Computer Science and Information Technology*, *3*(1), 17–24. https://doi.org/10.46576/device.v3i1.2183

Ranti Eka Putri, M. Wasito, & Ayu Nadia Lestari. (2023). Sosialisasi Pemanfaatan Media Sosial Sebagai Media Promosi Produk UMKM Desa Suka Damai. *JURPIKAT (Jurnal Pengabdian Kepada Masyarakat)*, *4*(3), 667–675. https://doi.org/10.37339/jurpikat.v4i3.1550

Supiyandi, S., Zen, M., Rizal, C., & Eka, M. (2022). Perancangan Sistem Informasi Desa

Tomuan Holbung Menggunakan Metode Waterfall. *JURIKOM (Jurnal Riset Komputer)*, *9*(2), 274. https://doi.org/10.30865/jurikom.v9i2.3986

Tasril, V., Zen, M., Fitriani, E. S., & Putra, A. D. (2023). Desain Ui/Ux Prototipe Pembelajaran Berbasis Game Kosakata Bahasa Inggris Dengan Metode Hcd Ui/Ux Design of English Vocabulary Game-Based Learning Prototype Using the Hcd Method. *Journal of Information Technology and Computer Science (INTECOMS)*, *6*(1), 1–8.

Xiao, J. (2023). How digital transformation improve government performance: the mediating role of partnering agility. Ieee Access, 11, 59274-59285. https://doi.org/10.1109/access.2023.3284793