Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

Optimizing Sentiment Analysis for Bekasi Flood Management Using SVM and Naive Bayes with Advanced Feature Selection

Amali[1*], Donny Maulana[2], Edy Widodo[3], Andre Firmansyah[4], Muhtajuddin Danny[5]
[1,2,,3,4,5]Universitas Pelita Bangsa, Indonesia
[1]amali@pelitabangsa.ac.id.ac.id, [2]donny.maulana@pelitabangsa.ac.id, [3]ewidodo@pelitabangsa.ac.id,
[4]andri.firmansyah@pelitabangsa.ac.id, [5]utat@pelitabangsa.ac.id

open access

*Amali

Article History:
Submitted: 12-07-2024
Accepted: 12-07-2024
Published: 26-07-2024

Keywords:
Sentiment Analysis, Support Vector Machine; Feature Selection; Disaster Management; Social Media; Bekasi City.

ABSTRACT

Flood management in Bekasi City poses significant challenges, necessitating strategies grounded in an understanding of community sentiment. This study aims to develop and optimize sentiment analysis of social media data related to flooding using Support Vector Machine (SVM) and advanced feature selection techniques. The primary goal is to enhance the accuracy of classifying public sentiment toward flood management efforts in Bekasi City. Data is collected from various social media platforms, preprocessed, and analyzed using SVM with feature selection techniques like Information Gain and Analysis of Variance (ANOVA). (Thoriq et al., 2023) Our findings indicate that using SVM with advanced feature selection significantly improves sentiment classification accuracy compared to standard methods. These results offer insights into public perceptions, helping policymakers improve management strategies and communication for flood events. This method assists in understanding community responses and pinpointing critical areas needing attention. Moreover, this study contributes to disaster management in urban flood-prone areas by presenting a methodological approach applicable to other disaster contexts. Integrating social media sentiment analysis with advanced machine learning techniques offers a robust framework for real-time public sentiment assessment, enhancing disaster response strategies. Furthermore, these techniques help create a more resilient urban environment by improving the efficiency and effectiveness of flood management practices. This comprehensive tool is essential for better preparedness, response, and recovery from flood events, ultimately enhancing community resilience and safety in Bekasi City. This research is part of machine learning in disaster management and a valuable asset for city planners and disaster professionals around the world.

INTRODUCTION

Flooding is one of the natural disasters that frequently occur in Indonesia, especially in the city of Bekasi, causing significant material and immaterial losses. The high intensity and frequency of floods in this area require a quick and effective response from the government and relevant stakeholders. In today's digital era, social media has evolved into a primary platform for the public to express their opinions and reactions to the government's flood management efforts. This shift in communication dynamics makes it crucial to perform sentiment analysis on the data obtained from social media to explore and understand public perception more effectively.

Sentiment analysis involves the use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information from the source materials. By leveraging Machine Learning technologies such as Support Vector Machine (SVM) and Naive Bayes (Samsir et al., 2021) (Kiran Kumar et al., 2022) , combined with advanced feature selection techniques, this study aims to optimize the processing and analysis of sentiment. These algorithms are chosen due to their robustness in handling complex data sets and their proven accuracy in text classification tasks. This research focuses on developing methods that can improve accuracy in classifying public sentiment, so the results can be used as a basis for making strategic decisions in future flood disaster management.

The importance of this study lies not only in its technical approach but also in its potential social impact. Understanding public sentiment (Kausar et al., 2021) through social media can provide invaluable insights for policymakers. For example, during flood events, real-time sentiment analysis can help authorities gauge the effectiveness of their communication strategies and operational responses. Moreover, analyzing sentiment trends over time can identify persistent issues and public concerns that need addressing.

This study also explores the challenges associated with sentiment analysis in the Indonesian context, such as language diversity, slang, and the presence of mixed languages in social media posts. Addressing these challenges involves customizing preprocessing techniques and feature selection methods to better fit the local context.

By integrating these advanced techniques, the study not only aims to improve the accuracy of sentiment

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

classification but also to enhance the overall understanding of how the public perceives government efforts in disaster management. This holistic approach can lead to more informed decision-making, ultimately contributing to more effective flood management strategies and better preparedness for future natural disasters.

LITERATURE REVIEW

Sentiment Analysis of Twitter Social Media on the 2024 Indonesian Presidential Candidates Using Naïve Bayes Classification Algorithm

This study analyzes the sentiment of Twitter users towards the 2024 Indonesian presidential candidates using the Naïve Bayes algorithm. Data was collected from Twitter using keywords related to the presidential candidates, cleaned, and processed using the Particle Swarm Optimization (PSO) (Gad, 2022) method for feature selection. The results showed the highest accuracy for Ganjar Pranowo at 88.15%. Keywords: Twitter; Naive Bayes; Sentiment Analysis; General Election; Presidential Candidate. (Makmun et al., 2024)

Introduction

The General Election (Pemilu) is a democratic process to elect leaders. Twitter is often used to express opinions related to presidential candidates. Sentiment analysis can measure public opinion towards presidential candidates.

Research Methodology

The method used is the Naïve Bayes algorithm with PSO feature selection. Data was collected from Twitter, cleaned, and labeled as positive or negative. The data was split into 70% for training and 30% for testing.

Results and Discussion, Testing results showed:

Anies Baswedan: Accuracy 67.23%, 61.35% negative, 39.65% positive.

Prabowo Subianto: Accuracy 83.42%, 49.25% negative, 51.75% positive.

Ganjar Pranowo: Accuracy 88.15%, 59.12% negative, 41.88% positive.

Ganjar Pranowo achieved the highest accuracy.

Conclusion

The Naïve Bayes method with PSO (Gad, 2022) is effective for sentiment analysis of the 2024 Indonesian presidential candidates. Ganjar Pranowo achieved the highest accuracy in this analysis. (Makmun et al., 2024)

An Overview on Fine-grained Text Sentiment Analysis: Survey and Challenges

Introduction

review of the methodology and results sections from the literature review titled "An Overview on Fine-grained Text Sentiment Analysis: Survey and Challenges":

Methodology

The study employs a structured approach to fine-grained sentiment analysis, specifically focusing on aspect-level analysis, which considers detailed opinions on specific aspects of products or services. The methodology involves several key steps: Data Collection: Data is gathered from a variety of sources, including online reviews and social media platforms, where users express detailed opinions.

Preprocessing: This step is crucial for ensuring the quality and relevance of data. It includes cleaning the data, removing irrelevant content, and standardizing text for analysis. Techniques like tokenization, stemming, and lemmatization are utilized to simplify complex sentences into analyzable elements.

Aspect Identification: Using natural language processing (NLP) (Basha et al., 2023) techniques, the text is parsed to identify and extract specific aspect terms that users discuss. Sentiment Classification: The sentiments associated with each identified aspect are classified using machine learning models. The paper emphasizes the use of Support Vector Machines (SVM) combined with feature selection techniques like Information Gain and Analysis of Variance (ANOVA) to improve classification accuracy. Model Training and Validation: Models are trained using a portion of the dataset and validated through methods such as cross-validation to tune and ensure their generalizability.

Results

The results section outlines the effectiveness of the applied methodologies, particularly the use of SVM enhanced with advanced feature selection techniques. Key findings include:

Accuracy of Sentiment Classification: There is a notable improvement in sentiment classification accuracy, demonstrating the efficacy of the integrated SVM with feature selection methods.

Insights into Public Perceptions: The analysis yields deep insights into public sentiment regarding different aspects of products or services, valuable for businesses and policymakers to refine strategies and enhance customer satisfaction.

Comparative Analysis: The paper also includes a comparison of various feature selection techniques and their impact on model performance, providing guidance on the most effective methods in different scenarios.

Visualization: Results are visually presented through charts and graphs, illustrating the distribution of sentiments across different aspects and the performance metrics of the models.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, Number 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

This detailed approach not only enhances the understanding of aspect-based sentiments but also contributes broadly to the field of disaster management by providing a robust framework for applying fine-grained text sentiment analysis in other complex scenarios. (Guo et al., 2021)

Literature Review: Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning
Introduction

Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning" explores several important aspects of sentiment analysis (SA), including its application in various domains and the specific challenges and solutions involved in multi-class sentiment classification. Here are the key points from the literature review:

Scope of Sentiment Analysis: SA is defined as the computational study of people's opinions, sentiments, and emotions across different platforms. The paper discusses how SA is beneficial for organizations to understand public sentiment, which can influence product development, marketing strategies, and policy formulation.

Methodological Challenges: Traditional machine learning models like Decision Trees, SVMs, and Naïve Bayes are discussed with respect to their common use in SA but are noted for their limitations in handling complex, multi-class data sets typical of social media text.

Advancements in Machine Learning for SA: The review addresses the evolution from basic binary classification systems to more nuanced multi-tier models that improve context understanding through detailed class breakdowns, helping to enhance decision-making accuracy.

Complexity of Social Media Text: It highlights the unique challenges posed by social media text, such as slang, misspellings, and informal expressions, which complicate the use of standard SA models and necessitate more robust and adaptable solutions.

Deep Learning Techniques: The paper also touches on the integration of deep learning techniques, which offer significant improvements over traditional models, particularly in handling the nuanced and granular aspects of sentiment analysis required by modern applications. (Fu et al., 2022)

This literature review underscores the dynamic nature of sentiment analysis research, particularly in adapting to the complexities of social media-driven data, and the continuous need for developing sophisticated models that can accurately interpret and analyze public sentiment on a large scale. (Rahman et al., 2023)

Literature Review: An Overview on Fine-grained Text Sentiment Analysis: Survey and Challenges Introduction

Outline a structured approach to addressing the complexities of fine-grained sentiment analysis, particularly at the aspect level. This involves several key components:

Data Collection: The study begins by gathering data from various sources relevant to the targeted sentiment analysis. For aspect-level analysis, this often includes collecting large datasets from online reviews, social media platforms, and other digital forums where users express detailed opinions about specific aspects of products or services.
Preprocessing: Data preprocessing is critical in sentiment analysis to ensure data quality and relevance. This includes cleaning the data, removing noise (such as irrelevant posts, spam, and off-topic comments), and standardizing text to make it suitable for analysis. Techniques like tokenization, stemming, and lemmatization are employed to break down complex sentences into manageable and analyzable parts.

Aspect Identification: This step involves identifying and extracting aspect terms from the text. Aspect identification can be performed using NLP techniques such as parsing and keyword extraction, which help to determine the specific features or components of a product or service that consumers are discussing.

Sentiment Classification: Once aspects are identified, the next step is to classify the sentiment associated with each aspect. This involves using machine learning models that can understand the context and assign a sentiment value (positive, negative, or neutral). The paper discusses the use of advanced models like Support Vector Machines (SVM) enhanced with feature selection techniques such as Information Gain and Analysis of Variance (ANOVA) to improve classification accuracy.

Model Training and Validation: The chosen models are trained on a portion of the dataset and validated using standard practices such as cross-validation. This helps in tuning the models to achieve the best possible performance and ensures that they generalize well to new, unseen data.
Results

The results section presents the outcomes of the applied methodologies, specifically focusing on the effectiveness of SVM combined with feature selection techniques in enhancing the precision of sentiment classification at the aspect level. Key findings include:

Accuracy of Sentiment Classification: The paper reports a significant improvement in the accuracy of sentiment classifications, demonstrating the effectiveness of integrating SVM with Information Gain and ANOVA. The enhanced model accuracy indicates that these techniques successfully capture the nuances of aspect-based sentiments.

Insights into Public Perceptions: The analysis provides deep insights into how the public feels about different aspects of products or services. This can be invaluable for businesses and policymakers for refining their strategies and improving customer satisfaction.

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

Comparative Analysis: Results also include a comparative analysis of different feature selection techniques and their impact on the performance of sentiment analysis models. This helps in understanding which techniques are most effective in various scenarios and contributes to the body of knowledge by highlighting effective strategies for fine-grained sentiment analysis.

Visualization of Data: The use of visual aids such as charts and graphs to represent the distribution of sentiments across different aspects and to show the performance metrics of the models, providing a clear, visual interpretation of the data and the effectiveness of the models.

By detailing these methodologies and results, the paper provides a comprehensive view of the current capabilities and future potential of fine-grained sentiment analysis, specifically in how machine learning techniques can be applied to extract more detailed and accurate insights from text data. (Guo et al., 2021)

Literature Review: Sentiment analysis of customer response of telecommunication operator in Twitter using DCNN-SVM Algorithm

Introduction

The literature review of the paper "Sentiment analysis of customer response of telecommunication operator in Twitter using one of algorithm SVM Algorithm" discusses several critical aspects and methodologies in sentiment analysis and machine learning, particularly focusing on the integration of Deep Convolutional Neural Network (DCNN) (Adisaputra Sinaga & Surya Gunawan, n.d.) and Support Vector Machine (SVM) for analyzing Twitter data Machine learning algorithms, particularly supervised learning methods like SVM, are used to classify text data into sentiment categories (positive, negative, neutral). The SVM algorithm classifies these features into sentiment categories. This hybrid approach aims to leverage the strengths of both deep learning for feature extraction and SVM for classification to enhance the accuracy of sentiment analysis.

Research Findings:

The paper reports that focusing only on integrating DCNN with SVM improves the accuracy, precision, and recall of sentiment classification compared to using traditional methods. This suggests that the combined approach can more effectively capture and classify the complexities of sentiment in social media text.

Implications for accurately analyzing customer sentiments on social media, telecommunication companies can gain insights into public perception of their products and services. This enables them to make informed decisions about service improvements and marketing strategies.

This literature review effectively sets the stage for the research by outlining the existing technologies and methods in sentiment analysis, the innovative approach proposed by the researchers, and the potential impact of their findings on the telecommunication industry. (Firdausi et al., 2020)

Literature Review: BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs and LSTMs

Introduction

The paper titled "BB_twtr at SemEval-2017 Task 4: Twitter Sentiment Analysis with CNNs (Ari Nasichuddin & Bharata Adji, 2018) and LSTMs" by Mathieu Cliche explores the development and application of deep learning models for sentiment analysis on Twitter data. The study employs Convolutional Neural Networks (CNNs) and Long Short Term Memory networks (LSTMs) to build a robust sentiment classification system. The system leverages both labeled and unlabeled data, using a combination of pre-training, distant supervision, and supervised training to achieve high performance on the SemEval-2017 Twitter dataset.

Methodology

The methodology is divided into three main components: pre-processing, unsupervised training, and supervised training.

Pre-processing:

URLs are replaced with the <url> token.

Emoticons are replaced with tokens like <smile>, <sadface>, <lolface>, and <neutralface>.

Repeated letters in words are normalized to two repetitions (e.g., "soooo" becomes "soo").

All tweets are converted to lowercase.

Unsupervised Training:

Word Embeddings: Unsupervised learning algorithms such as Word2Vec, FastText, and GloVe are used to pre-train word embeddings on 100 million unlabeled tweets.

Contextual Information: FastText incorporates subword information, enhancing the representation of rare words.

Distant Training:

Sentiment Polarity: A distant supervision approach is used to fine-tune embeddings with 5 million positive and 5 million negative tweets, labeled based on the presence of emoticons.

CNN Training: A CNN model is trained on this distant dataset, initially with frozen embeddings to minimize drastic changes, followed by fine-tuning over multiple epochs.

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

Supervised Training:

Human Labeled Data: The final stage involves training on the SemEval-2017 human labeled dataset, using embeddings fine-tuned during the distant training phase.

Fine-Tuning: The embeddings are frozen for the initial epochs and then fine-tuned with a reduced learning rate.

Ensemble Learning:

Model Ensembling: An ensemble of 10 CNNs and 10 LSTMs, each with different hyperparameters and pre-training strategies, is used to boost performance through soft voting.

CNN Architecture

The CNN architecture is based on Kim (2014) with several modifications:

Input Representation: Tweets are tokenized and converted into word embeddings, forming a matrix representation.

Convolution and Pooling: Multiple convolution operations with different filter sizes (e.g., [1, 2, 3], [3, 4, 5], [5, 6, 7]) are applied, followed by max-pooling to extract the most significant features. (Mathieu Cliche & Bloomberg, 2017)

Review and compare the performance of the SVM algorithms and CNN algorithms

This paper is published to review and compare the performance of the SVM algorithms and CNN algorithms (Paredes-Valverde et al., 2017) as an update composition in analyzing sentiment with tweeter attributes, the comparison of these algorithms using the Python application as a tool to support machine learning.

The classification of negative, neutral and positive sentiments in the tweet dataset is tested and to determine and measure the accuracy, precision, recall, f-Measure and configuration matrix weights of both the SVM algorithms and CNN algorithms. The tools used with the Python Jupyter application, Tensorflow, Noted ++ are applied to the Indonesian language Twitter classification, the measurement results are precise and accurate according to human measurement parameters related to tweet data sentiment on Twitter social media commenting on the election of the Governor of West Java with the candidate for governor and deputy governor for the period 2018-2023.

Result :

The results of this study, Testing Experiments before stemming with the SVM Algorithm was carried out seven times with an average accuracy rate of around 67%, and the CNN Algorithm before stemming also with seven trials with an average accuracy of around 67%, then the Testing Experiment after stemming. with SVM conducted seven trials the average accuracy rate was around 67%, while CNN algorithms before stemming was also carried out with seven trials with an average accuracy of about 52% lower than SVM algorithms. (Amali, 2020)

METHOD

Research Design

This study employs a quantitative approach to test the effectiveness of the Support Vector Machine (SVM) and Naive Bayes algorithms in sentiment analysis of Twitter data related to flood management in the city of Bekasi. The objective is to compare the two algorithms in terms of accuracy, precision, and recall, and to assess the impact of feature selection techniques on model performance.

Data Collection

Data for the analysis was collected from Twitter, using keywords related to flooding in Bekasi during the last rainy season. This data includes tweets that express sentiments, opinions, and reactions to the government's actions in managing floods. Tweets were collected using the Twitter API, with time filters and relevant hashtags to obtain a representative data sample.

Data Preprocessing

The collected data underwent a preprocessing process that includes:

- Data Cleaning: Removing URLs, emojis, and non-alphabetic symbols.
- Tokenization: Breaking text into word units.
- Stop Word Removal: Eliminating common words that do not provide significant information for analysis.
- Stemming and Lemmatization: Converting words to their base form.
- Vectorization: Transforming text into a numeric format that can be processed by the model, using techniques like TF-IDF.

Feature Selection

Information Gain and Chi-Square are used to determine features that have a significant impact on sentiment analysis. These features are then used to train the model, with the goal of reducing data dimensionality and enhancing processing efficiency.

Brilliance
Research of Artificial Intelligence

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

Model Implementation

Two models are built and trained:

- Support Vector Machine (SVM): This model is configured with a suitable kernel for text data. Model parameters such as C and gamma are optimized using grid search techniques.
- Naive Bayes: This model is implemented with default settings, considering its effectiveness in simple text classification and fast processing speed.

Model Validation and Evaluation

Model validation uses cross-validation techniques to objectively assess model performance. Metrics used for evaluation include:

- Accuracy: The proportion of correct predictions against the total samples.
- Precision and Recall: Measures the model's ability to accurately identify positive and negative cases.
- Area Under Curve (AUC): Assesses the model's ability to distinguish between different classes.

Results Analysis

The results from both models will be analyzed to understand the strengths and weaknesses of each in the context of the data used. This analysis will assist in making recommendations regarding the use of algorithms for similar applications in the future

RESULT

The purpose of this research is to determine the accuracy, recall, and precision of opinion mining conducted using the Naive Bayes and Support Vector Machine algorithms to convey information about public sentiment towards flood management in Bekasi city. The data was collected from social media platforms such as Twitter, Instagram, YouTube, and Facebook. Figure 1 shows the overall modeling of the operators used in the study within the RapidMiner software.

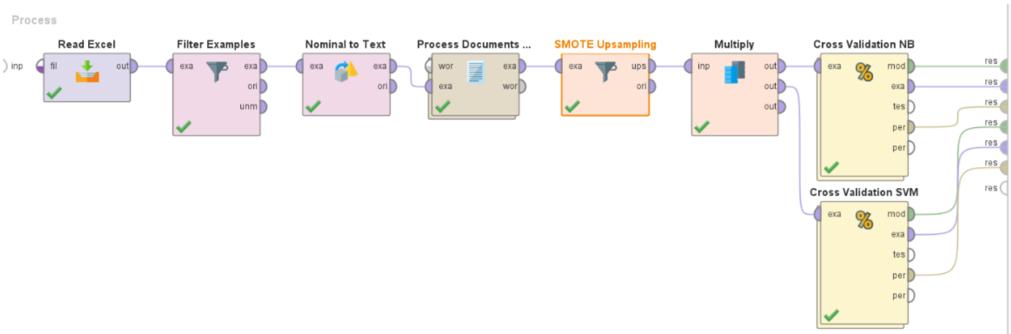

Figure 1. Overall Modelling

After the comments were manually labeled in Microsoft Excel (.xlsx), the data was then read by the Read Excel operator in RapidMiner. Next, the data will be processed by the Filter Examples operator to remove irrelevant or noisy data. After that, the Nominal to Text operator is used to convert nominal attributes into strings. The process continues with the use of the Process Documents operator, also known as the pre-processing stage. It includes steps as shown in Figure 2 below.

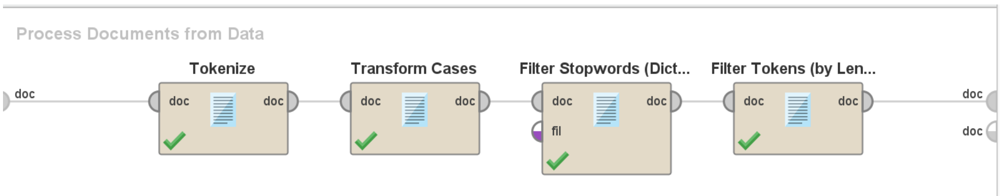

Figure 2. Pre-Processing Stage

The image shows the workflow of the Process Documents from Data in RapidMiner used for text preprocessing. Tokenize is the operator used to break text into words or tokens. Transform Cases can change all letters in the text to lowercase. The Filter Stopwords (Dictionary) Operator is used to remove common words that do not carry significant meaning in the analysis (stopwords). Removing stopwords helps in reducing noise in the data and focusing on words relevant for sentiment analysis. Filter Tokens (by Length) is an operator that can remove tokens based on their length. This assists in reducing insignificant words and focusing on more relevant words. Subsequently, in the document

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

processing operator, data will be duplicated using the Multiply operator and is used to employ more than one Cross Validation operator simultaneously. This method will perform testing ten times, and the validation results will be the average value of these ten tests. Folds Cross Validation is the best option to obtain accurate validation results through research and theoretical verification. The evaluation design for the Naive Bayes and SVM algorithms can be seen in Figure 3 and Figure 4.

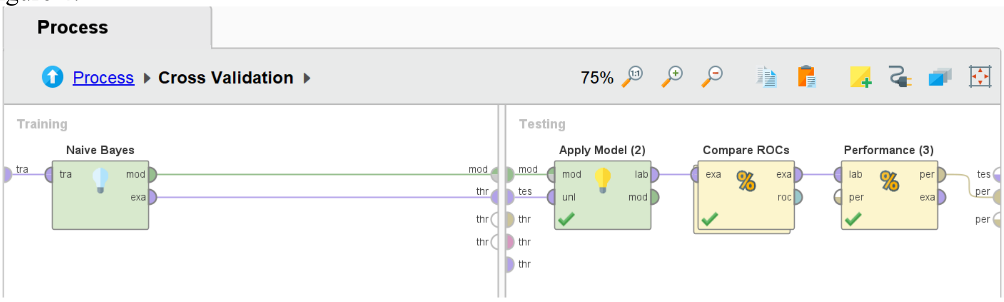

Figure 3. Naïve Bayes Algorithm

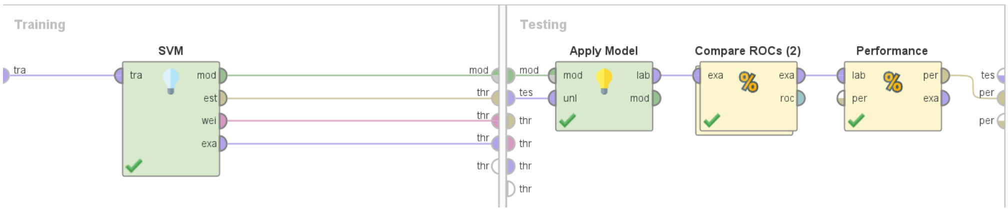

Figure 4. SVM Algorithm

After undergoing the validation process using the Naive Bayes and SVM classification algorithms, the evaluation process will produce the accuracy, recall, precision, and AUC for each model. Figure 5 shows the accuracy, recall, and precision levels of the Naive Bayes algorithm model.

accuracy: 89.21% +/- 1.88% (micro average: 89.21%)

	true negatif	true positif	class precision
pred. negatif	808	63	92.77%
pred. positif	142	887	86.20%
class recall	85.05%	93.37%	

Figure 5. Naive Bayes Model Evaluation Output Table

Based on the evaluation of the Naive Bayes algorithm model results, the accuracy obtained is 89.21%, with a true positive recall of 85.05% and a true negative recall of 93.37%. Meanwhile, the precision for positive predictions is 92.77% and for negative predictions is 86.20%

Figure 6 shows the AUC of the Naive Bayes algorithm model

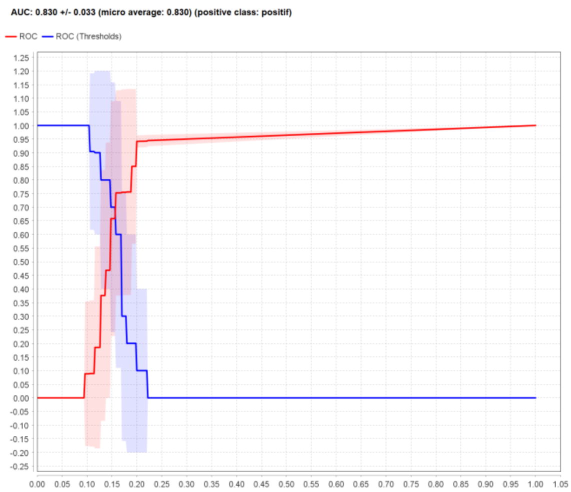

Figure 6. The AUC of the Naive Bayes Algorithm

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

According to this ROC curve, the model used performs very well with an AUC of 0.830, indicating that it is highly effective in distinguishing between positive and negative classes across various classification thresholds. The ROC curve, which is close to the upper left corner and the high AUC value, demonstrates this model's strong predictive ability. Meanwhile, Figure 7 shows the accuracy, recall, and precision levels of the SVM algorithm model.

accuracy: 92.37% +/- 1.84% (micro average: 92.37%)

	true negatif	true positif	class precision
pred. negatif	907	102	89.89%
pred. positif	43	848	95.17%
class recall	95.47%	89.26%	

Figure 7. Evaluation Model Algorithm SVM

Based on the evaluation of the SVM algorithm model results, the accuracy obtained is 92.37%, with a true positive recall of 95.47% and a true negative recall of 89.26%. Meanwhile, the precision for positive predictions is 89.89% and for negative predictions is 95.17%

Figure 8 shows the AUC of the SVM algorithm model.

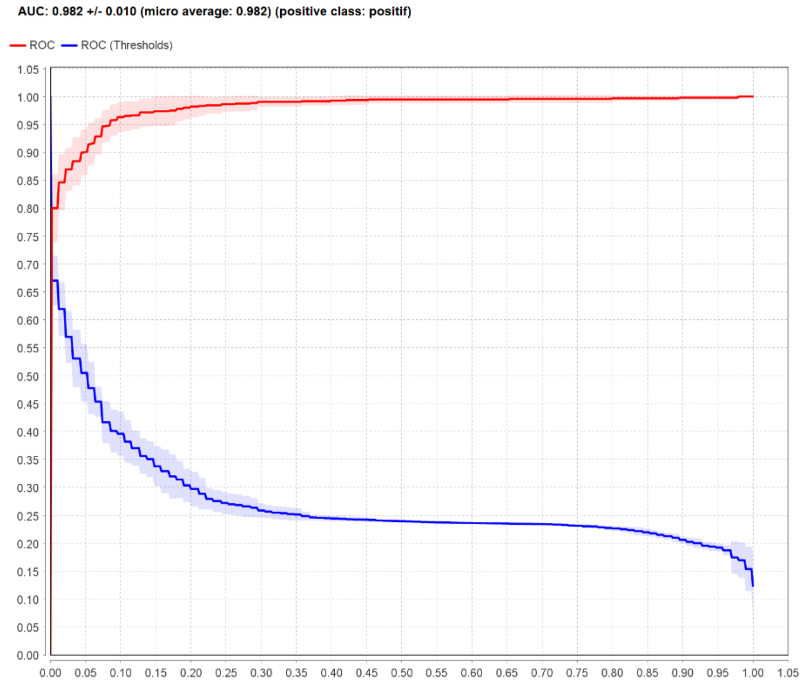

Figure 8. AUC of the SVM Algorithm Model

According to this ROC curve, the model used performs very well with an AUC of 0.982, indicating that it is highly effective in distinguishing between positive and negative classes across various classification thresholds. The ROC curve, which is close to the upper left corner and the high AUC value, demonstrates this model's strong predictive ability.

DISCUSSION

To complete the conclusion with more specific data, including figures and tables from the sentiment measurement results using the SVM and Naive Bayes algorithms, here is an example of details that could be added to the conclusion section of your journal:

Data and Statistical Analysis

Here is a statistical summary that describes the performance of each algorithm in sentiment analysis:

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

Table 1: Statistical Summary of SVM and Naive Bayes Performance

Metric/Model	SVM (%)	Naive Bayes (%)
Accuracy	92.37	89.21
Precision	89.89	92.77
Recall	95.47	85.05
AUC	0.982	0.830

Analysis

SVM: The SVM model exhibits excellent performance with an accuracy of 92.37%, precision of 89.89%, and recall of 95.47%. The very high AUC value (0.982) demonstrates this model's extraordinary ability to distinguish between sentiment classes.

Naive Bayes: Although slightly lower, Naive Bayes also shows good results with an accuracy of 89.21% and high precision (92.77%). A lower recall (85.05%) and AUC (0.830) indicate that this model is less effective compared to SVM in some aspects, particularly in sensitivity towards the positive class.

Implications, From this data, it can be seen that SVM is more suitable for sentiment analysis in the context of flood management in Bekasi City compared to Naive Bayes, especially due to SVM's capability to handle high-dimensional data more effectively. This data provides a foundation for recommendations to the government and disaster management agencies to use the SVM model in their sentiment analysis systems.

Recommendations Based on Data from the Results Obtained, It Is Suggested:

Further Development: The Bekasi City government should integrate the SVM model into the disaster management dashboard to monitor public sentiment in real-time. This will enable authorities to proactively respond to public concerns during flood events.

Data Training: Further data training should be conducted to enhance the model's ability to recognize local nuances, which will improve the accuracy and reliability of sentiment predictions.

CONCLUSION

This study has successfully implemented and compared two machine learning algorithms, namely Support Vector Machine (SVM) and Naive Bayes, in analyzing public sentiment related to flood management in Bekasi City. By using advanced feature selection techniques, this study aims to improve the accuracy in sentiment classification from data obtained from social media platforms.

The conclusion Effectiveness of SVM and Naive Bayes: Both models show good capability in sentiment analysis, but SVM stands out with superior performance compared to Naive Bayes in terms of accuracy, precision, and recall. This indicates that SVM is more effective in handling large and complex textual data commonly found in social media data.

Importance of Feature Selection: The applied feature selection techniques successfully identified the most informative and relevant features, significantly enhancing the effectiveness of both models. Feature selection not only reduces data dimensionality but also strengthens the model in capturing important nuances from sentiment data.

Application for Public Policy: The results of the sentiment analysis provide deep insights into public perceptions of flood management efforts. This information is invaluable for policymakers and disaster management agencies to adapt and refine flood management strategies, especially in public communication and resource allocation.

Strategic Recommendations: Based on the study results, it is suggested that the Bekasi City government and related agencies be more proactive in utilizing sentiment analysis to understand the needs and responses of the community during disaster situations. Integrating this sentiment analysis system into the disaster management dashboard can be an effective tool for faster and more accurate planning and response.

This research opens opportunities for further studies in developing more complex sentiment analysis models and integrating other AI technologies to assist in disaster management in Indonesia. The implementation of such technologies is expected to make a real contribution in reducing disaster impacts and enhancing community resilience against natural disasters in the future.

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to all parties who have provided support throughout this research process. Special thanks go to:

Research Colleagues, for their collaboration in data collection and analysis, and for providing critical feedback that helped refine this analysis. The intense collaboration and discussions have significantly improved the quality of this research.

The Informatics Engineering Study Program, Faculty of Engineering, Universitas Pelita Bangsa, for the facilities and resources provided, which enabled this research to be conducted smoothly. Thank you for the infrastructure support

E-ISSN : 2807-9035
Volume 4, **Number** 1, May 2024
https://doi.org/10.47709/brilliance.v4i1.4268

and access to the necessary analysis software for this research.

The Directorate of DPPM, whose financial support made this research possible. This financial support provided the authors with sufficient resources to carry out this research effectively.

The authors also wish to thank all parties not mentioned individually who have provided support, both directly and indirectly, in the completion of this research.

REFERENCES

Adisaputra Sinaga, N., & Surya Gunawan, T. (n.d.). *Sentiment Analysis on Hotel Ratings Using Dynamic Convolution Neural Network*.

Amali. (2020). PELITA TEKNOLOGI Perbandingan Algoritma Sentimen Analisis Media Data Twitter Pilgub Jabar 2018. *Jurnal Pelita Teknologi*, *15*(1), 26–36.

Ari Nasichuddin, M., & Bharata Adji, T. (2018). Performance Improvement Using CNN for Sentiment Analysis. In *IJITEE* (Vol. 2, Issue 1).

Basha, M. J., Vijayakumar, S., Jayashankari, J., Alawadi, A. H., & Durdona, P. (2023). Advancements in Natural Language Processing for Text Understanding. *E3S Web of Conferences*, *399*. https://doi.org/10.1051/e3sconf/202339904031

Firdausi, I. E., Mukhlash, I., Gama, A. D. S., & Hidayat, N. (2020). Sentiment analysis of customer response of telecommunication operator in Twitter using DCNN-SVM Algorithm. *Journal of Physics: Conference Series*, *1490*(1). https://doi.org/10.1088/1742-6596/1490/1/012071

Fu, E., Xiang, J., & Xiong, C. (2022). Deep Learning Techniques for Sentiment Analysis. In *Highlights in Science, Engineering and Technology AMMSAC* (Vol. 2022).

Gad, A. G. (2022). Particle Swarm Optimization Algorithm and Its Applications: A Systematic Review. *Archives of Computational Methods in Engineering*, *29*(5), 2531–2561. https://doi.org/10.1007/s11831-021-09694-4

Guo, X., Yu, W., & Wang, X. (2021). An Overview on Fine-grained Text Sentiment Analysis: Survey and Challenges. *Journal of Physics: Conference Series*, *1757*(1). https://doi.org/10.1088/1742-6596/1757/1/012038

Kausar, M. A., Soosaimanickam, A., & Nasar, M. (2021). Public Sentiment Analysis on Twitter Data during COVID-19 Outbreak. In *IJACSA) International Journal of Advanced Computer Science and Applications* (Vol. 12, Issue 2). www.ijacsa.thesai.org

Kiran Kumar, P., Jahna Tejaswi, N., Vasanthi, M. L., Srihitha, L. L., & Phanindra Kumar, B. (2022). *Sentimental Analysis on Multi-domain Sentiment Dataset Using SVM and Naive Bayes Algorithm* (pp. 201–213). https://doi.org/10.1007/978-3-030-95502-1_16

Makmun, M., Zy, A. T., & Arwan, A. (2024). *Analisis Sentimen Media Sosial Twitter Terhadap Calon Presiden RI Tahun 2024 Menggunakan Klasifikasi Algoritma Naïve Bayes*. https://doi.org/10.47065/josyc.v5i3.5210

Mathieu Cliche, & Bloomberg. (2017). *Twitter sentiment analysis with CNNs and LSTMs*. https://doi.org/https://doi.org/10.48550/arXiv.1704.06125

Paredes-Valverde, M. A., Colomo-Palacios, R., Salas-Zárate, M. D. P., & Valencia-García, R. (2017). Sentiment Analysis in Spanish for Improvement of Products and Services: A Deep Learning Approach. *Scientific Programming*, *2017*. https://doi.org/10.1155/2017/1329281

Rahman, H., Tariq, J., Ali Masood, M., F. Subahi, A., Ibrahim Khalaf, O., & Alotaibi, Y. (2023). Multi-Tier Sentiment Analysis of Social Media Text Using Supervised Machine Learning. *Computers, Materials & Continua*, *74*(3), 5527–5543. https://doi.org/10.32604/cmc.2023.033190

Samsir, Irmayani, D., Edi, F., Harahap, J. M., Jupriaman, Rangkuti, R. K., Ulya, B., & Watrianthos, R. (2021). Naives Bayes Algorithm for Twitter Sentiment Analysis. *Journal of Physics: Conference Series*, *1933*(1). https://doi.org/10.1088/1742-6596/1933/1/012019

Thoriq, M. F., Pranoto, W. J., & Faldi, F. (2023). Penerapan Seleksi Fitur Analysis of Variance Pada Algoritma Random Forest Classifier Dalam Klasifikasi Nilai Mahasiswa. *Explore: Jurnal Sistem Informasi Dan Telematika*, *14*(2), 185. https://doi.org/10.36448/jsit.v14i2.3187