Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

Sentiment and Toxicity Analysis of Biometric Authentication and Facial Recognition Technology Content Reviews using Cross-Industry Standard Process for Data-Mining

Yerik Afrianto Singgalen

Faculty of Business Administration and Communication, Tourism Study Program, Atma Jaya Catholic University of Indonesia, Jakarta, Indonesia
Email: yerik.afrianto@atmajaya.ac.id
Correspondence Author Email: yerik.afrianto@atmajaya.ac.id
Submitted: 10/05/2024; Accepted: 29/05/2024; Published: 30/05/2024

Abstract−This study investigates sentiment analysis methodologies within the framework of CRISP-DM (Cross-Industry Standard Process for Data Mining), aiming to discern the efficacy of various algorithms in sentiment classification tasks. The research uses a structured approach to evaluate SVM, NBC, DT, and K-NN algorithms with the SMOTE oversampling technique, uncovering distinct performance metrics and limitations. Results indicate SVM achieving 59.88% accuracy, NBC at 59.25%, DT with 52.09%, and K-NN obtaining 54.80%, highlighting the differential precision, recall, and f-measure. Additionally, content analysis identifies pertinent themes such as Biometric security, Cloud storage, and Emotion Analysis, enriching sentiment dynamics comprehension. The toxicity scores of analyzed videos reveal nuanced sentiment nuances, with the first video exhibiting Toxicity: 0.13227 and the second scoring Toxicity: 0.12794. This study underscores the significance of informed algorithm selection and evaluation methodologies within CRISP-DM, fostering optimized sentiment analysis outcomes while acknowledging diverse topical nuances.

Keywords: Sentiment; Reviews; Toxicity; Biometric; Security

1. INTRODUCTION

Utilizing biometric data, including face-recognition technology, across various sectors has sparked controversy regarding security. In recent years, the proliferation of biometric systems in sectors ranging from finance to law enforcement has raised concerns about the vulnerability of sensitive personal data [1]–[4]. Critics argue that the reliance on biometric identifiers, such as facial features, exposes individuals to heightened risks of identity theft and unauthorized surveillance [5]–[8]. Moreover, the potential for data breaches and misuse underscores the imperative for robust security protocols and stringent regulatory frameworks to safeguard privacy rights and mitigate potential abuses [9]–[14]. Despite promises of enhanced efficiency and convenience, deploying biometric technologies necessitates vigilant oversight to ensure adherence to ethical principles and respect for individual autonomy.

The debate surrounding the development of technologies utilizing biometric data is evident in responses to content addressing biometric data and security systems in videos. Advocates highlight the potential of biometric technologies to bolster security measures through enhanced authentication processes, minimizing the risks associated with traditional methods like passwords or PINs [15]–[17]. Moreover, proponents argue that biometric systems offer unparalleled accuracy and efficiency in identifying individuals, thus facilitating seamless access to restricted areas or sensitive information [18]–[20]. However, detractors raise concerns regarding the susceptibility of biometric data to breaches and misuse, emphasizing the need for robust safeguards to protect individuals' privacy and prevent unauthorized access [21], [22]. Furthermore, critics caution against the potential for biometric systems to perpetuate discrimination or bias, particularly in cases where algorithms exhibit inaccuracies or biases against certain demographic groups [23]–[25]. In conclusion, while the development of biometric technologies holds promise for improving security measures, it is imperative to address the associated ethical and privacy considerations to ensure responsible and equitable implementation.

This research aims to identify viewer sentiments towards videos concerning biometric data and security to analyze viewer responses regarding digital technology and its utilization across various sectors. By examining viewer reactions, the study seeks to gauge public perception and attitudes toward integrating biometric technologies into security systems and the implications for diverse industries [26]–[28]. Understanding viewer sentiment provides valuable insights into societal acceptance, concerns, and expectations regarding digital technologies, informing policy decisions and industry practices [29], [30]. Ultimately, this research contributes to the discourse surrounding biometric data usage's ethical, social, and technological dimensions, fostering informed debates and responsible technological development.

The method employed to address the research problem is CRISP-DM (Cross-Industry Standard Process for Data Mining), a widely recognized framework for data mining projects. CRISP-DM provides a systematic approach encompassing six phases: business understanding, data understanding, data preparation, modeling, evaluation, and deployment [31]. This structured methodology allows for comprehensive data exploration and analysis, facilitating informed decision-making and the development of practical solutions [32]. Adhering to the CRISP-DM framework, this research navigates the complexities of the research process, ensuring methodological rigor and maximizing the likelihood of achieving meaningful insights and outcomes.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

The urgency of this research lies in its potential to address pressing societal challenges and inform strategic decision-making in an increasingly digitalized world. As technology advances rapidly, integrating biometric data and security systems pervades numerous sectors, ranging from finance to healthcare [33], [34]. Understanding public sentiment and attitudes towards these technologies is crucial for policymakers, industry leaders, and other stakeholders to navigate ethical, legal, and social implications effectively [35], [36]. By elucidating the complexities surrounding adopting and utilizing biometric technologies, this research fosters responsible innovation and ensures equitable outcomes for individuals and society.

Theoretical and practical implications of this research extend across interdisciplinary domains, offering valuable insights into the ethical, social, and technological dimensions of biometric data utilization. From a theoretical standpoint, the findings advance knowledge in digital ethics, privacy studies, and technology adoption theories by enriching existing frameworks with empirical evidence and nuanced understandings of public attitudes [37], [38]. Furthermore, the practical implications of this research are far-reaching, informing policy formulation, industry practices, and the design of biometric systems to align with societal values and expectations [39], [40]. This research facilitates informed decision-making and responsible innovation by bridging the gap between theory and practice, fostering a more equitable and sustainable digital future.

Examining similar research and acknowledging its limitations is essential for contextualizing the contributions and scope of this study. Existing literature on biometric data and security systems offers valuable insights into public perceptions, technological advancements, and regulatory frameworks, providing a foundation for further inquiry [41]–[46]. However, limitations such as sample biases, methodological constraints, and rapidly evolving technologies underscore the need for cautious interpretation and continued inquiry [47]–[50]. By building upon existing research while addressing its limitations, this study aims to offer novel perspectives and contribute to the ongoing discourse surrounding the ethical, social, and technological implications of biometric data utilization in diverse contexts.

2. RESEARCH METHODOLOGY

2.1 Gap Analysis of Biometric Authentication and Face Recognition Technology

Gap analysis is conducted to discern disparities within previous research addressing biometric data and security, aiming to identify areas warranting further investigation and refinement. This process unveils overlooked dimensions, unresolved questions, and emerging trends by scrutinizing existing literature, thereby guiding research objectives and methodological approaches. Through meticulous examination, this research pinpoints gaps in knowledge and methodologies, facilitating the advancement of scholarship and developing more comprehensive and nuanced understandings of biometric data and security systems. Consequently, gap analysis enriches the scholarly discourse and enhances the relevance and impact of subsequent research endeavors.

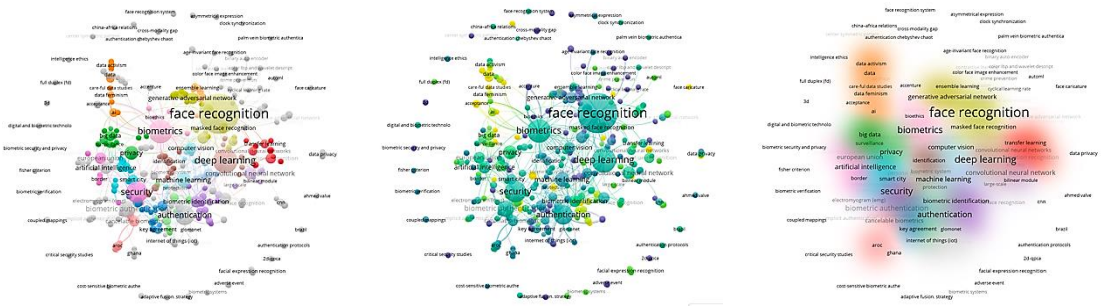

Figure 1. Network, Density, and Overlay Visualization of Biometric Data Security

Figure 1 shows the network, density, and overlay visualization of biometric data security. The results of the gap identification indicate that studies focusing on biometric data security using sentiment analysis and toxicity assessment through CRISP-DM are still relatively scarce. Despite the growing interest in biometric technologies and security systems, a notable dearth of research employing comprehensive analytical frameworks like CRISP-DM to explore these technologies' sentiments and toxicity levels remains. This observation underscores the need for further investigations that leverage robust methodologies to comprehensively assess public attitudes, perceptions, and potential risks about biometric data security. Addressing this research gap advances our understanding of biometric data utilization's ethical, social, and technical dimensions, informing policy development and technological innovation in this domain.

Drawing upon the insights from gap analysis, this research exhibits distinct advantages in addressing the lacunae within the current scholarly landscape. This study offers a comprehensive understanding of public perceptions and concerns regarding biometric data security by adopting a multifaceted approach that integrates sentiment analysis and toxicity assessment within the CRISP-DM framework. Furthermore, utilizing such a methodological framework enables systematic data exploration, modeling, and evaluation, thereby enhancing the

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

robustness and reliability of the findings. Consequently, this research stands poised to make significant contributions to the field by bridging existing gaps in knowledge and methodology, ultimately fostering informed decision-making and responsible innovation in biometric data security.

2.2 Cross-Industry Standard Process for Data Mining (CRISP-DM)

The CRISP-DM framework is the backbone for analyzing viewer sentiments towards video content concerning biometric data and security. This methodological approach encompasses several phases, including business understanding, data understanding, data preparation, modeling, evaluation, and deployment, facilitating a systematic and iterative sentiment analysis process. This research methodically collects and preprocess data through this framework, identifies relevant features, and develops predictive models to discern underlying sentiment patterns. Moreover, the evaluation phase enables the assessment of model performance and the refinement of analytical techniques to ensure the accuracy and reliability of the sentiment analysis results. Thus, leveraging the CRISP-DM framework in analyzing viewer sentiments offers a rigorous and structured methodology for deriving meaningful insights into public perceptions and attitudes toward biometric data and security issues.

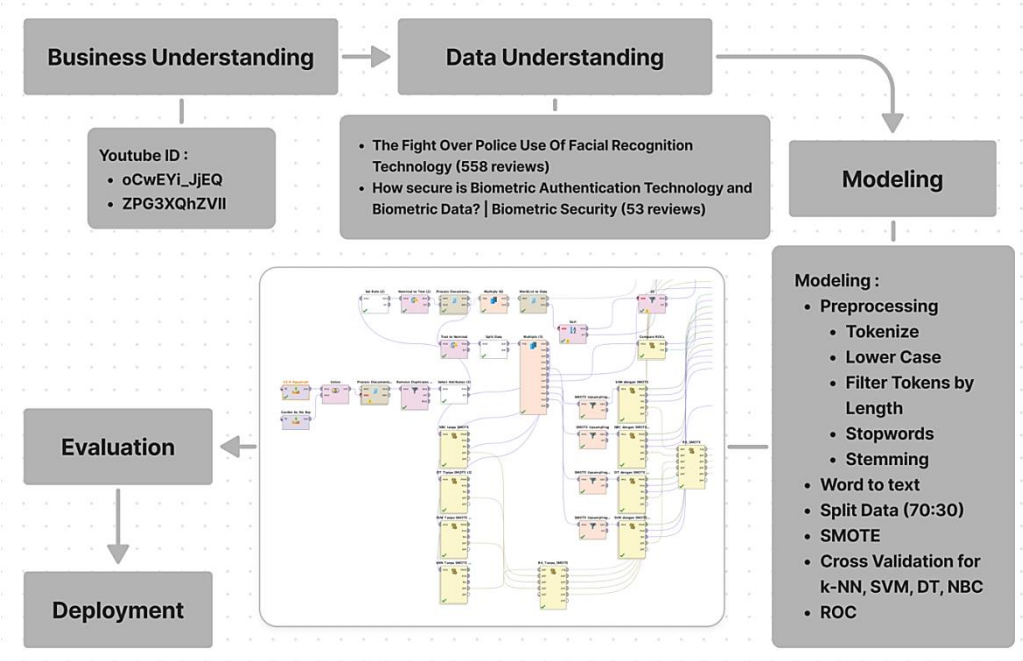

Figure 2. Implementaiton of CRISP-DM Framework

Figure 2 shows the implementation of the CRISP-DM framework. CRISP-DM presents several advantages in the contextual collection and processing of textual data. Firstly, its structured approach facilitates systematic data gathering by identifying relevant sources and variables pertinent to the research objectives. Secondly, the framework provides a systematic methodology for preprocessing textual data, including tasks such as data cleaning, tokenization, and stemming, ensuring the consistency and quality of the dataset. CRISP-DM's iterative nature allows for continual refinement of data processing techniques, enabling this research to adapt to evolving data characteristics and analytical requirements. Consequently, leveraging the CRISP-DM framework enhances the efficiency and effectiveness of data collection and processing in textual analysis, ultimately contributing to more robust and insightful research outcomes in biometric data and security.

The relevance and superiority of CRISP-DM in sentiment and toxicity analysis stand as primary considerations. Firstly, the framework's systematic methodology ensures a structured approach to data preprocessing, modeling, and evaluation, enhancing the accuracy and reliability of sentiment and toxicity assessments. Secondly, CRISP-DM's iterative nature allows for continual refinement of analytical techniques, enabling this research to adapt to evolving data characteristics and analytical requirements in the dynamic landscape of sentiment and toxicity analysis. Consequently, leveraging CRISP-DM affords a robust and comprehensive framework for conducting nuanced and insightful analyses of sentiment and toxicity, ultimately contributing to a deeper understanding of public perceptions and attitudes toward biometric data and security issues.

2.2.1 Business Understanding

In the business understanding phase, it is imperative to comprehend the contextual nuances of the data and discussions that center on biometric data and security. Understanding the business context entails delineating the

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

objectives, stakeholders, and constraints pertinent to the research endeavor, laying the groundwork for informed decision-making and resource allocation. Focusing on biometric data and security necessitates a comprehensive grasp of the regulatory landscape, technological advancements, and societal concerns surrounding these domains, ensuring the relevance and applicability of subsequent analyses and insights. Consequently, prioritizing a thorough understanding of the business context at this initial phase facilitates formulating research goals and strategies aligned with the overarching objectives of enhancing security measures and safeguarding privacy rights in the digital age.

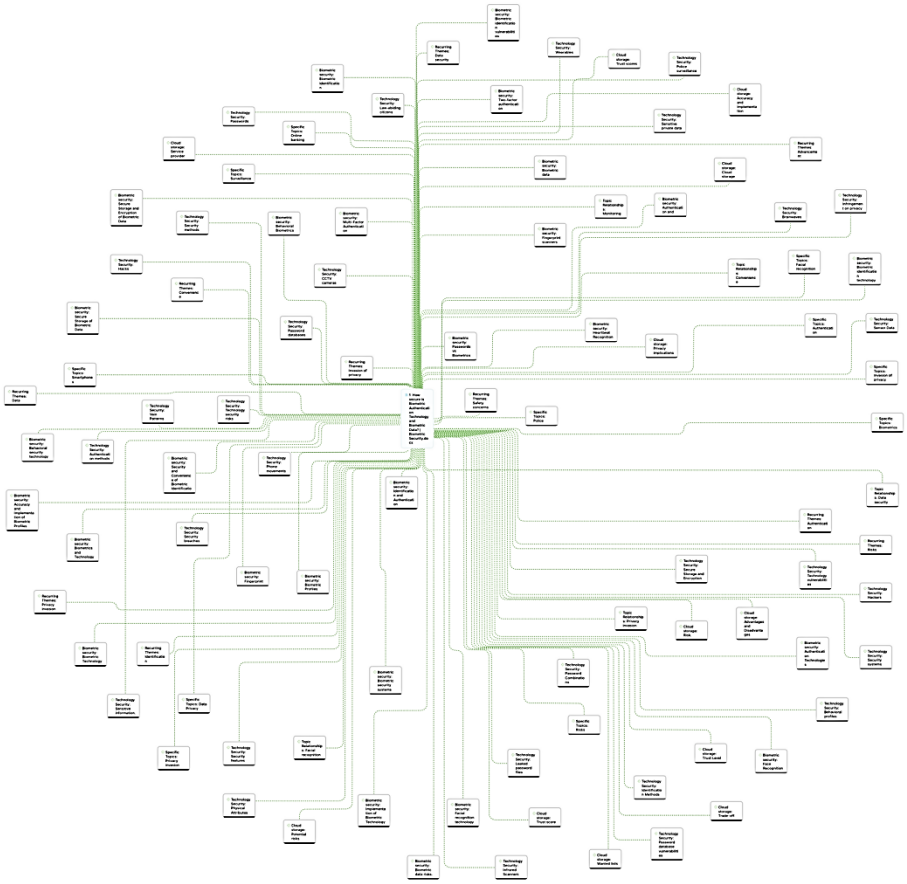

Figure 3. Topics Mentioned in the Video (ZPG3XQhZVII)

Figure 3 shows the topics in the video entitled "How secure is Biometric Authentication Technology and Biometric Data? | Biometric Security" (oCwEYi_JjEQ). The video titled "How secure is Biometric Authentication Technology and Biometric Data? | Biometric Security" (id=ZPG3XQhZVII) delves into various topics relevant to biometric security and technology. The video covers themes such as biometric security, cloud storage, emotion analysis, emotion conveyance, keywords for emotions, negative attitudes or reception, recurring themes, regulation, specific topics, and technology security, and the video comprehensively explores the intricate interplay between biometric authentication technology and security concerns. By addressing these diverse topics, the video offers viewers a nuanced understanding of the complexities inherent in utilizing biometric data and the imperative for robust regulatory frameworks to ensure privacy protection and technological integrity. Consequently, the video is a valuable resource for individuals to understand biometric security and its implications in contemporary digital landscapes.

In addition, topics related to biometric authentication discussed in the video encompass various issues and technologies. These include authentication methods, behavioral profiles, brainwaves, CCTV cameras, hackers, hacks, identification methods, infrared scanners, privacy infringement, law-abiding citizens, leaked password files, password combinations, password database vulnerabilities, passwords, phone movements, physical attributes, police surveillance, secure storage and encryption, security breaches, security features, security methods, security systems, sensitive information, sensitive private data, sensor data, technology security risks, technology vulnerabilities, vein patterns, and wearables. This comprehensive array of topics underscores the multifaceted nature of discussions surrounding biometric authentication and its implications for privacy, security, and technological advancement. Consequently, it emphasizes the necessity for comprehensive and informed approaches to address biometric authentication technologies' myriad challenges and opportunities.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

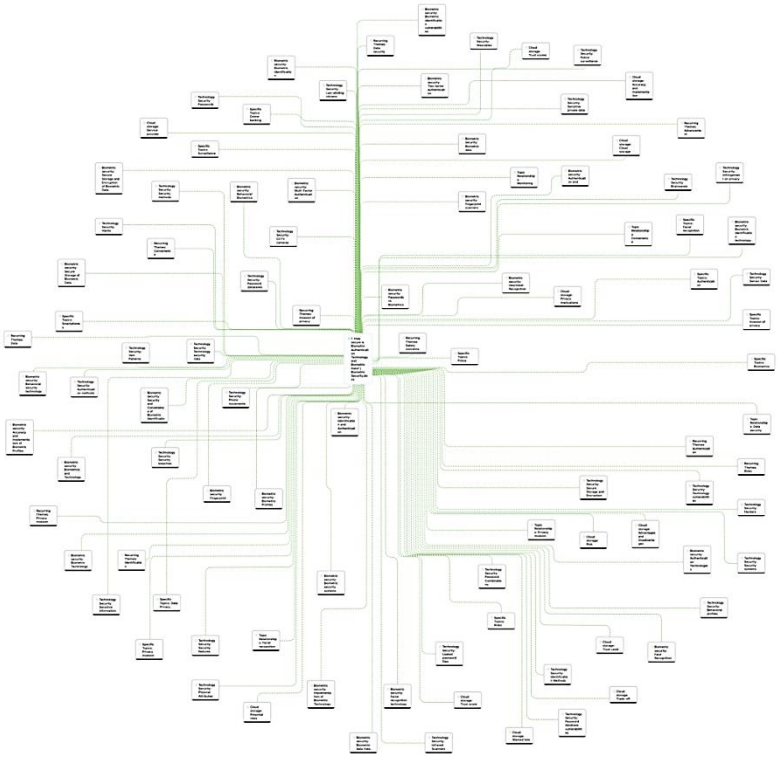

Figure 4. Emotion Conveyance in the Video (oCwEYi_JjEQ)

Figure 4 shows the topics in the video entitled "The Fight Over Police Use Of Facial Recognition Technology" (oCwEYi_JjEQ). The emotion analysis of the video titled "The Fight Over Police Use Of Facial Recognition Technology" (id=oCwEYi_JjEQ) reveals a diverse array of sentiments among viewers, including concern, determination, discomfort, excitement, fear, frustration, hope, opposition, relief, skepticism, uncertainty, and worry. These emotions reflect the multifaceted nature of discussions surrounding using facial recognition technology by law enforcement agencies, highlighting the potential benefits and risks associated with its implementation. While some viewers express hope and excitement about the technology's potential to enhance public safety and security, others voice concerns and skepticism regarding its potential for abuse, privacy invasion, and discriminatory practices. The presence of emotions such as determination, opposition, and frustration underscores the intensity of the debate surrounding the regulation and oversight of facial recognition technology, emphasizing the need for balanced and informed decision-making to address societal concerns effectively.

Emotion conveyance is characterized by a range of sentiments and linguistic elements, including concern, criticism, disagreement, doubt, informative language use, lack of control, negative tone, neutral tone, skepticism, and word choice. These elements collectively shape the tone and emotional impact of the video, influencing viewers' perceptions and attitudes towards the contentious issue of police utilization of facial recognition technology. Critical and skeptical tones reflect the topic's complexity and controversy, while informative language contributes to a deeper understanding of the underlying issues at stake. Moreover, using word choice and neutral tones enhances the objectivity and credibility of the discourse, facilitating balanced and nuanced discussions. Overall, the emotional conveyance of the video engages viewers intellectually and emotionally, prompting reflection and debate on the ethical, legal, and societal implications of facial recognition technology in law enforcement contexts.

Upon grasping the discussion context, progression to the data understanding phase becomes imperative. This transition marks a pivotal juncture in the research process, where efforts are directed toward comprehending available data sources' characteristics, quality, and relevance. By delving into data understanding, this research gleans insights into the structure, format, and potential biases inherent in the data, laying a solid foundation for subsequent analysis and interpretation. Moreover, this phase enables this research to discern patterns, relationships, and anomalies within the data, formulating research questions and hypotheses. Thus, advancing to the data understanding phase is essential for fostering methodological rigor and ensuring the validity and reliability of research findings.

2.2.2 Data Understanding

During the data understanding phase, it is imperative to identify the quantity and sources of textual data and the processes involved in data cleansing and extraction. This initial step lays the groundwork for subsequent analysis

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

by ensuring the data is representative, relevant, and of sufficient volume to yield meaningful insights. Moreover, identifying data sources allows this research to assess the credibility and reliability of the information, thereby enhancing the validity of the findings. Additionally, implementing robust data cleansing and extraction procedures is essential for mitigating noise, inconsistencies, and biases in the dataset, ultimately facilitating accurate and reliable analysis. Thus, prioritizing the identification and preparation of textual data sets the stage for rigorous and insightful research outcomes in biometric authentication and security.

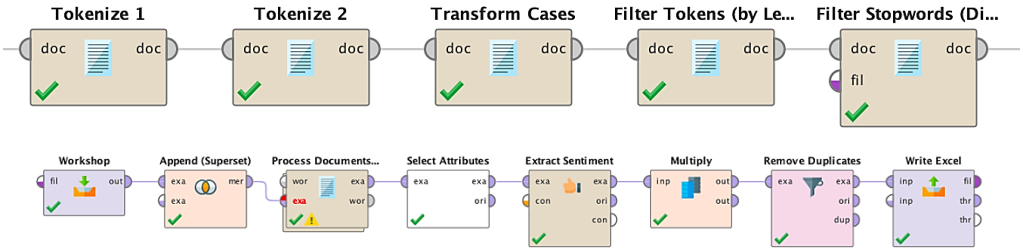

Figure 5. Data Cleaning and Extract Sentiment

Figure 5 shows the data cleaning and extract sentiment process in Rapidminer. The data cleaning involves several key steps, including tokenization, transforming cases, filtering tokens by length, and removing stopwords in both English and Indonesian. With 558 data points extracted from the video with id oCwEYi_JjEQ and 53 from the video with id ZPG3XQhZVII, this process ensures the standardization and refinement of textual data for subsequent analysis. Tokenization involves breaking down the text into individual words or tokens while transforming cases to ensure uniformity by converting all text to lowercase. Filtering tokens by length helps remove irrelevant or insignificant words, while removing stopwords eliminates common words that do not carry meaningful information. By implementing these techniques, the data cleaning process enhances the quality and usability of the dataset, facilitating more accurate and insightful analysis of the content discussed in the videos.

Figure 6. Frequently used Words (Video id ZPG3XQhZVII)

Figure 6 shows the words of video d ZPG3XQhZVII. Specific terms recur with notable frequency. Among these, "can" appears most frequently, occurring 11 times, followed by "data" and "nt," each appearing seven times. Additionally, terms such as "biometric," "need," "password," and "technology" are frequently utilized, each appearing six times. These findings underscore the salience of themes related to biometric authentication, data security, and technological advancements in the discourse captured by the video. Furthermore, the prevalence of terms like "hackers," "security," and "biometrics" suggests a heightened awareness of cybersecurity threats and the importance of robust security measures in safeguarding sensitive information. Consequently, the frequent recurrence of these words reflects the central focus and critical concerns addressed in the video regarding the utilization of biometric technology and its implications for security and privacy.

Figure 7. Frequently used Words (Video id oCwEYi_JjEQ)

Figure 7 shows the words of video d oCwEYi_JjEQ. Specific terms are recurrently employed. Among these, "can" is the most frequently used word, appearing 88 times, followed closely by "people" and

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

"recognition," each appearing 87 and 76 times, respectively. Notably, terms such as "facial recognition," "police," and "privacy" feature prominently in the discourse, suggesting a predominant focus on the utilization of facial recognition technology by law enforcement agencies and its implications for privacy rights. Additionally, the repeated mention of words like "technology," "world," and "criminals" underscores broader societal concerns regarding the impact of technological advancements on security and justice systems. The frequent recurrence of these terms reflects the depth and breadth of discussions surrounding facial recognition technology and its societal ramifications, emphasizing the need for informed discourse and regulatory measures to address ethical and legal considerations.

Upon understanding the characteristics of the data, the subsequent process involves the implementation of algorithms for modeling. This pivotal step translates theoretical understanding into practical application, where various algorithms are employed to analyze, interpret, and derive insights from the data. By utilizing appropriate modeling techniques, this research uncovers patterns, relationships, and trends within the dataset, thereby facilitating the generation of meaningful predictions or classifications. Moreover, implementing algorithms allows for the refinement and validation of hypotheses, enabling this research to draw reliable conclusions and make informed decisions based on the empirical evidence obtained. Thus, implementing algorithms for modeling represents a critical stage in the research process, bridging the gap between theory and practice to facilitate robust analysis and interpretation of data.

2.2.3 Modeling

During the modeling phase, algorithms such as k-NN, DT, SVM, and NBC are tested to analyze and predict outcomes based on the dataset. Additionally, the SMOTE operator is employed to address data imbalance issues. This process involves generating synthetic samples for the minority class, thus equalizing the distribution of data points across different classes. By implementing these algorithms and techniques, this research explores various modeling approaches' effectiveness in accurately representing and predicting real-world phenomena. Moreover, utilizing the SMOTE operator enhances the robustness of the analysis by mitigating the impact of data imbalance, thereby improving the overall reliability and validity of the modeling results.

The division of data into training and testing sets is typically set at 30% for training data and 70% for testing data, with a higher percentage allocated to testing data to analyze the performance of machine learning models. This allocation ensures sufficient data for testing the model's generalizability and predictive accuracy on unseen data. By allocating a more significant proportion of the dataset to testing, this research effectively evaluates the model's ability to generalize to new observations and assess its performance under real-world conditions. This approach enhances the reliability and validity of the model evaluation process, enabling this research to make informed decisions regarding the model's effectiveness and suitability for practical applications.

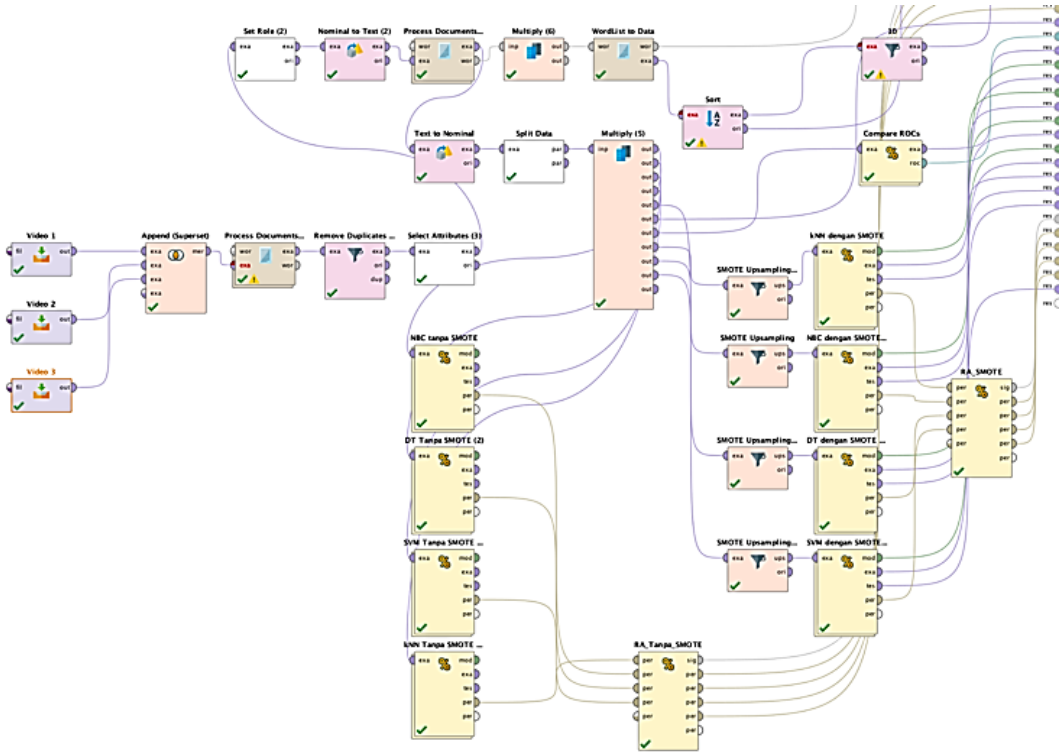

Figure 6. Implementation of k-NN, DT, NBC, and SVM with SMOTE

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

Figure 6 shows the Implementation of k-NN, DT, NBC, and SVM with SMOTE. Based on the results of implementing the k-NN, DT, NBC, and SVM algorithms in RapidMiner, it becomes apparent which algorithm performs best in sentiment classification. Through rigorous testing and evaluation, one algorithm performs better in accurately classifying sentiment within the dataset. This outcome underscores the significance of methodical experimentation and analysis in identifying the most effective algorithm for a specific task, providing valuable insights into optimizing machine learning models for sentiment analysis applications.

The results of the modeling process will undergo evaluation to generate pertinent recommendations. This evaluation encompasses a comprehensive analysis of the model's performance metrics, including accuracy, precision, recall, and F1-score, among others, to assess its effectiveness in achieving the desired objectives. Additionally, the evaluation process involves comparing the model's performance against predefined benchmarks or industry standards to determine its suitability and reliability in practical applications. Through rigorous evaluation, stakeholders derive actionable insights and recommendations to inform decision-making processes and optimize outcomes in relevant domains.

2.2.4 Evaluation

In the evaluation stage, the model is assessed based on the values of the confusion matrix, including accuracy, precision, recall, F-measure, and Area Under the Curve (AUC). These metrics comprehensively understand the model's performance across various classification or prediction tasks. Accuracy measures the overall correctness of predictions, while precision quantifies the proportion of correctly predicted positive instances among all instances predicted as positive. Conversely, recall represents the proportion of correctly predicted positive instances among all positive instances. F-measure combines precision and recall into a single metric, offering a balanced assessment of the model's performance. Additionally, the AUC metric evaluates the model's ability to distinguish between different classes, providing insights into its discriminative power. Together, these evaluation metrics offer valuable insights into the strengths and limitations of the model, facilitating informed decision-making and optimization efforts.

Subsequently, viewer comments on the video will be analyzed based on toxicity scores according to the Perspective API model. This approach involves natural language processing techniques to assess toxicity or harmfulness in viewer comments, enabling a nuanced understanding of the sentiment and tone expressed. Leveraging advanced algorithms and linguistic analysis, the Perspective API model quantifies the toxicity in comments, ranging from mild disagreement to severe harassment or abuse. This analysis provides valuable insights into the nature of viewer engagement and sentiment surrounding the video content, informing content creators and platform moderators about potential issues or areas for improvement.

2.2.5 Deployment

During the deployment phase, the content analysis findings are discussed based on the results of axial coding from the Atlas. Ti application to identify aspects related to biometric security, cloud storage, emotion analysis, emotion conveyance, keywords for emotions, negative attitudes or receptions, recurring themes, regulation, specific topics, technology security, and topic relationships. This systematic approach allows for a comprehensive examination of the identified themes and patterns within the analyzed content, providing valuable insights into the interconnectedness and significance of various topics about the subject matter. Additionally, by aligning the discussion with the results of axial coding, the analysis process remains grounded in empirical evidence, enhancing the credibility and rigor of the findings.

The toxicity score calculation results are adjusted according to toxicity, severe toxicity, identity attack, insult, profanity, and threat. This comprehensive approach enables a nuanced assessment of the harmfulness and aggression levels in the analyzed content, encompassing various dimensions of negative interaction and potential harm. By incorporating multiple metrics, including toxicity and subcategories, the evaluation process becomes more robust and tailored to capture the diverse manifestations of inappropriate or harmful behavior within the analyzed data. This meticulous consideration of different aspects of toxicity enhances the accuracy and reliability of the toxicity scoring system, providing valuable insights for content moderation and community management efforts.

3. RESULT AND DISCUSSION

The discussion is divided into content analysis, toxicity score assessment, and model performance evaluation in sentiment classification. This segmentation allows for comprehensively examining the analyzed content's characteristics and identifying potentially harmful or inappropriate elements through toxicity scoring. Additionally, it facilitates an in-depth assessment of the effectiveness and accuracy of the developed models in classifying sentiments expressed within the content. By systematically addressing these aspects, a more thorough understanding of the content's nature and the model's performance is attained, aiding in informed decision-making and further refining the analytical processes.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

3.1 Toxicity and Content Analysis

Identifying topics related to the video reveals several critical areas of concern, including convenience, data security, facial recognition, monitoring, and privacy invasion. These topics underscore the multifaceted nature of discussions surrounding biometric authentication technology and the broader implications for individuals' privacy and security. By shedding light on issues such as the balance between convenience and security, the risks associated with data breaches, the potential implications of facial recognition technology, the challenges of monitoring systems, and the ethical considerations surrounding privacy invasion, the video contributes to a nuanced understanding of the complexities inherent in biometric security discussions. As such, it is a valuable resource for individuals navigating the intricate landscape of biometric technology and its societal impacts.

Identifying specific topics from video elucidates key focal points encompassing authentication, biometrics, data privacy, facial recognition, invasion of privacy, online banking, police, risks, smartphones, and surveillance. These topics encapsulate the diverse concerns and considerations about biometric authentication technology and its implications for privacy, security, and societal dynamics. By delineating these specific focus areas, the video provides viewers with a nuanced understanding of the intricate interplay between technological advancements, regulatory frameworks, and individual rights in biometric security. Consequently, it serves as a comprehensive resource for individuals seeking to navigate the complexities of biometric technology and its broader societal impacts.

Identifying regulatory topics from the video highlights crucial considerations such as balanced views, cautious language, comfort with regulation, responsibility, lack of regulation, potential misuse, profound implications, promises, scrutiny, specific issues, and urgency for regulation. These regulatory themes underscore the complex interplay between technological advancements, ethical considerations, and legal frameworks governing biometric authentication technology. By addressing these regulatory topics, the video offers viewers insights into the necessity for comprehensive and balanced regulatory approaches to mitigate risks, safeguard individual rights, and foster responsible innovation in biometric security. Consequently, it underscores the imperative for policymakers, industry stakeholders, and regulatory bodies to prioritize developing and implementing robust regulatory frameworks to address biometric technology's profound implications effectively.

Identifying negative attitudes based on the identified topics reveals a spectrum of sentiments ranging from the boldest stance to suggestions of eventual comfort. Viewers express feelings of being creeped out, critical, and upset, highlighting fundamental concerns about biometric authentication technology's compatibility and potential misuse. The uncertain tone suggests a lack of confidence or trust in such technologies' efficacy and ethical implications. At the same time, references to being tracked evoke apprehension about surveillance and privacy invasion. Furthermore, virtue signaling suggests a perceived insincerity or superficiality in discussions surrounding biometric security, indicating skepticism or distrust toward the motives driving the discourse. Collectively, these negative attitudes underscore the need for careful consideration of ethical, social, and regulatory dimensions in developing and deploying biometric authentication technology, aiming to address concerns and build trust among stakeholders.

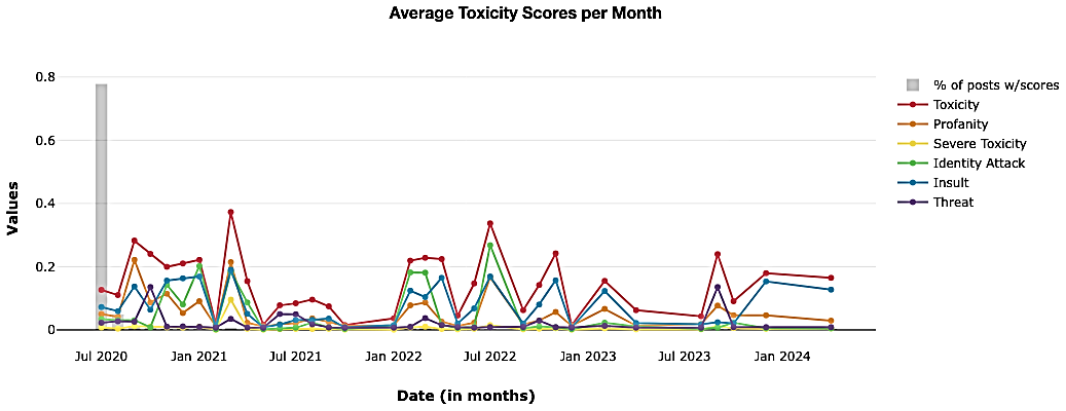

Figure 7. Toxicity Score of the First Video (oCwEYi_JjEQ)

Figure 7 shows the toxicity score of the first video (oCwEYi_JjEQ). The toxicity scores derived from the first video indicate the following: Toxicity score is 0.13227 with a threshold of 0.84032, Severe Toxicity is 0.00813 with a threshold of 0.44787, Identity Attack is 0.03762 with a threshold of 0.57071, Insult is 0.07484 with a threshold of 0.70658, Profanity is 0.05251 with a threshold of 0.67056, and Threat is 0.02290 with a threshold of 0.52254. These scores provide quantitative insights into potentially harmful or inappropriate language in the analyzed content, facilitating the identification and mitigation of toxic elements to ensure a safer and more respectful online environment.

The toxicity scores presented in the analysis provide a quantitative assessment of the potentially harmful or inappropriate content within the first video. Each toxicity category, including Toxicity, Severe Toxicity,

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

Identity Attack, Insult, Profanity, and Threat, is assigned a score, along with corresponding thresholds. These scores help identify and categorize toxic behaviors or language types within the video content. For instance, higher scores in categories like Severe Toxicity, Insult, or Profanity may indicate a more significant presence of offensive or disrespectful language. In comparison, lower scores in categories like Threat may suggest a lower likelihood of explicit threats or harassment. Analyzing these scores assists in understanding the overall tone and nature of the content, enabling appropriate actions to address any issues and ensure a safer online environment.

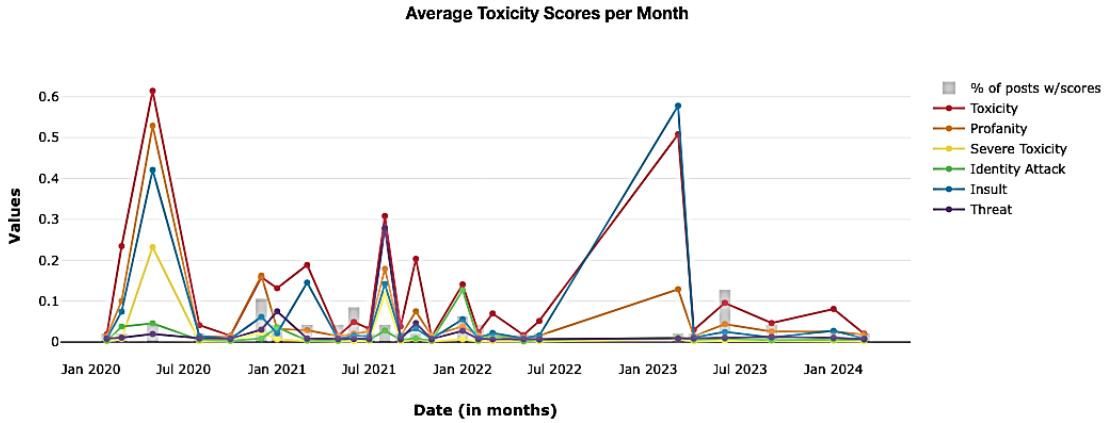

Figure 8. Toxicity Score of the Second Video (ZPG3XQhZVII)

Figure 8 shows the toxicity score of the first video (ZPG3XQhZVII). The analysis of toxicity scores from the second video reveals specific metrics indicating the presence of potentially harmful or inappropriate content. Each toxicity category, including Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat, is assigned a score along with respective thresholds. For instance, Toxicity is measured at 0.12794 with a threshold of 0.95638, Severe Toxicity at 0.01995 with a threshold of 0.45895, Identity Attack at 0.01868 with a threshold of 0.27992, Insult at 0.06490 with a threshold of 0.67254, Profanity at 0.07277 with a threshold of 0.95877, and Threat at 0.02783 with a threshold of 0.54744. These scores are quantitative measures to identify and categorize various toxic behaviors or language types within the video content. Analyzing these scores enables a deeper understanding of the overall tone and nature of the content, facilitating appropriate actions to address any issues and promote a safer online environment.

The provided toxicity score analysis presents numerical values representing different aspects of potentially harmful content within the second video. These metrics, including Toxicity, Severe Toxicity, Identity Attack, Insult, Profanity, and Threat, offer insights into the level and nature of toxic behavior or language in the video. The scores are accompanied by corresponding thresholds, indicating the point at which content may be considered toxic. For instance, a higher Toxicity score implies a greater likelihood of harmful content, while Severe Toxicity and Threat scores suggest more severe or threatening language or behavior. Identity Attack, Insult, and Profanity scores reflect specific types of toxic communication. By analyzing these scores, one assesses the overall toxicity of the video and takes appropriate measures to address any issues or concerns, such as content moderation or community guidelines enforcement.

3.2 Model Performance in Sentiment Classification

Based on the results of sentiment classification, differences in algorithm performance based on the confusion matrix, f-measure, and AUC were identified when utilizing the SMOTE operator compared to without using the SMOTE operator. The confusion matrix provides insights into classification accuracy by revealing the number of true positive, true negative, false positive, and false pessimistic predictions. Meanwhile, the f-measure considers precision and recall, offering a balanced classification performance assessment. Additionally, the AUC metric evaluates the ability of the model to distinguish between positive and negative sentiment classes. By comparing these performance metrics between models with and without SMOTE, the impact of oversampling on sentiment classification accuracy and effectiveness is comprehensively understood, facilitating informed decision-making in model selection and deployment strategies.

The provided metrics elucidate the performance of the SVM algorithm with SMOTE. The algorithm exhibited an accuracy of 59.88% +/- 7.45%, with a micro average of 59.91%. The confusion matrix revealed that out of the total positive instances, 173 were correctly classified, while 130 were misclassified as negative. Similarly, out of the total negative instances, 87 were correctly classified, with 44 misclassified as positive. The AUC metrics demonstrated values of 0.679 +/- 0.044 (optimistic), 0.643 +/- 0.051 (standard), and 0.607 +/- 0.060 (pessimistic), with the positive class being Negative. Precision was recorded at 65.09% +/- 8.51%, and recall at 40.09% +/- 13.95%, with a corresponding f-measure of 49.02% +/- 12.72%, all with the positive class being Negative. These metrics collectively portray the performance characteristics of the SVM algorithm with SMOTE, providing insights into its effectiveness and limitations in sentiment classification tasks.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

SVM Using SMOTE

```
PerformanceVector:
accuracy: 59.88% +/- 7.45% (micro average: 59.91%)
ConfusionMatrix:
True:   Positive        Negative
Positive:       173     130
Negative:       44      87
AUC (optimistic): 0.679 +/- 0.044 (micro average: 0.679) (positive class: Negative)
AUC: 0.643 +/- 0.051 (micro average: 0.643) (positive class: Negative)
AUC (pessimistic): 0.607 +/- 0.060 (micro average: 0.607) (positive class: Negative)
precision: 65.09% +/- 8.51% (micro average: 66.41%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       173     130
Negative:       44      87
recall: 40.09% +/- 13.95% (micro average: 40.09%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       173     130
Negative:       44      87
f_measure: 49.02% +/- 12.72% (micro average: 50.00%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       173     130
Negative:       44      87
```

DT Using SMOTE

```
PerformanceVector:
accuracy: 52.09% +/- 2.61% (micro average: 52.07%)
ConfusionMatrix:
True:   Positive        Negative
Positive:       15      6
Negative:       202     211
AUC (optimistic): 0.974 +/- 0.030 (micro average: 0.974) (positive class: Negative)
AUC: 0.521 +/- 0.023 (micro average: 0.521) (positive class: Negative)
AUC (pessimistic): 0.068 +/- 0.025 (micro average: 0.068) (positive class: Negative)
precision: 51.11% +/- 1.76% (micro average: 51.09%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       15      6
Negative:       202     211
recall: 97.27% +/- 3.18% (micro average: 97.24%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       15      6
Negative:       202     211
f_measure: 66.99% +/- 1.93% (micro average: 66.98%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       15      6
Negative:       202     211
```

k-NN Using SMOTE

```
PerformanceVector:
accuracy: 54.80% +/- 6.18% (micro average: 54.84%)
ConfusionMatrix:
True:   Positive        Negative
Positive:       74      53
Negative:       143     164
AUC (optimistic): 0.807 +/- 0.086 (micro average: 0.807) (positive class: Negative)
AUC: 0.589 +/- 0.072 (micro average: 0.589) (positive class: Negative)
AUC (pessimistic): 0.371 +/- 0.089 (micro average: 0.371) (positive class: Negative)
precision: 53.42% (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       74      53
Negative:       143     164
recall: 75.71% +/- 28.14% (micro average: 75.58%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       74      53
Negative:       143     164
f_measure: 62.60% (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       74      53
Negative:       143     164
```

NBC Using SMOTE

```
PerformanceVector:
accuracy: 59.25% +/- 5.74% (micro average: 59.22%)
ConfusionMatrix:
True:   Positive        Negative
Positive:       156     116
Negative:       61      101
AUC (optimistic): 0.740 +/- 0.073 (micro average: 0.740) (positive class: Negative)
AUC: 0.538 +/- 0.081 (micro average: 0.538) (positive class: Negative)
AUC (pessimistic): 0.469 +/- 0.092 (micro average: 0.469) (positive class: Negative)
precision: 62.54% +/- 8.35% (micro average: 62.35%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       156     116
Negative:       61      101
recall: 46.56% +/- 11.82% (micro average: 46.54%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       156     116
Negative:       61      101
f_measure: 52.72% +/- 9.27% (micro average: 53.30%) (positive class: Negative)
ConfusionMatrix:
True:   Positive        Negative
Positive:       156     116
Negative:       61      101
```

Figure 9. Performance of SVM, DT, k-NN, and NBC Using SMOTE

Figure 9 shows the performance of each algorithm using SMOTE. The analysis of the DT algorithm with SMOTE underscores its performance characteristics. The algorithm exhibited an accuracy of 52.09% +/- 2.61%, with a micro average of 52.07%. The confusion matrix illustrates that only 15 were correctly classified among the positive instances, while six were misclassified as unfavorable. Conversely, out of the total negative instances, 211 were correctly classified, with 202 misclassified as positive. Regarding AUC metrics, the optimistic, standard, and pessimistic values stood at 0.974 +/- 0.030, 0.521 +/- 0.023, and 0.068 +/- 0.025, respectively, with the positive class being Negative. The algorithm's precision was recorded at 51.11% +/- 1.76%, while recall and f-measure were observed at 97.27% +/- 3.18% and 66.99% +/- 1.93%, respectively, all with the positive class being Negative. These findings provide insights into the DT algorithm's performance with SMOTE, highlighting its strengths and weaknesses in sentiment classification tasks.

The performance evaluation of the NBC algorithm with SMOTE reveals notable insights. The algorithm achieved 59.25% +/- 5.74% accuracy, with a micro average of 59.22%. The confusion matrix depicts that out of the total positive instances, 156 were correctly classified, while 116 were misclassified as unfavorable. Similarly, out of the total negative instances, 101 were correctly classified, with 61 misclassified as positive. Regarding AUC metrics, the optimistic, standard, and pessimistic values were recorded at 0.740 +/- 0.073, 0.538 +/- 0.081, and 0.469 +/- 0.092, respectively, with the positive class being Negative. The precision of the algorithm stood at 62.54% +/- 8.35%, while recall and f-measure were observed at 46.56% +/- 11.82% and 52.72% +/- 9.27%, respectively, all with the positive class being Negative. These metrics provide insights into the performance characteristics of the NBC algorithm with SMOTE, delineating its effectiveness and limitations in sentiment classification tasks.

Analyzing the K-NN algorithm with SMOTE demonstrates its performance characteristics in sentiment classification tasks. The algorithm achieved an accuracy of 54.80% +/- 6.18%, with a micro average of 54.84%. Examination of the confusion matrix reveals that among the positive instances, 74 were correctly classified, while 53 were misclassified as unfavorable. Conversely, out of the total negative instances, 164 were correctly classified, with 143 misclassified as positive. Regarding AUC metrics, the optimistic, standard, and pessimistic values stood at 0.807 +/- 0.086, 0.589 +/- 0.072, and 0.371 +/- 0.089, respectively, with the positive class being Negative. The algorithm's precision for the positive class was observed at 53.42%. Furthermore, the algorithm exhibited a recall of 75.71% +/- 28.14% and an f-measure of 62.60% for the positive class. These findings provide insights into the performance of the K-NN algorithm with SMOTE, highlighting its strengths and areas for improvement in sentiment classification tasks.

Journal of Computer System and Informatics (JoSYC)
ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

The shortcomings of sentiment classification algorithms, including SVM, NBC, DT, and K-NN with or without SMOTE, vary but include several common challenges. Firstly, these algorithms may struggle to handle noisy or unstructured data, leading to inaccurate classifications. Secondly, they might exhibit limited generalization capabilities, especially when dealing with sentiment expressions outside the training data's scope. Additionally, these algorithms may be susceptible to overfitting, where they memorize the training data instead of learning underlying patterns, leading to poor performance on unseen data. Moreover, they might face challenges in capturing context and nuances in language, resulting in misinterpretations of sentiment. The computational complexity of some algorithms, mainly when applied to large datasets, hinders scalability and efficiency. Lastly, the effectiveness of these algorithms is influenced by imbalanced datasets, where one sentiment class significantly outweighs the other, leading to biased predictions. Addressing these limitations requires ongoing research and advancements in algorithm design, feature engineering, and data preprocessing techniques.

The limitations of this research encompass several aspects that warrant consideration. Firstly, the scope of the study may be confined to specific datasets, platforms, or contexts, which could restrict the generalizability of the findings to broader settings. Additionally, the methodology employed, such as the choice of algorithms, feature selection techniques, or evaluation metrics, may influence the results' robustness and comparability with other studies. Moreover, the availability and quality of data, including potential biases or inaccuracies, could impact the reliability and validity of the analyses. Furthermore, the study's duration and resources allocated might have constrained the depth or breadth of the investigation, limiting the exploration of alternative approaches or examining long-term trends. Additionally, external factors, such as technological advancements or changes in user behavior, may have occurred during the research period, which could affect the relevance and applicability of the findings. Lastly, ethical considerations, such as privacy concerns or the potential misuse of sentiment analysis outputs, should be acknowledged and addressed to ensure responsible research conduct and societal impact mitigation.

4. CONCLUSION

In conclusion, employing the CRISP-DM framework has facilitated this research's structured and systematic approach to sentiment analysis. Through meticulous evaluation, it has been revealed that each algorithm, including SVM, NBC, DT, and K-NN with SMOTE, exhibits distinct strengths and weaknesses in sentiment classification tasks. While SVM demonstrates an accuracy of 59.88% +/- 7.45%, NBC achieves 59.25% +/- 5.74%, DT scores 52.09% +/- 2.61%, and K-NN attains 54.80% +/- 6.18%. Despite the merits, these algorithms also exhibit limitations, such as varying precision, recall, and f-measure levels, underscoring the need for a comprehensive understanding of the performance metrics. Moreover, topics identified through content analysis, including Biometric security, Cloud storage, Emotion Analysis, Emotion Conveyance, Keywords for Emotions, Negative attitude or reception, Recurring Themes, Regulation, Specific Topics, Technology Security, and Topic Relationships, enrich the understanding of sentiment dynamics. Additionally, the toxicity score of the first video indicates Toxicity: 0.13227, Severe Toxicity: 0.00813, Identity Attack: 0.03762, Insult: 0.07484, Profanity: 0.05251, and Threat: 0.02290. Meanwhile, the toxicity score of the second video shows Toxicity: 0.12794, Severe Toxicity: 0.01995, Identity Attack: 0.01868, Insult: 0.06490, Profanity: 0.07277, and Threat: 0.02783. Therefore, this study underscores the importance of informed algorithm selection and evaluation methodologies within the CRISP-DM framework to optimize sentiment analysis outcomes while considering the diverse topics prevalent in content analysis.

REFERENCES

[1] A. N. Uwaechia and D. A. Ramli, "A Comprehensive Survey on ECG Signals as New Biometric Modality for Human Authentication: Recent Advances and Future Challenges," *IEEE Access*, vol. 9, pp. 97760–97802, 2021, doi: 10.1109/ACCESS.2021.3095248.

[2] R. Ryu, S. Yeom, S. H. Kim, and D. Herbert, "Continuous Multimodal Biometric Authentication Schemes: A Systematic Review," *IEEE Access*, vol. 9, pp. 34541–34557, 2021, doi: 10.1109/ACCESS.2021.3061589.

[3] A. Ali, M. Testa, L. Markhasin, T. Bianchi, and E. Magli, "Adversarial Learning of Mappings onto Regularized Spaces for Biometric Authentication," *IEEE Access*, vol. 8, pp. 149316–149331, 2020, doi: 10.1109/ACCESS.2020.3016599.

[4] M. Ingale, R. Cordeiro, S. Thentu, Y. Park, and N. Karimian, "ECG Biometric Authentication: A Comparative Analysis," *IEEE Access*, vol. 8, pp. 117853–117866, 2020, doi: 10.1109/ACCESS.2020.3004464.

[5] H. J. Mun and M. H. Lee, "Design for Visitor Authentication Based on Face Recognition Technology Using CCTV," *IEEE Access*, vol. 10, no. November, pp. 124604–124618, 2022, doi: 10.1109/ACCESS.2022.3223374.

[6] R. Arjona and I. Baturone, "A post-quantum biometric template protection scheme based on learning parity with noise (LPN) commitments," *IEEE Access*, vol. 8, pp. 182355–182365, 2020, doi: 10.1109/ACCESS.2020.3028703.

[7] A. A. Al-Saggaf, "A Post-Quantum Fuzzy Commitment Scheme for Biometric Template Protection: An Experimental Study," *IEEE Access*, vol. 9, pp. 110952–110961, 2021, doi: 10.1109/ACCESS.2021.3100981.

[8] X. Zhang, D. Cheng, P. Jia, Y. Dai, and X. Xu, "An Efficient Android-Based Multimodal Biometric Authentication System with Face and Voice," *IEEE Access*, vol. 8, pp. 102757–102772, 2020, doi: 10.1109/ACCESS.2020.2999115.

[9] H. Y. Kwon and M. K. Lee, "Comments on 'PassBio: Privacy-Preserving User-Centric Biometric Authentication,'"

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

IEEE Trans. Inf. Forensics Secur., vol. 17, pp. 2816–2817, 2022, doi: 10.1109/TIFS.2022.3195380.
[10] W. Yan, J. Tang, and S. Stucki, "Design and Implementation of a Lightweight Deep CNN-Based Plant Biometric Authentication System," *IEEE Access*, vol. 11, no. August, pp. 79984–79993, 2023, doi: 10.1109/ACCESS.2023.3296801.
[11] W. El-Shafai, F. A. H. E. Mohamed, H. M. A. Elkamchouchi, M. Abd-Elnaby, and A. Elshafee, "Efficient and Secure Cancelable Biometric Authentication Framework Based on Genetic Encryption Algorithm," *IEEE Access*, vol. 9, pp. 77675–77692, 2021, doi: 10.1109/ACCESS.2021.3082940.
[12] S. Hinatsu, D. Suzuki, H. Ishizuka, S. Ikeda, and O. Oshiro, "Evaluation of PPG Feature Values Toward Biometric Authentication Against Presentation Attacks," *IEEE Access*, vol. 10, pp. 41352–41361, 2022, doi: 10.1109/ACCESS.2022.3167667.
[13] L. Wu, L. Meng, S. Zhao, X. Wei, and H. Wang, "Privacy-Preserving Cancelable Biometric Authentication Based on RDM and ECC," *IEEE Access*, vol. 9, pp. 90989–91000, 2021, doi: 10.1109/ACCESS.2021.3092018.
[14] R. Ryu, S. Yeom, D. Herbert, and J. Dermoudy, "A Comprehensive Survey of Context-Aware Continuous Implicit Authentication in Online Learning Environments," *IEEE Access*, vol. 11, no. February, pp. 24561–24573, 2023, doi: 10.1109/ACCESS.2023.3253484.
[15] D. Palma, F. Blanchini, G. Giordano, and P. L. Montessoro, "A Dynamic Biometric Authentication Algorithm for Near-Infrared Palm Vascular Patterns," *IEEE Access*, vol. 8, pp. 118978–118988, 2020, doi: 10.1109/ACCESS.2020.3005460.
[16] Q. N. Tran, B. P. Turnbull, M. Wang, and J. Hu, "A Privacy-Preserving Biometric Authentication System With Binary Classification in a Zero Knowledge Proof Protocol," *IEEE Open J. Comput. Soc.*, vol. 3, no. January, pp. 1–10, 2021, doi: 10.1109/ojcs.2021.3138332.
[17] J. Zhao *et al.*, "A Secure Biometrics and PUFs-Based Authentication Scheme with Key Agreement for Multi-Server Environments," *IEEE Access*, vol. 8, pp. 45292–45303, 2020, doi: 10.1109/ACCESS.2020.2975615.
[18] K. Eledlebi, C. Y. Yeun, E. Damiani, and Y. Al-Hammadi, "Empirical Studies of TESLA Protocol: Properties, Implementations, and Replacement of Public Cryptography Using Biometric Authentication," *IEEE Access*, vol. 10, pp. 21941–21954, 2022, doi: 10.1109/ACCESS.2022.3152895.
[19] S. Vhaduri, S. V. Dibbo, and W. Cheung, "HIAuth: A Hierarchical Implicit Authentication System for IoT Wearables Using Multiple Biometrics," *IEEE Access*, vol. 9, pp. 116395–116406, 2021, doi: 10.1109/ACCESS.2021.3105481.
[20] B. Nakisa, F. Ansarizadeh, P. Oommen, and S. Shrestha, "Technology Acceptance Model: A Case Study of Palm Vein Authentication Technology," *IEEE Access*, vol. 10, no. November, pp. 120436–120449, 2022, doi: 10.1109/ACCESS.2022.3221413.
[21] P. Bauspieb *et al.*, "BRAKE: Biometric Resilient Authenticated Key Exchange," *IEEE Access*, vol. 12, no. January, pp. 46596–46615, 2024, doi: 10.1109/ACCESS.2024.3380915.
[22] A. Pradhan, J. He, H. Lee, and N. Jiang, "Multi-Day Analysis of Wrist Electromyogram-Based Biometrics for Authentication and Personal Identification," *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 5, no. 4, pp. 553–565, 2023, doi: 10.1109/TBIOM.2023.3299948.
[23] A. Rahman *et al.*, "Multimodal EEG and Keystroke Dynamics Based Biometric System Using Machine Learning Algorithms," *IEEE Access*, vol. 9, pp. 94625–94643, 2021, doi: 10.1109/ACCESS.2021.3092840.
[24] J. Coetzer, J. P. Swanepoel, and R. Sabourin, "Optimal human-machine collaboration for enhanced cost-sensitive biometric authentication," *SAIEE Africa Res. J.*, vol. 112, no. 2, pp. 110–119, 2021, doi: 10.23919/saiee.2021.9432899.
[25] V. Kumar, A. Mohammed Ali Al-Tameemi, A. Kumari, M. Ahmad, M. W. Falah, and A. A. Abd El-Latif, "PSEBVC: Provably Secure ECC and Biometric Based Authentication Framework Using Smartphone for Vehicular Cloud Environment," *IEEE Access*, vol. 10, no. July, pp. 84776–84789, 2022, doi: 10.1109/ACCESS.2022.3195807.
[26] G. Li and H. Sato, "Sensing In-Air Signature Motions Using Smartwatch: A High-Precision Approach of Behavioral Authentication," *IEEE Access*, vol. 10, pp. 57865–57879, 2022, doi: 10.1109/ACCESS.2022.3177905.
[27] R. Zhang, Z. Yan, X. Wang, and R. H. Deng, "VOLERE: Leakage Resilient User Authentication Based on Personal Voice Challenges," *IEEE Trans. Dependable Secur. Comput.*, vol. 20, no. 2, pp. 1002–1016, 2023, doi: 10.1109/TDSC.2022.3147504.
[28] A. Sedik *et al.*, "Deep learning modalities for biometric alteration detection in 5g networks-based secure smart cities," *IEEE Access*, vol. 9, pp. 94780–94788, 2021, doi: 10.1109/ACCESS.2021.3088341.
[29] M. Mwapasa *et al.*, "'Are we getting the biometric bioethics right?'–the use of biometrics within the healthcare system in Malawi," *Glob. Bioeth.*, vol. 31, no. 1, pp. 67–80, 2020, doi: 10.1080/11287462.2020.1773063.
[30] A. Thiel, "Biometric payment and gendered kinds in Ghana," *Tapuya Lat. Am. Sci. Technol. Soc.*, vol. 4, no. 1, 2021, doi: 10.1080/25729861.2021.1924486.
[31] I. Z. P. Hamdan and M. Othman, "Predicting Customer Loyalty Using Machine Learning for Hotel Industry," *J. Soft Comput. Data Min.*, vol. 3, no. 2, pp. 31–42, 2022.
[32] I. Maskanah, A. Primajaya, and A. Rizal, "Segmentasi Pelanggan Toko Purnama dengan Algoritma K-Means dan Model RFM untuk Perancangan Strategi Pemasaran," *J. INOVTEK Polbeng - Seri Inform.*, vol. 5, no. 2, pp. 218–228, 2020, doi: 10.35314/isi.v5i2.1443.
[33] C. A. Fidas and D. Lyras, "A Review of EEG-Based User Authentication: Trends and Future Research Directions," *IEEE Access*, vol. 11, no. February, pp. 22917–22934, 2023, doi: 10.1109/ACCESS.2023.3253026.
[34] M. Suorsa and P. Helo, "Information security failures identified and measured–ISO/IEC 27001:2013 controls ranked based on GDPR penalty case analysis," *Inf. Secur. J.*, vol. 33, no. 3, pp. 285–306, 2024, doi: 10.1080/19393555.2023.2270984.
[35] L. Laishram, J. T. Lee, and S. K. Jung, "Face De-Identification Using Face Caricature," *IEEE Access*, vol. 12, no. November 2023, pp. 19344–19354, 2024, doi: 10.1109/ACCESS.2024.3356550.
[36] M. Montenegro de Wit and M. Canfield, "'Feeding the world, byte by byte': emergent imaginaries of data productivism," *J. Peasant Stud.*, vol. 51, no. 2, pp. 381–420, 2024, doi: 10.1080/03066150.2023.2232997.
[37] J. Wei, "Video face recognition of virtual currency trading system based on deep learning algorithms," *IEEE Access*,

ISSN 2714-8912 (media online), ISSN 2714-7150 (media cetak)
Volume 5, No. 3, May 2024, Page 591-604
https://ejurnal.seminar-id.com/index.php/josyc
DOI 10.47065/josyc.v5i3.5167

vol. 9, pp. 32760–32773, 2021, doi: 10.1109/ACCESS.2021.3060458.
[38] A. Wibowo, W. Alawiyah, and Azriadi, "The importance of personal data protection in Indonesia's economic development," *Cogent Soc. Sci.*, vol. 10, no. 1, p., 2024, doi: 10.1080/23311886.2024.2306751.
[39] A. Magunna, "Charting waters: the private sector's evolving governance role in Southeast Asian maritime security," *Aust. J. Int. Aff.*, pp. 1–20, 2024, doi: 10.1080/10357718.2024.2337013.
[40] R. Wevers, "Denormalising surveillance through curation in Face Value: Surveillance and Identity in the Age of Digital Face Recognition," *Media Pract. Educ.*, vol. 24, no. 2, pp. 182–198, 2023, doi: 10.1080/25741136.2023.2210425.
[41] L. S. Luevano, L. Chang, H. Heydi Mendez-Vazquez, Y. Martinez-Diaz, and M. Gonzalez-Mendoza, "A Study on the Performance of Unconstrained Very Low Resolution Face Recognition: Analyzing Current Trends and New Research Directions," *IEEE Access*, vol. 9, pp. 75470–75493, 2021, doi: 10.1109/ACCESS.2021.3080712.
[42] P. C. P. Neto, J. R. Pinto, F. Boutros, N. Damer, A. F. Sequeira, and J. S. Cardoso, "Beyond Masks: On the Generalization of Masked Face Recognition Models to Occluded Face Recognition," *IEEE Access*, vol. 10, no. July, pp. 86222–86233, 2022, doi: 10.1109/ACCESS.2022.3199014.
[43] N. Li *et al.*, "Chinese Face Dataset for Face Recognition in an Uncontrolled Classroom Environment," *IEEE Access*, vol. 11, no. August, pp. 86963–86976, 2023, doi: 10.1109/ACCESS.2023.3302919.
[44] H. O. Shahreza and S. Marcel, "Comprehensive Vulnerability Evaluation of Face Recognition Systems to Template Inversion Attacks via 3D Face Reconstruction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 12, pp. 14248–14265, 2023, doi: 10.1109/TPAMI.2023.3312123.
[45] L. I. U. Jinjin, L. I. Qingbao, M. Liu, and T. Wei, "CP-GAN: A cross-pose profile face frontalization boosting pose-invariant face recognition," *IEEE Access*, vol. 8, pp. 198659–198667, 2020, doi: 10.1109/ACCESS.2020.3033675.
[46] H. Yang and X. Han, "Face recognition attendance system based on real-time video processing," *IEEE Access*, vol. 8, pp. 159143–159150, 2020, doi: 10.1109/ACCESS.2020.3007205.
[47] M. Zhang, R. Liu, D. Deguchi, and H. Murase, "Masked Face Recognition With Mask Transfer and Self-Attention Under the COVID-19 Pandemic," *IEEE Access*, vol. 10, pp. 20527–20538, 2022, doi: 10.1109/ACCESS.2022.3150345.
[48] H. H. Nguyen, S. Marcel, J. Yamagishi, and I. Echizen, "Master Face Attacks on Face Recognition Systems," *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 4, no. 3, pp. 398–411, 2022, doi: 10.1109/TBIOM.2022.3166206.
[49] P. Terhorst, M. Huber, N. Damer, F. Kirchbuchner, K. Raja, and A. Kuijper, "Pixel-Level Face Image Quality Assessment for Explainable Face Recognition," *IEEE Trans. Biometrics, Behav. Identity Sci.*, vol. 5, no. 2, pp. 288–297, 2023, doi: 10.1109/TBIOM.2023.3263186.
[50] Z. Huang, J. Zhang, and H. Shan, "When Age-Invariant Face Recognition Meets Face Age Synthesis: A Multi-Task Learning Framework and a New Benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 45, no. 6, pp. 7917–7932, 2023, doi: 10.1109/TPAMI.2022.3217882.