Exploring Sentiments and Topics in News Coverage of Police and Law Enforcement

Hsin Hua Sung

Code Repository - https://osf.io/9taj4/?view_only=c9116fc66f8643198c87457fa2631747

Abstract

This study employs text mining and machine learning to analyse news coverage of police and law enforcement in the UK, aiming to explore sentiment, trending topics, and their potential correlations. Data was collected from The Guardian, followed by sentiment analysis and topic modelling. Support Vector Machines (SVM), Naïve Bayes (NB), and Neural Networks (NN) were used to train the model for sentiment prediction. Latent Dirichlet Allocation (LDA) and Structural Topic Model (STM) were adopted to identify prominent topics. The findings reveal a prevalence of negative sentiment in the news coverage. The NN model exhibited the highest accuracy in sentiment prediction. Positive sentiment was observed for proactive police action, while negative sentiment emerged in relation to frustrations with the criminal justice system. This study assists law enforcement agencies in better understanding public concerns and enhancing community-police relations.

1 Introduction

The media plays a crucial role in shaping public opinion and attitudes towards law enforcement. News coverage serves as a primary source of information about the criminal justice system, influencing individuals' perceptions, trust, and willingness to cooperate with the police (Gross and Aday, 2003; Intravia et al., 2018).

Negative portrayals of police misconduct in the media significantly impact public attitudes and erode trust (Graziano, 2018). Studies have shown that increased exposure to negative media coverage related to the police is associated with more negative attitudes towards law enforcement (Wu, 2010; Wu et al., 2011; Sun et al., 2016). This suggests that addressing negative depictions in the media could potentially contribute to improving public perception and trust in law enforcement. When individuals perceive police actions as lawful, they are more inclined to cooperate with law enforcement (Dowler, 2002; Tyler, 2004).

The media's depiction of the police has generated conflicting findings in previous research (Dowler, 2002; Graziano, 2018). Some suggest a positive portrayal of the police in the media (Eschholz et al., 2002), while others view it as negative (Doyle, 2006; Chermak et al., 2006). Weitzer and Tuch (2004) conducted surveys to investigate the public perception of negative news coverage regarding police misconduct by asking respondents about their frequency of encountering such news in the media. This approach has been widely adopted in subsequent studies; however, it heavily depends on respondents' memory and is susceptible to biases.

In light of previous controversies and limitations, this study analyses news coverage of police and law enforcement in the UK, aiming to address the following research questions:

- What was the general sentiment in news coverage of police and law enforcement, and how did it evolve over time?
- What were the predominant topics in news coverage of police and law enforcement, and how did they change over time?
- Was there a correlation between fluctuations in sentiment and the topics discussed in news coverage of police and law enforcement?

To investigate these questions, The Guardian was selected as the data source due to its reputation, wide readership, and representation of British news. Sentiment analysis and topic modelling techniques were applied to analyse news coverage of police and law enforcement. This study can assist policymakers in gaining a deeper understanding of public concerns, enabling them to foster a positive image of the police and enhance community-police relations.

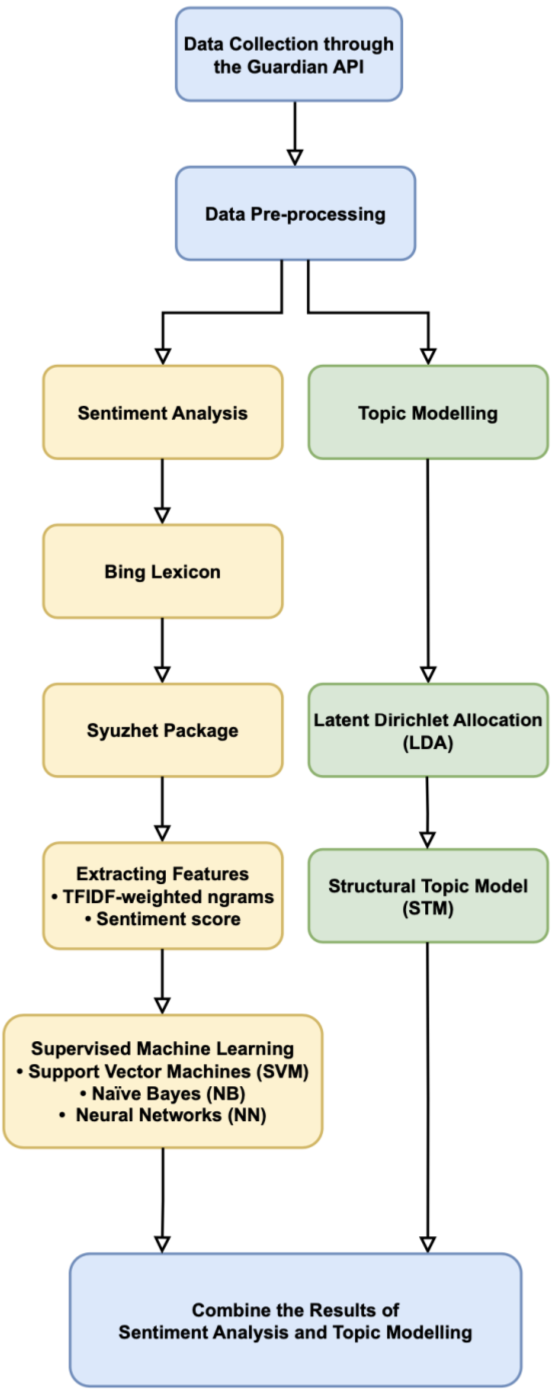

Figure 1: This study's methodology and workflow.

2 Methods

The study's methodology consists of four primary stages: data collection, data pre-processing, sentiment analysis, and topic modelling, as shown in Figure 1.

2.1 Data Collection

Firstly, The Guardian's Application Programming Interface (API) was utilised to communicate with its server and gather news articles related to 'police' and 'law enforcement' from January 2021 to March 2023. This initial search yielded 24,863 articles with various attributes including title, body text, publication date, section, word count, URL, etc. To narrow our focus on the UK region, the articles were filtered by section, resulting in 3,259 news items related to British police and law enforcement.

2.2 Data Pre-processing

Pre-processing plays a vital role in ensuring data quality and preparing it for analysis. It involves balancing the reduction of analysis complexity while preserving relevant content. To ensure the selected articles contain the necessary components for meaningful analysis, we specifically included news articles with titles, body text, and publication dates, while eliminating any duplicate content and noise. Tokenisation was applied to enhance the data's readiness for analysis. Punctuation marks, numbers, symbols, and English stop words were removed to eliminate distractions. Additionally, all text was converted to lowercase and stemmed to ensure consistency across the dataset. After pre-processing, 3,194 cleaned UK news articles were obtained as the basis for subsequent analysis.

2.3 Sentiment Analysis

2.3.1 Lexicon-based Analysis

To explore the sentiment in news articles, the Bing lexicon and the syuzhet package were used to compute a sentiment score for each article. The Bing lexicon is a pre-existing sentiment dictionary that contains a large collection of words and phrases annotated with their associated sentiment scores. The algorithm counts the number of positive and negative words in each article's text and calculates a score based on the difference between the two. Articles with a sentiment score less than zero were classified as negative, those with a score of zero were considered neutral, and articles with a score greater than zero were categorised as positive.

To capture a more nuanced understanding of sentiment, the syuzhet package was employed. The syuzhet package is a popular R package used for emotion detection in textual data (Misuraca et al., 2020). It is designed to uncover and visualise the emotional patterns and shifts present in texts. This study defined specific cutoff points to categorise articles based on their sentiment scores. Articles with sentiment scores lower than or equal to -0.1 were classified as negative, sentiment scores ranging from -0.1 to 0.1 were considered neutral, and sentiment scores greater than 0.1 were classified as positive. The initial analysis allowed for classifying and labelling the sentiment of each article as positive, negative, or neutral, capturing the sentiment distribution of news articles, and further exploring how sentiment changed over time.

2.3.2 Supervised Machine Learning

After labelling and classifying the data, we employed supervised machine learning techniques to construct a sentiment prediction model by extracting relevant features. Our model utilised two primary features: TFIDF-weighted n-grams and sentiment scores.

TFIDF-weighted n-grams are a well-established feature representation in natural language processing (NLP) and text mining. This approach involves analysing the frequency of n-grams (sequences of n words) within a document and assigning weights based on their significance in distinguishing that document from others in the corpus (Nasser et al., 2021). The weights are calculated using the TFIDF (Term Frequency-Inverse Document Frequency) formula, which considers both the term's frequency in a document (TF) and its importance relative to the entire corpus (IDF) (Elberrichi and Aljohar, 2007; Hassan et al., 2020). By incorporating TFIDF-weighted n-grams as features, the model could capture the distinctive word patterns that contribute to sentiment analysis.

Sentiment scores were incorporated as additional features to provide insights into the overall sentiment of an article. These scores reflect the polarity or intensity of sentiment associated with specific words or phrases (Sarkar, 2018). By integrating sentiment scores as features, the model learned to identify meaningful connections between specific linguistic expressions and sentiment, enhancing its ability to predict sentiment accurately.

After feature engineering, the news article was split into 80% training and 20% testing datasets. Support Vector Machines (SVM), Naïve Bayes (NB), and Neural Networks (NN) were employed to train the model.

SVM is a machine learning algorithm that separates data points into different classes using a hyperplane, which is effective for high-dimensional data with a clear separation between classes. NB is another popular algorithm for classification tasks that leverages Bayes' theorem to predict the probability of a data point belonging to a particular class. It assumes that the features are independent, making it computationally efficient and effective for text classification tasks. NN are a class of machine learning algorithms that simulate the structure and function of the human brain to identify patterns in data. They can handle complex nonlinear relationships between features and are effective for a wide range of classification tasks (Mohammad et al., 2016; Shahi and Pant, 2018).

After model training, confusion matrices were used to compare and assess the performance of these algorithms on the testing dataset.

2.4 Topic Modelling

Latent Dirichlet Allocation (LDA) and Structural Topic Model (STM) were utilised to identify the underlying topics and observe how they changed over time. LDA is a probabilistic model that assumes each document is a mixture of topics, with each topic represented as a probability distribution over words (Blei et al., 2003). It is a widely used and well-established technique that can be effective for NLP and text mining applications. STM is an extension of LDA that incorporates document-level covariates to model the relationships between topics and these covariates (He et al., 2020). This allows for more flexible modelling and can lead to improved topic coherence and interpretability.

Finally, the results of topic modelling were combined with sentiment analysis to explore potential correlations between the identified topics and sentiment fluctuations.

3 Results

3.1 Sentiment Analysis Results

Among the 3,194 UK news articles considered, the Bing lexicon was utilised to assign a sentiment score to each article. The resulting scores range from -214 to 301 (Mean = -16.14). Subsequently, the syuzhet package was employed to calculate sentiment scores for each article, yielding scores ranging from -62.30 to 162.45 (Mean = -3.79). Figure 2 illustrates the distribution of sentiment categories observed in the analysed news coverage. It is worth noting that the syuzhet package captured a greater number of positive emotions for semantic context in the database.

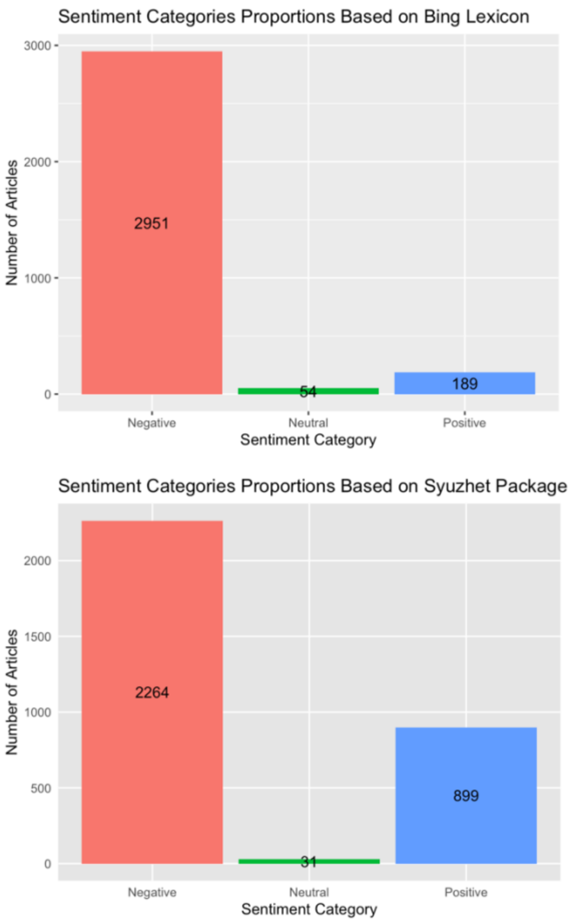

Figure 2: Sentiment categories calculated by Bing lexicon and syuzhet package.

In Figure 3, we observe the temporal variations in sentiment scores. When utilising the Bing lexicon, the majority of sentiment scores were negative, except for a peak of positive sentiment in September 2022. In comparison, sentiment scores derived from the syuzhet package display peaks of positive sentiment in April 2021 and September 2022, while fluctuating between positive and negative during other time periods.

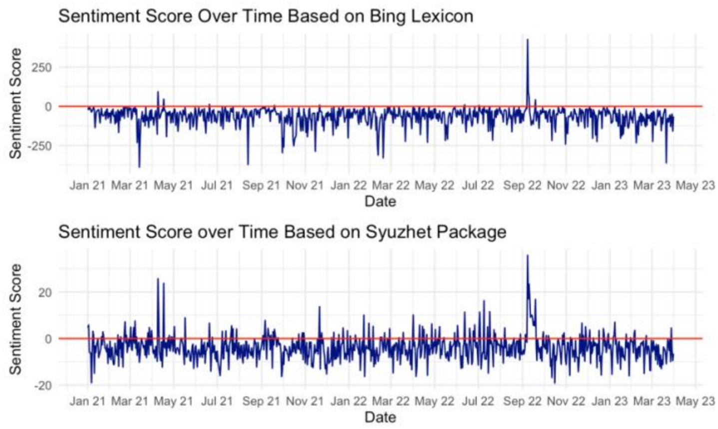

Figure 3: Temporal variations in sentiment scores.

3.2 Machine Learning Models

After sentiment classification and labelling, the article features were extracted to construct a machine learning model capable of predicting sentiment classes as either negative, neutral, or positive. To assess the effectiveness of the trained models using SVM, NB, and NN algorithms, a confusion matrix was employed.

The confusion matrix presents the number of true positives, true negatives, false positives, and false negatives for each class, providing multiple metrics to evaluate the model's performance. Precision, recall, and F1 score are the primary statistics utilised to measure the model's effectiveness. Precision measures the proportion of true positives among instances predicted as positive, while recall measures the proportion of true positives that the model correctly predicts. The F1 score combines precision and recall by calculating their harmonic mean, giving a comprehensive measure of the model's overall performance (Goutte and Gaussier, 2005).

The performance of the three machine learning models is presented in Table 1, revealing the following findings:

- The SVM algorithm achieved an accuracy of 0.8917 and demonstrated high accuracy for both negative and positive classes. However, it had a recall of zero for the neutral category, indicating that the model was unable to recognise instances belonging to this class.

- The NB algorithm exhibited the lowest precision among the three models, with a value of 0.7080. This value was slightly lower than the uninformative rate of 0.7096, suggesting that the model's performance was only slightly better than random guessing. Additionally, the kappa value was low (0.0169), indicating poor agreement between the predicted and actual categories.

- The NN algorithm outperformed the other models, achieving an overall accuracy of 0.9796. It performed well for both negative and positive classes, with precision, recall, and F1 scores above 0.94. However, the model did not correctly predict any samples as neutral. Overall, the NN model demonstrated high accuracy and effectiveness, but more data is required to better handle the neutral class.

Algorithm	Accuracy	Precision (Negative) (Neutral) (Positive)	Recall (Negative) (Neutral) (Positive)	F1 (Negative) (Neutral) (Positive)
SVM	0.8917	0.9017 N/A 0.8616	0.9535 0.0000 0.7654	0.9269 N/A 0.8107
NB	0.7080	0.7147 0.0000 N/A	0.9978 0.0000 0.0000	0.8329 N/A N/A
NN	0.9796	0.9933 N/A 0.9465	0.9889 0.0000 0.9888	0.9911 N/A 0.9672

Table 1: Performance evaluation metrics for the machine learning models.

3.3 Topic Modelling Results

To identify potential topics in news related to police and law enforcement, LDA and STM were employed for topic modelling. By analysing the keywords associated with each topic, the categories are inferred as shown in Table 2.

Method	Key words	Topic Categories
LDA	"incidents", "die", "fire"	Topic 1: Incidents and Emergencies
	"govern", "protest", "right", "work"	Topic 2: Protests and Governance
	"investig", "met", "forc", "report", "case", " alleg"	Topic 3: Investigations
	"royal", "minister", "state", "service"	Topic 4: Royalty, Government and Countries
	"court", "murder", "kill", "charge", "sentence"	Topic 5: Legal Issues and Trials
STM	"court", "alleg", "case", "charg", "evid", "victim", "sexual"	Topic 1: Legal Issues and Trials
	"incid", "fire", "arrest", "hospit"	Topic 2: Incidents and Emergencies
	"protest", "govern", "crime", "women"	Topic 3: Protests and Governance
	"family", "friend", "love", "murder", "kill"	Topic 4: Family and Social Issues
	"royal", "state", "minist", "countri"	Topic 5: Royalty, Government and Countries

Table 2: Topic categories inferred from identified keywords.

Figure 4 displays the distribution of topic categories in the analysed news coverage. The results of LDA modelling indicate that the most widely reported topic was Topic 1 (Incidents and Emergencies), followed by Topic 2 (Protests and Governance) and Topic 5 (Legal Issues and Trials). In comparison, the results of STM reveal that Topic 3 (Protests and Governance) had the largest proportion of reports, followed by Topic 1 (Legal Issues and Trials).

Figure 4: Distribution of topic categories in the coverage.

After identifying the potential topics, a temporal analysis was conducted to examine their coverage over time, as shown in Figure 5. According to LDA modelling, Topic 2 (Protests and Governance) was the most reported topic in early 2021. Subsequently, the focus shifted to topics related to Topic 1 (Incidents and Emergencies). In early 2022, Topic 3 (Investigations) and Topic 5 (Legal Issues and Trials) received increased coverage. Topic 4 (Royalty, Government, and Countries) reached its peak in the second half of 2022.

By contrast, the changes in topics mined from STM indicate that Topic 3 (Protests and Governance) remained the primary reporting topic throughout the period. The coverage of Topic 1 (Legal Issues and Trials), Topic 2 (Incidents and Emergencies), and Topic 3 (Protests and Governance) remained relatively stable over time, with no significant variations. Topic 5 (Royalty, Government, and Countries) was reported less frequently, but it experienced a peak in the second half of 2022.

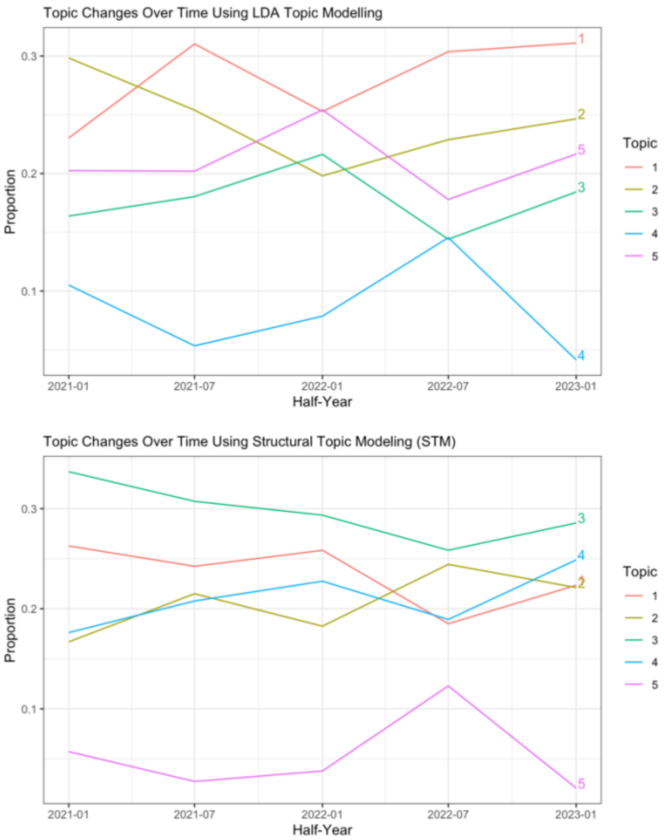

Figure 5: Temporal variations of topic categories.

3.4 Sentiment-Topic Connections

In the final stage, we aim to investigate potential relationships between sentiment and topics, providing explanations for fluctuations in sentiment. Based on the preliminary sentiment analysis conducted using the syuzhet package, notable peaks of positive sentiment were observed in April 2021 and September 2022. Further examination reveals that these peaks were primarily driven by news related to the British royal family.

To narrow our focus on news articles more closely related to the police, we excluded keywords associated with the royal family. As a result, we discovered that news topics involving active police actions in investigating cases, such as confronting and arresting suspects or urgently searching for missing persons, generally evoked positive sentiment. Additionally, news topics discussing the equipment utilised by the police in law enforcement, such as stun guns, also elicited positive sentiment. Conversely, news topics with extremely low sentiment scores were associated with negative events, including assault, murder, stalking, abuse, violence, and potential failures of the police and criminal justice system in bringing perpetrators to justice. These types of topics were found to be associated with negative emotions.

4 Discussion

4.1 Sentiment Analysis

The results of the sentiment analysis reveal the distribution of sentiment and the temporal changes in the coverage of police and law enforcement. The use of the syuzhet package, which incorporates semantic context, enhanced the identification of positive sentiments within the dataset. By incorporating semantic context, the sentiment analysis approach offered a more nuanced understanding of sentiment in news articles.

Sentiment classification in machine learning models demonstrates varying levels of performance. The SVM algorithm achieved high accuracy in identifying negative and positive classes but encountered difficulties in recognising neutral classes. The NB algorithm exhibited the lowest accuracy among all the models, slightly better than random guessing.

On the other hand, the NN algorithm demonstrated high accuracy in classifying negative and positive classes, but it failed to correctly predict any samples as neutral. These findings emphasise the potential of machine learning models, especially NN algorithms, in effectively classifying sentiment in news articles. However, a larger dataset containing neutral class data may be necessary to improve performance in the neutral class.

4.2 Topic Modelling

Employing LDA and STM techniques allows for the identification of underlying topics within news articles. The results from LDA modelling indicate that the most widely reported topics were 'Incidents and Emergencies,' 'Protests and Governance,' and 'Legal Issues and Trials.' Similarly, STM analysis reveals that 'Protests and Governance' and 'Legal Issues and Trials' were the most proportionately reported topics.

Understanding these topics provide insights for law enforcement agencies, enabling them to make informed decisions and take effective action in response to public concerns and priorities. Further temporal analysis of the topics reveals their dynamic nature in news coverage related to policing. Recognising the patterns of topic coverage changes allow policymakers to adapt their strategies to effectively address evolving public interests and emerging issues.

4.3 Sentiment-Topic Connections

Lastly, this study investigates the relationship between sentiment and police-related topics. The findings indicate that news articles highlighting proactive policing actions and discussions on police equipment generally evoked positive sentiment. Conversely, articles associated with negative events and failures of the criminal justice system elicited negative sentiment.

These findings help law enforcement agencies understand public perceptions and identify areas for improvement to enhance the overall image of the police. Specifically, emphasising proactive police actions in news coverage could contribute to fostering positive public sentiment.

5 Limitations

5.1 Data-related Limitations

While the Guardian is a representative news media outlet in the UK, relying solely on this single data source may introduce the possibility of reporting bias. It is essential to consider including additional data sources to mitigate this limitation.

5.2 Methodological Limitations

Analysing sentiment using pre-existing sentiment dictionaries and automated algorithms may not fully capture the complexity of human subjectivity and interpretation. Incorporating human annotation or expert review can provide more nuanced insights and enhance the accuracy of sentiment analysis. Additionally, it is crucial to exercise caution when interpreting negative sentiment in crime-related news, as it should not be automatically attributed to negative sentiment towards the police in general.

The machine learning models used for sentiment classification were evaluated on a limited dataset, emphasising the need for validation on larger and more diverse datasets to ensure generalisability and reliability. The imbalanced distribution of sentiment categories, particularly the scarcity of neutral samples, may impact the model's performance and accuracy.

The interpretation of topic modelling results heavily relies on the keywords associated with each topic, which may introduce the possibility of oversight or misclassification of important subtopics. Furthermore, the selection of the number of topics extracted from the data can significantly influence the results, and different researchers may choose different numbers of topics, leading to variations in research outcomes.

6 Conclusions

In this study, the sentiment and topics covered in news articles related to British police and law enforcement were examined. The findings reveal that a majority of these articles express negative sentiment, and the choice of sentiment analysis tool plays a crucial role in capturing emotions in context. Notably, neural network algorithms showed high accuracy in classifying sentiment.

Moreover, key topics within the articles were identified, shedding light on the public's interests in police-related matters. The study also explores the relationship between sentiment and topics, revealing positive emotions associated with proactive police actions and discussions on police equipment, while negative emotions arose from adverse incidents and shortcomings in the criminal justice system.

Overall, this study provides valuable information about sentiment and topical interests surrounding policing and law enforcement. Law enforcement agencies and policymakers can utilise these findings to better understand public concerns and expectations, making informed decisions to foster public trust and confidence in law enforcement.

Moving forward, future research could address the mentioned limitations and consider incorporating a wider range of sources, including other news media platforms and social media content, to improve the reliability and generalisability of the study.

References

Aaron Doyle. 2006. *How Not to Think about Crime in the Media.* Canadian Journal of Criminology and Criminal Justice, 48(6):867–885.

Adel H. Mohammad, Tariq Alwada‘n, and Omar Al-Momani. 2016. *Arabic Text Categorization Using Support vector machine, Naïve Bayes and Neural Network.* GSTF Journal on Computing (JoC), 5(1):16.

Cyril Goutte and Eric Gaussier. 2005. *A Probabilistic Interpretation of Precision, Recall and F-Score, with Implication for Evaluation*. In David E. Losada and Juan M. Fernández-Luna, editors, Advances in Information Retrieval, pages 345–359, Berlin, Heidelberg. Springer.

David M. Blei, Andrew Y. Ng, and Michael I. Jordan. 2003. *Latent dirichlet allocation.* Journal of machine Learning research, 3:993-1022.

Ivan Y. Sun, Yuning Wu, Ruth Triplett, and Shun-Yung K. Wang. 2016. *The impact of media exposure and political party orientation on public perceptions of police in Taiwan.* Policing: An International Journal of Police Strategies & Management, 39(4):694–709.

Jonathan Intravia, Kevin T. Wolff, and Alex R. Piquero. 2018. *Investigating the Effects of Media Consumption on Attitudes Toward Police Legitimacy.* Deviant Behavior, 39(8):963–980.

Kenneth Dowler. 2002. *Media Influence on Citizen Attitudes Toward Police Effectiveness.* Policing and Society, 12(3):227–238.

Kimberly Gross and Sean Aday. 2003. *The Scary World in Your Living Room and Neighborhood: Using Local Broadcast News, Neighborhood Crime Rates, and Personal Experience to Test Agenda Setting and Cultivation.* Journal of Communication, 53(3):411-426.

Lifeng He, Dongmei Han, Xiaohang Zhou, and Zheng Qu. 2020. *The Voice of Drug Consumers: Online Textual Review Analysis Using Structural Topic Model.* International Journal of Environmental Research and Public Health, 17(10):3648.

Lisa M. Graziano. 2018. *News media and perceptions of police: a state-of-the-art-review.* Policing: An International Journal, 42(2):209–225.

Michelangelo Misuraca, Alessia Forciniti, Germana Scepi, and Maria Spano. 2020. *Sentiment Analysis for Education with R: packages, methods and practical applications.*

Nidal Nasser, Lutful Karim, Ahmed E. Ouadrhiri, Asmaa Ali, and Nargis Khan. 2021. *n-Gram based language processing using Twitter dataset to identify COVID-19 patients.* Sustainable Cities and Society, 72:103048.

Noha Y. Hassan, Wael H. Gomaa, Ghada A. Khoriba, and Mohammed H. Haggag. 2020. *Credibility Detection in Twitter Using Word N-gram Analysis and Supervised Machine Learning Techniques.* International Journal of Intelligent Engineering and Systems, 13(1):291–300.

Ronald Weitzer and Steven A. Tuch. 2004. *Race and Perceptions of Police Misconduct.* Social Problems, 51(3):305–325.

Sarah Eschholz, Brenda S. Blackwell, Marc Gertz, and Ted Chiricos. 2002. *Race and attitudes toward the police: Assessing the effects of watching "reality" police programs.* Journal of Criminal Justice, 30(4):327–341.

Steven Chermak, Edmund McGarrell, and Jeff Gruenewald. 2006. *Media coverage of police misconduct and attitudes toward police.* Policing: An International Journal of Police Strategies & Management, 29(2):261–281.

Tej B. Shahi and Ashok K. Pant. 2018. *Nepali news classification using Naïve Bayes, Support Vector*

Machines and Neural Networks. In 2018 International Conference on Communication information and Computing Technology (ICCICT), pages 1–5.

Tom R. Tyler. 2004. *Enhancing police legitimacy.* The annals of the American academy of political and social science 593(1): 84-99.

Yuning Wu, Ivan Y. Sun, and Brad W. Smith. 2011. *Race, Immigration, and Policing: Chinese Immigrants' Satisfaction with Police*. Justice Quarterly, 28(5):745–774.

Yuning Wu. 2010. *'College students' evaluation of police performance: A comparison of Chinese and Americans.* Journal of Criminal Justice 38(4): 773-780.

Zakaria Elberrichi and Badr Aljohar. 2007. *N-grams in Texts Categorization.* Scientific Journal of King Faisal University (Basic and Applied Sciences), 8(2):25-39.