Check for
updates

Social Sensing for Sentiment Analysis of Policing Authority Performance in Smart Cities

Tariq Malik[1,2], Ahsen Tahir[2,3]*, Ahsan Bilal[1], Kia Dashtipour[4], Muhammad Ali Imran[2] and Qammer H. Abbasi[2]

[1]Punjab Safe Cities Authority, Lahore, Pakistan, [2]James Watt School of Engineering, University of Glasgow, Glasgow, United Kingdom, [3]Department of Electrical Engineering, University of Engineering and Technology, Lahore, Pakistan, [4]School of Computing, Edinburgh Napier University, Edinburgh, United Kingdom

High-tech services in smart cities, ubiquity of smart phones, and proliferation of social media platforms have enabled social sensing, either through direct human observers or through humans as sensor carriers and operators, such as through the use of smart phones, cameras, etc. We performed a sentiment analysis (SA) and mined public opinion on the civil services and policing authority in a smart city. The establishment of high-tech policing in Lahore, Pakistan, known as the Punjab Safe Cities Authority (PSCA), Lahore, along with integrated command and control centers and various equipments, such as 8,000 cameras, monitoring sensors, etc., has resulted in a requirement for its performance evaluation and social media–enabled opinion mining to determine the broader impact on communities. Social sensing of civil services has been enabled through the presence of the PSCA on Facebook, Twitter, YouTube, and Web TV. The SA of the local civil services is not possible without taking into account the local language. In this article, we utilize machine learning techniques to perform multi-class SA of public opinion on policing authority and the provided civil services in both the local languages Urdu and English. The support vector machine provides the highest performance multi-classification accuracy of 86.87% for positive, negative, and neutral sentiments. The temporal sentiments are determined over time from January 2020 to July 2021, with an overall positive sentiment of 62.40% and a negative sentiment of 13.51%, which shows high satisfaction of policing authority and the provided civil services.

Keywords: sentiment analysis, machine learning, natural language processing, artificial intelligence, social sensing

1 INTRODUCTION

The internet provides a large repository of structured and unstructured data. The analysis of data to extract potential and relevant information is a challenge. In the current era, social media platforms like Twitter, Facebook, LinkedIn, etc. have assumed an important role in the expression of public opinion toward various events, news, and performance of the government, civil services, and public organizations. Many civil services and organizations are utilizing social media as a tool for publicizing products and services. Social media also enables the collection and analysis of their clients' feedback, comments, and reviews for improvement of the service provisions. Social sensing of civil services has become an important tool and a requirement in future smart cities for enabling near-real-time feedback and proper management of services.

In the public eyes, civil services are experiencing a crisis of trust and confidence in the developing world. This crisis provides an important lesson to be learnt for local governments and police departments. Each civil service agency must function as an open system, using a continuous feedback loop to provide performance evaluations and constantly improve service delivery, thus reinvigorating public trust in the provided services. Considering the wide range of services provided by the police in ensuring the safety and security of the public, it is crucial to review the metrics that reflect the whole spectrum of these services and the resulting public feedback and sentiments, such as, is the police ensuring justice or is the police department maintaining its integrity and the concept of internal accountability to remove impediments toward a just society and an impartial policing culture?

Sentiment analysis (SA) is being used as a driver to make information much more meaningful for stakeholders and consumers, making it a revered concept in artificial intelligence (AI) and machine learning (ML). The benefits of public SA on the conditioning of an organization and individual outweigh the costs and time–effort put into developing an SA platform through the social media presence. Recently, public sentiments toward police have been analyzed by Oh et al. (2021) and Ellis (2020). The work by Oh et al. (2021) takes into account 82 metropolitan areas in the United States and determines different factors, such as violent crime rate and racial heterogeneity, which affect public sentiments toward the police. Ellis (2020) provided public perception on the excessive use of force by the Australian police related to a YouTube video and determined that the social media has evolved into a legitimate platform for assessing police performance. However, public offices and departments in the developing world are yet to fully realize the importance of public opinion mining and SA through social media. There is a significant research gap and opportunity to enable SA services in the developing world, where public opinion is usually expressed in local languages, and determine issues that may be specific to the region, country, or the developing world as a whole.

We believe that opinion extraction with respect to civil services, and the law and order departments in Pakistan, both in the local language Urdu and in English, is the first of its kind to the best of our knowledge. Mining the opinions from the public enables us in measuring the gap between the public views and civil services performance of the policing department. Furthermore, utilizing opinion extraction techniques based on ML to automate and future-proof the process has been one of the main elements of the whole study and practice. In Pakistan, Lahore, there has been the establishment of the Punjab Safe Cities Authority, Lahore (PSCA), an integrated command control and communication center, and integration of police with high-tech resources along with equipment to root out crimes, traffic monitoring problems, and traffic violation penalization, and deal with emergency services in the province. A large number of cameras, around 8,000, have been installed all over Lahore city for monitoring and surveillance. The PSCA maintains their social media platforms like Facebook, Twitter, Instagram, and YouTube channel along with Web TV and Radio FM facilities. They provide people up-to-date information regarding law and order and daily traffic alerts, working toward improving their services by analyzing and monitoring public feedback in the form of comments and reviews. Therefore, the foundation for the data mining and the platform from which we can get a better understanding of one of the parts of the civil service, i.e., law and management, is the PSCA. The SA on the PSCA social media data and public feedback about services in the form of mobile applications such as women safety app may on paper just seem to be an ordinary process, but in hindsight, it is the first of its kind of data on public safety and the law and order domain for a developing nation. The main aim is to enhance performance of policing services and highlight the need at grass-rooted and higher levels of hierarchy. In this article, we only focus on the SA results (positive, neutral, and negative) of a target word given in comments, reviews, and feedback.

The major contribution of this article is as follows:

- We create a large-scale data set for SA of civil services and policing authority both in the local language Urdu and in English.
- Many of the comments about local services are written in the local language. Currently, there is a lack of work in SA for Urdu language. We perform SA based on Urdu and English corpus.
- We verify the effectiveness and performance of our model by applying ML models and show that SA of civil services can achieve a good 86.6% accuracy utilizing both English and the local language Urdu.

The rest of sections are organized as follows: **Section 2** lists the "Related Work"; "Data Collection" and processing is described and elaborated in **Section 3**; the experimental results are described in **Section 4**; and the "Results" and their impacts are discussed in **Section 5**. We conclude this work in **Section 6**.

2 RELATED WORK

There is considerable focus on ML techniques for sensing in general (Boulila et al., 2021; Tahir et al., 2020; Tahir et al., 2021; Chebbi et al., 2016) and on social sensing for SA in particular (Han et al., 2020; Cresswell et al., 2021; Oh et al., 2021; Ellis, 2020). Hussain et al. (2021) have utilized a hybrid lexicon and deep learning–based ensemble for SA of social media posts and tweets to determine public sentiments related to COVID-19 vaccination in the United States and United Kingdom. Cresswell et al. (2021) proposed AI-based social media analyses of Facebook and Twitter to determine public sentiments on contact-tracing apps for COVID-19 in the United Kingdom. The authors used both lexicon and deep learning models for improving the performance of sentiment classification. Furthermore, SA on pre- and post-COVID-19 public sentiments has been performed to determine the general reservations and issues faced by the public during and after the pandemic (Nemes and Kiss, 2021). Han et al. (2020) utilized the probability characteristics and Fisher kernel function

derived from Latent Semantic Analysis to determine the sentiments of a Twitter data set. The authors used the support-vector machine (SVM) to determine the classification performance of the proposed features. Boiy and Moens (2009) performed sentiment classification of news forums, reviews, and blog comments in English, French, and Dutch, utilizing SVM, multinomial Naive Bayes (NB), and maximum entropy classifiers. The authors utilize a limited labelled training set with parsed features to improve the classification performance of the ML techniques.

Furthermore, Muhammad et al. (2016) performed elaborate contextual SA on different social media genres by determining the polarities in a context captured by the interaction of words local to the text. A number of ML models have been utilized in research for sentiment classification of text. SVM has been used in different configurations including hybrid SVM algorithms. The work in Mullen and Collier (2004) utilized the comparison of hybrid SVM versus unigram-based SVM to get optimal performance. Moreover, the authors added more features to the data set and topic information to get overall better performance. SVM has been utilized in social media experiments and studies to add more depth and value (Han et al., 2020). Other ML techniques such as NB have been utilized in Singh et al. (2013) to determine the context for movie reviews along with Twitter tweet sentiments regarding movies or media in the study. Aspect-based SA work focuses on a context word and an aspect word whose polarity is to be classified. The sentiment of the aspect word is reflected by its surrounding words in a sentence. Yadav et al. (2021) utilized the sentiment of aspect words which relied heavily on their positions in a sentence and performed SA of the sentences into different polarities.

Deep learning techniques have also been extensively utilized in research for SA. Zhang et al. (2021) presented conversational SA by adapting long short-term memory with confidence gates to model the conversational dynamics between two speakers. The authors also made available a conversational data set ScenarioSA (Zhang et al., 2021) for public use. Moreover, deep neural networks such as CNNs have been used in sentiment classification. The authors Dos Santos and Gatti (2014) focused on short texts including Twitter tweets and single sentences and proposed a deep CNN model for binary classification of sentiments into positive and negative classes.

Oh et al. (2021) utilized SA to determine public perception of policing authority in 82 metropolitan areas of the United States and reported a number of factors of public interest, which shape the general public's opinion on policing performance. The authors found that economic disadvantage, racial heterogeneity, and high crime rate shape public sentiments toward the police. Ellis (2020) interviewed both police and non–police subjects to determine perspectives on the use of excessive force related to a YouTube video in Australia. The authors determined the influence of social media on the police narrative and public perception and concluded that the impact of social media–based public discourse on policing performance cannot be dismissed.

However, the SA of public opinion on civil services and policing is a regionally localized phenomena, and public perceptions may usually be conveyed in non-English local

languages. The current work on natural language processing of non-English languages has a considerable focus on regional languages, such as Arabic, Bangla, and Chinese. Most of the comments and feedback on social media in Pakistan are based in Urdu and to some extent in English. The normalization of text for achieving consistency is an important aspect of processing. Sharf and Rahman (2017) performed a comparative study on normalization of Bangla, Roman Urdu, Arabic, Dutch, Japanese, and Chinese. The authors proposed a model for normalization of Roman Urdu text. Mehmood et al. (2020) presented three-word embedding for Roman Urdu obtained from GloVe, fastText, and Word2Vec. The authors evaluated the integrity of the approaches and provided the first data set for Roman Urdu annotated with positive, neutral, and negative sentiments for SA. The work also provides benchmark results for the data set utilizing various machine and deep learning approaches for SA. The next section discusses the methodology used in our work for sentiment classification of social media data.

3 METHODOLOGY

3.1 Data Collection

The PSCA has not only visual data in the form of videos and images obtained from the 8,000 cameras but also input streams for textual data of a huge size that may be classified into the following:

- social media feedback and comments,
- textual call data for emergency hotline 15,
- discreet feedback on various mobile and web applications, and
- textual data from news channels and sources provided by the media-monitoring cell of the PSCA.

We have the provision of not only using publicly available data on various social media platforms such as YouTube, Facebook, and Twitter but also a feedback repository that is made up of discreet feedback in the form of cellular calls that have been converted into textual files for complaints against the emergency 15 hotline operators. Moreover, mobile applications having a considerably large user base also have feedback-gathering mechanisms from regular users, as well as new users, and the data can be in the form of English, Roman Urdu, Urdu, or a hybrid of English and Roman Urdu, etc. Data cleansing was done by employing simple parsing techniques to remove emojis and unnecessary special characters having low entropy, an example being searching and replacing emojis.

Sorting of data into a header is done based on the main script being used, i.e., English, Roman Urdu, Urdu (using language detection techniques based on the most occurring words).

3.2 Data Extraction

Data are collected from the social media platforms, specifically the data selected from videos on YouTube that are projecting

FIGURE 1 | YouTube comments. (Publicly Available).
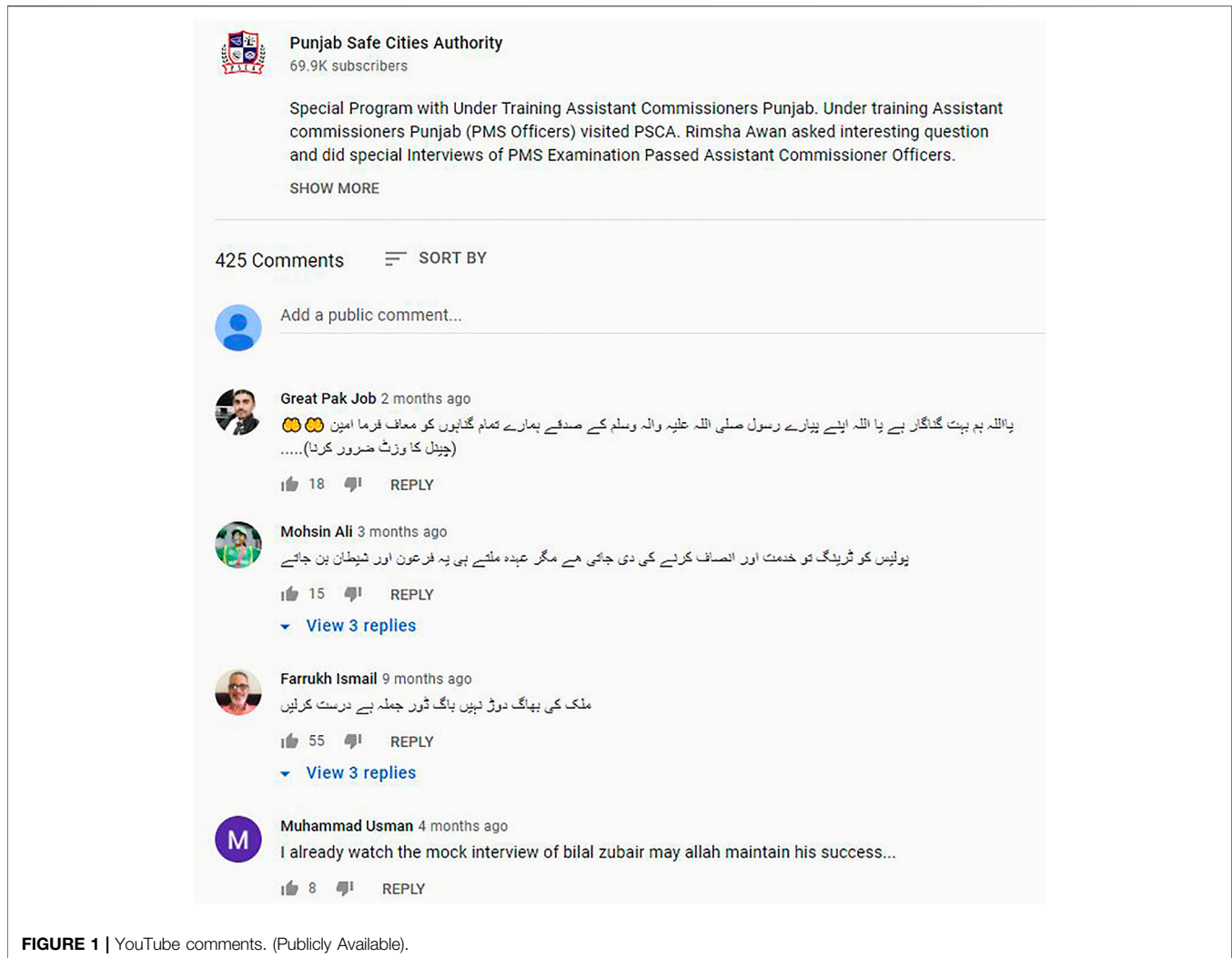
service delivery on behalf of the PSCA. Taking into consideration a certain threshold of views, all the comments are scrapped from the period of uploading the videos till the first iteration of results for this specific article; discreetly, the date and period for this data collection was January 10, 2020 till June 20, 2021. The comments are all considered as feedback in aspect and as our main focus of SA to infer the general feedback of each video.

The number of comments scrapped was 2,000 from the overall YouTube videos, and after the data cleansing round, about 1,600 were available for the initial iteration of analysis. Most of the data were about a general feeling expressed, a question, a remark, and at times, criticism and appraisal of the content being projected. The data were mined from the public outlets related to the PSCA. The scrapping of data in vast amounts on the PSCA content on different platforms such as YouTube, Facebook, etc., was done utilizing python and scrappy. Data were also obtained from internal storage systems and databases. Microsoft SQL servers maintaining the data were queried and the relevant data of feedback from

the applications and complaints were compiled into the CSV file format for readability due to size (**Figure 1**).

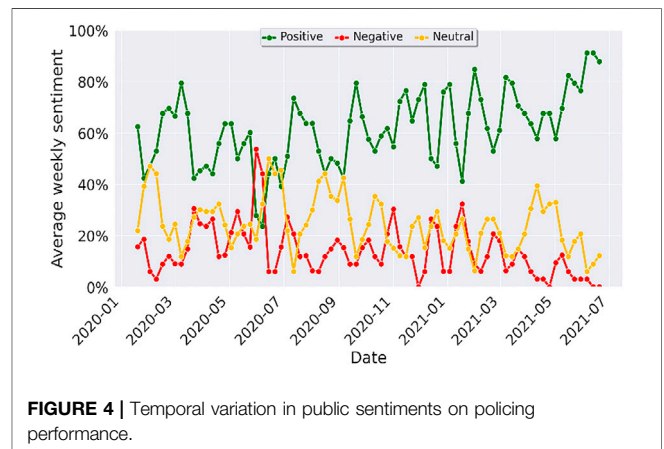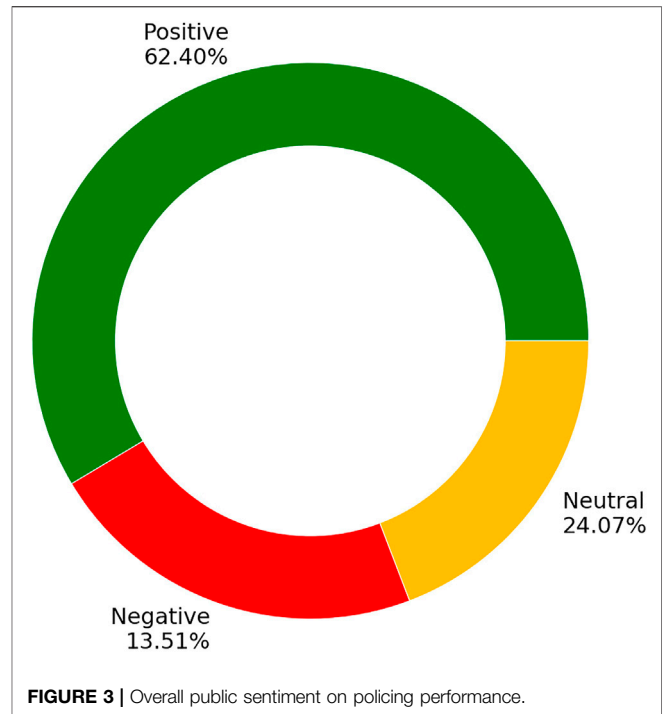# 4 SENTIMENT ANALYSIS OF SMART CITY POLICING PERFORMANCE

Simple parsing techniques were employed to remove emojis and unnecessary special characters having low entropy, an example being searching and replacing emojis. The data were sorted into a header based on the main script being used, i.e., English, Roman Urdu, and Urdu (using language detection techniques based on the most occurring words). All the data were collectively compiled into the CSV file in the form of whole comment texts; in the first phase, 2000 comments from YouTube were scrapped. **Figure 2** illustrates the data collection process.

A basic sentiment scale was used to first of all label all the comments and feedback into a class of positive, negative, and neutral sentiments. The various aspects obtained from the data include casual opinion, gender discrimination, law and order,

DATA BLOCK DIAGRAM

FIGURE 2 | Data scrapping and storage steps.

FIGURE 3 | Overall public sentiment on policing performance.

FIGURE 4 | Temporal variation in public sentiments on policing performance.

false-negative. For example, a given comment/feedback may have words and an aspect that point toward a positive overall score, but in hindsight, it might relate to a sarcastic approach using positively tagged words but actually presenting a negative opinion.

5 RESULTS AND DISCUSSION

The SA of the data was performed for positive, negative, and neutral sentiments with 1,600 comments. A number of ML algorithms were utilized including SVM, logistic regression (LR), multi-layer perceptron (MLP), NB, k-Nearest Neighbor,

civil services, women safety, and safe cities. The topic modeling technique used is latent Dirichlet allocation (Blei et al., 2003).

The topics further help us propagate the actual sentiment behind a certain feedback that may initially be false-positive or

TABLE 1 | Performance results for ML models.

ML model	Accuracy (%)	Precision/recall/F-measure
SVM	86.87	0.87
LR	84.06	0.84
MLP	82.76	0.83
NB	74.65	0.75
KNN	76.32	0.76
RF	77.93	0.78

FIGURE 5 | SVM performance on test data set

FIGURE 6 | LR performance on test data set
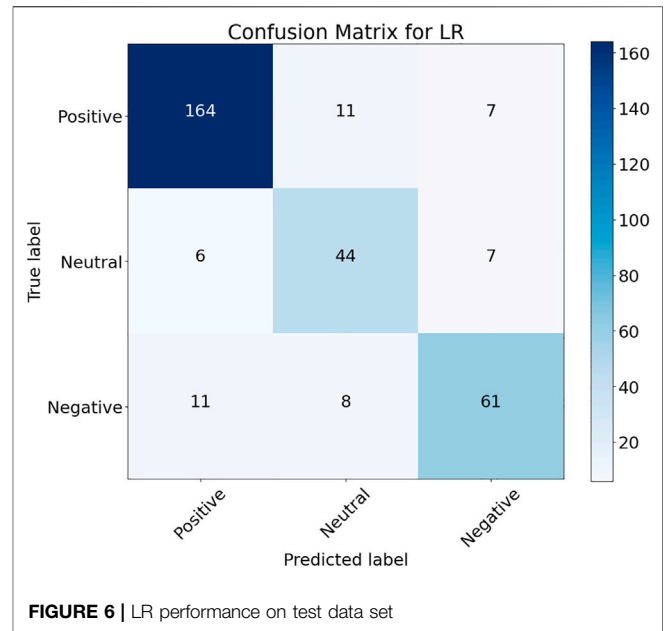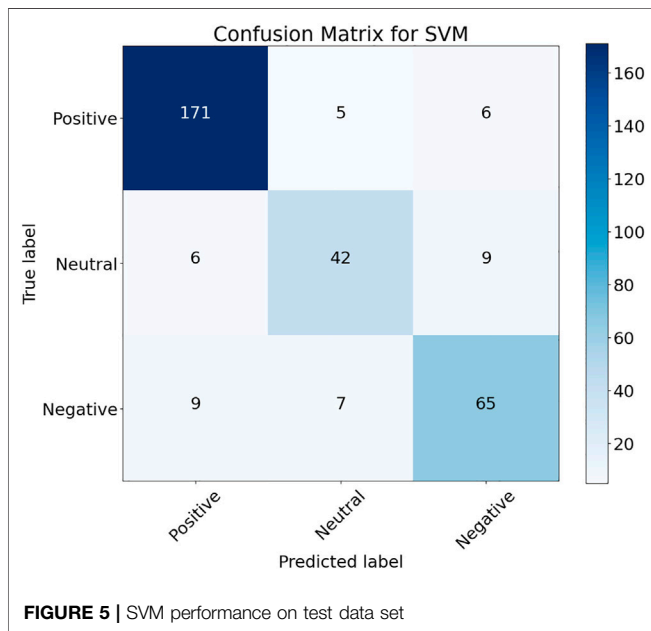
and Random Forest (RF). The hold out validation with 80–20% split was used, with training performed with 80% of the data and testing with the remaining 20% of the data. The performance measures used were accuracy, precision, recall, and F-measure.

The overall public sentiment was positive, with 62.40% comments as positive sentiment, 13.51% as negative, and 24.07% as neutral, as is illustrated in **Figure 3**. The positive public sentiment indicates higher satisfaction with the services provided by the civil services.

The temporal sentiments over time are illustrated in **Figure 4**, which shows weekly variations in public sentiments. The temporal sentiment is also mainly positive followed by neutral and negative sentiments. The sentiment varies around an average value and remains around an average level over time. In July 2020, the public sentiments become temporarily negative toward the service provided due to some of the concerns on women safety. However, the problem seemed to have resolved and the sentiment became overwhelmingly positive again.

The SVM ML model along with all the other models trained on the data are illustrated in **Table 1**. The NB, *k*-nearest neighbors (KNN), and RF had an accuracy of 74.65, 76.32, and 77.93%, respectively, with the NB providing the lowest accuracy, precision, recall, and F-measure. An accuracy of more than 80% was achieved

with the MLP, LR, and SVM algorithms, with SVM providing the highest accuracy of 86.54%. LR also provided a high accuracy of 85.32% with almost comparable precision, recall, and F-measure with respect to the best-performing SVM algorithm. The two best-performing models SVM and LR and their confusion matrices are given in **Figure 5** and **Figure 6**, respectively. The confusion matrices provide results for the 20% test data set, with SVM providing a better classification of sentiments for the positive and negative sentiments, with 164 and 61 correct classifications from a total of 182 positive and 81 negative comments in the test data set as illustrated in **Figure 5**. Moreover, LR works better for the neutral sentiment with 44 correct classifications as compared to 42 for SVM from a total of 57 neutral comments in the test samples, as shown in the confusion matrix in **Figure 6**. A total number of 320 samples from 1,600 were utilized for the test data set with a total test data set of 20% and training data set of 80%.

6 CONCLUSION

This article presents automated AI-based natural language processing of public comments on policing and civil services. This helps to ascertain the satisfaction of the public with policing issues like safety and the services provided. The overall public sentiments along with the temporal variations in the sentiments are provided and discussed in the article. The ML models utilized for determining the public sentiment are given, and their results illustrated. SVM provides the highest accuracy at 86.87% followed by LR at 84.06%. The public sentiments are overwhelmingly positive with 62.40% and negative with 13.51%. The neutral comments have a higher percentage than the negative comments, comparatively. Overall, the public satisfaction with policing and the civil services provided remains high.

DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/Supplementary Material; further inquiries can be directed to the corresponding author.

REFERENCES

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet Allocation. *J. machine Learn. Res.* 3, 993–1022.

Boiy, E., and Moens, M.-F. (2009). A Machine Learning Approach to Sentiment Analysis in Multilingual Web Texts. *Inf. Retrieval* 12, 526–558. doi:10.1007/s10791-008-9070-z

Boulila, W., Sellami, M., Driss, M., Al-Sarem, M., Safaei, M., and Ghaleb, F. A. (2021). Rs-dcnn: A Novel Distributed Convolutional-Neural-Networks Based-Approach for Big Remote-Sensing Image Classification. *Comput. Electronics Agric.* 182, 106014. doi:10.1016/j.compag.2021.106014

Chebbi, I., Boulila, W., and Farah, I. R. (2016). "Improvement of Satellite Image Classification: Approach Based on Hadoop/mapreduce," in Proceeding of the 2016 2nd International Conference on Advanced Technologies for Signal and Image Processing (ATSIP), Monastir, Tunisia, 21-23 March 2016 (IEEE), 31–34. doi:10.1109/ATSIP.2016.7523046

Cresswell, K., Tahir, A., Sheikh, Z., Hussain, Z., Domínguez Hernández, A., Harrison, E., et al. (2021). Understanding Public Perceptions of COVID-19 Contact Tracing Apps: Artificial Intelligence-Enabled Social Media Analysis. *J. Med. Internet Res.* 23, e26618. doi:10.2196/26618

Dos Santos, C., and Gatti, M. (2014). "Deep Convolutional Neural Networks for Sentiment Analysis of Short Texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics* (Dublin: Dublin City University and Association for Computational Linguistics), 69–78. Technical Papers.

Ellis, J. R. (2020). More Than a Trivial Pursuit: Public Order Policing Narratives and the 'social media Test'. *Crime, Media, Cult.* 17 (2), 185–207. doi:10.1177/1741659020918634

Han, K.-X., Chien, W., Chiu, C.-C., and Cheng, Y.-T. (2020). Application of Support Vector Machine (SVM) in the Sentiment Analysis of Twitter DataSet. *Appl. Sci.* 10, 1125. doi:10.3390/app10031125

Hussain, A., Tahir, A., Hussain, Z., Sheikh, Z., Gogate, M., Dashtipour, K., et al. (2021). Artificial Intelligence-Enabled Analysis of Public Attitudes on Facebook and Twitter toward COVID-19 Vaccines in the United Kingdom and the United States: Observational Study. *J. Med. Internet Res.* 23, e26627. doi:10.2196/26627

Mehmood, F., Ghani, M. U., Ibrahim, M. A., Shahzadi, R., Mahmood, W., and Asim, M. N. (2020). A Precisely Xtreme-Multi Channel Hybrid Approach for Roman Urdu Sentiment Analysis. *IEEE Access* 8, 192740–192759. doi:10.1109/access.2020.3030885

Muhammad, A., Wiratunga, N., and Lothian, R. (2016). Contextual Sentiment Analysis for Social media Genres. *Knowledge-Based Syst.* 108, 92–101. doi:10.1016/j.knosys.2016.05.032

Mullen, T., and Collier, N. (2004). "Sentiment Analysis Using Support Vector Machines with Diverse Information Sources," in Proceedings of the 2004 Conference on Empirical Methods in Natural Language Processing, EMNLP 2004, A meeting of SIGDAT, a Special Interest Group of the ACL, held in conjunction with ACL 2004, Barcelona, Spain, 25-26 July 2004

(Stroudsburg, PA: Association for Computational Linguistics (ACL)). undefined.

Nemes, L., and Kiss, A. (2021). Social media Sentiment Analysis Based on COVID-19. *J. Inf. Telecommunication* 5, 1–15. doi:10.1080/24751839.2020.1790793

Oh, G., Zhang, Y., and Greenleaf, R. G. (2021). Measuring Geographic Sentiment toward Police Using Social media Data. *Am. J. Criminal Justice*, 1–17. doi:10.1007/s12103-021-09614-z

Sharf, Z., and Rahman, S. U. (2017). Lexical Normalization of Roman Urdu Text. *Int. J. Computer Sci. Netw. Security* 17, 213–221.

Singh, V. K., Piryani, R., Uddin, A., Waila, P., and Marisha (2013). "Sentiment Analysis of Textual Reviews; Evaluating Machine Learning, Unsupervised and Sentiwordnet Approaches," in Prceeding of the 2013 5th international conference on knowledge and smart technology (KST), Chonburi, Thailand, 31 Jan.-1 Feb. 2013 (IEEE), 122–127. doi:10.1109/KST.2013.6512800

Tahir, A., Ahmad, J., Morison, G., Larijani, H., Gibson, R. M., and Skelton, D. A. (2021). Hrnn4f: Hybrid Deep Random Neural Network for Multi-Channel Fall Activity Detection. *Prob. Eng. Inf. Sci.* 35, 37–50. doi:10.1017/s0269964819000317

Tahir, A., Morison, G., Skelton, D. A., and Gibson, R. M. (2020). Hardware/software Co-design of Fractal Features Based Fall Detection System. *Sensors* 20, 2322. doi:10.3390/s20082322

Yadav, R. K., Jiao, L., Granmo, O.-C., and Goodwin, M. (2021). "Human-level Interpretable Learning for Aspect-Based Sentiment Analysis," in *35th AAAI Conference* (AAAI Press).

Zhang, Y., Tiwari, P., Song, D., Mao, X., Wang, P., Li, X., et al. (2021). Learning Interaction Dynamics with an Interactive LSTM for Conversational Sentiment Analysis. *Neural Networks* 133, 40–56. doi:10.1016/j.neunet.2020.10.001

AUTHOR CONTRIBUTIONS

TM: methodology and conceptualization. AT and AB: formal analysis and writing manuscript. KD, MI, and QA: investigation. All the authors have read and agreed to the published version of the manuscript.