IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

CRIME DETECTION USING SENTIMENT ANALYSIS

Bezan Nanavatti, Kavita Thakur, Rutvik Korade, Sunit Lohade
ALL INDIA SHRI SHIVAJI MEMORIAL SOCIETY
INSTITUTE OF INFORMATION TECHNOLOGY
PUNE, MAHARASHTRA

Abstract: Crime is one of the biggest, and most pervasive problems in our society, and preventing it is a critical duty. Every day, many crimes are committed in large numbers. This makes it an absolute necessity to monitor and make note of all crimes. A database for them that can be used as a resource in the future. In order to predict and solve future crimes, it is currently difficult to keep proper records of criminal activity. The goal of this research is to examine a dataset that includes values of a large number of crimes, in correspondence with sentiment values, and make predictions about intensity and chance of crimes that may occur.

I. INTRODUCTION

The greatest problem, or in a sense, tragedy, to befall humanity is crime. Crimes occur on a regular basis and is a menacing factor in almost every human settlement and environment. To add to this, the amount only increases exponentially. Little towns, large cities, and rural areas all experience crime. Due to this ungodly rise in crime, cases must be resolved at a significantly quicker pace.

We can, at times, anticipate the type and/or intensity of crime[s] that might occur in a specific location, by utilizing a machine learning system with past data or sentiment or some arithmetic logic as it's preface.

II. LITERATURE SURVEY

- Pratibha, A. Gahalot, Uprant, S. Dhiman and L. Chouhan, "Crime Prediction and Analysis," 2nd International Conference on Data, Engineering and Applications (IDEA),
- Kumar, A. Verma, G. Shinde, Y. Sukhdeve and N. Lal, "Crime Prediction Using K-Nearest Neighboring Algorithm," 2020 International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE),
- N. H. M. Shamsuddin, N. A. Ali and R. Alwee, "An overview on crime prediction methods," 2017 6th ICT International Student Project Conference (ICT-ISPC)
- W. Safat, S. Asghar and S. A. Gillani, "Empirical Analysis for Crime Prediction and Forecasting Using Machine Learning and Deep Learning Techniques,"
- J. Kiran and K. Kaishveen., "Prediction Analysis of Crime in India Using a Hybrid Clustering Approach,"
- M. Al Boni and M. S. Gerber, "Predicting crime with routine activity patterns inferred from social media,"

- Gerber, Matthew. (2014). Predicting Crime Using
- Twitter and Kernel Density Estimation, Decision Support Systems.
- X. Chen, Y. Cho and S. Y. Jang, "Crime prediction using Twitter sentiment and weather," 2015 Systems and Information Engineering Design Symposium

III. PROPOSED WORK

The ideal goal and aim of this work is to classify places or areas with the largest number of crimes by applying sentiment analysis provided by an open source dataset of information, namely, “Twitter messages”, to estimate crime rate or density of the same. By giving data on crime at various and all levels in terms of area and location, this work will prove to be useful in aiding an envisioned fall or drop, in crime rates.

Multiple unnerving incidents have left people as a whole, feeling devastatingly fearful for the rest of their lives, maybe because of that one time when they were made to do something they didn't want to do [an example]. Our work's major goal is to primarily concentrate or focus on the functionality and accessibility of social media in promoting and hopefully ensuring the safety of everyone in the world, with a particular reference to the accessibility of social-media website and application-- Twitter.

IV. METHODOLOGY

TWITTER ANALYSIS

Social media sites like Twitter encourage users to actively participate in discussions and share their opinions. Social media can be used to find out what people think and feel about a particular event. There are many platforms for collecting and analyzing opinion-based data that aim to get people's opinions on various topics. There are numerous opinion-based data collection and analysis platforms that attempt to elicit people's thoughts on various topics. Implementation of Sentimental Analysis of Tweets helps create Reports on the tweets retrieved using the Twitter API, which is given by Twitter. Because of this Twitter API, there are a variety of tools for emotive or sentimental-analysis of data on social media. This work made use of a number of readily available libraries, and all data sources and values are open sourced and not copy-righted.

GRAPH

An interactive graph is generated via some social graph model, to introduce a visual representation of the output data of our project. The input would refer to social media data, like data from tweets. An interaction graph represents how social network users interact with each other. In social media, entities [or users] and their interactions are recognized, and an interaction graph is constructed using a vertex set for entities, an edge set for interactions, and an attribute set for both, the vertex (entity) and edge (interaction) attributes.

FINAL REPORT SUMMARIZATION

If there are disproportionately many or purely neutral tweets, it means that people are actually less engaged in the topic at hand, and are unwilling to take sides. Although people's attitudes can change depending on the circumstances, it is also important to keep in mind that the experiments or project’s results are susceptible to change depending on the data. For instance, rape news becomes the most popular topic of the year in 2020. The fact that neutral tweets make up more than 66% of some inquiries makes the viewpoints plainly limited. Considering an example of an analysis performed in relevance to "How safe some areas are", we could conclude that one area is far safer than another.

V. ALGORITHM

Support Vector Machine (SVM) is an administered, AI calculation is included and can be utilized for both grouping and relapse difficulties. Nonetheless, it is generally utilized in order to prevent computational errors and issues. With respect to this calculation, each piece of information is represented as a point in an multi-dimensional space with the value of each component being the value of a certain arrangement. When that happens, we can execute grouping by locating the hyper-plane that clearly divides the groups.

The coordinates of a single perception are essentially what support vectors are. Support Vector Machine is the method that isolates the two classes (hyper-plane/line) the best. A help vector machine constructs a hyper plane or collection of hyper planes in a high- or infinite dimensional space, which can be used for grouping, relapsing, or other tasks like anomalies location. Intuitively, a good partition is produced by the hyper plane with the greatest distance from the nearest prepared information point of any class (referred to as the practical edge), as, generally speaking, the larger the edge, the fewer errors the classifier makes in its predictions. It commonly happens that the sets to separate are not clearly distinguishable in that area, despite the fact that the first problem may be described in a small dimensional space.

VI. ARCHITECTURE

The user signs in via a home page or opens a local application to access our software, after which he/she uploads the data (or tweets), and then implements sentimental analysis of comments or values stored in the database in the form of positive and negative text analysis, based on which a modified graph is formed. To add to this, we can use available data on crimes committed along with the sentiment of the area to provide our uses with a map entailing the crime as well as sentiment in a proportionally defined map with colors or indicators which describe the “best-of” or “worst-of” both factors (crimes reported and sentiment value of the same area).

An example of sentiment analysis in a tabular format is given following this paragraph, with the positive and the negative sentiment values loaded in percentages over the total tweets in the area. (Tentative values are used here)

Table- Analysis in tabular format:

S. No.	Name of Cities	Positive Analysis (%)	Negative Analysis (%)
1	Lucknow	9	1
2	Delhi	14	20
3	Kolkata	9	11
4	Pune	12	8
5	Mumbai	18	2
6	Goa	9	11
7	Chennai	13	9

A diagrammatic format which represents the working of our application is presented as follows:

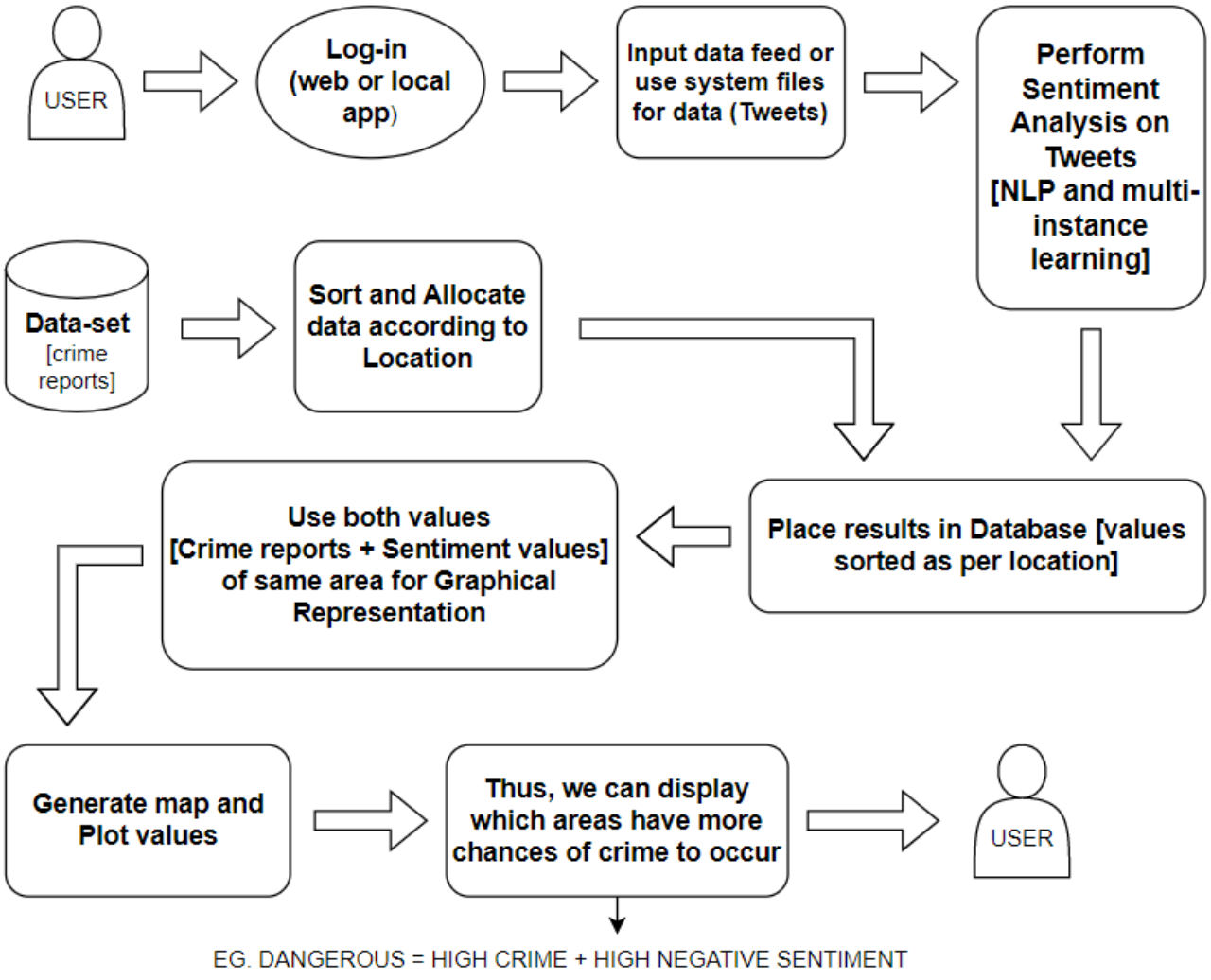

Figure 1. Architecture/Working of our system

VII. CONCLUSION

In this study, we examined a number of AI evaluations that can assist us in sorting through and breaking down the enormous amount of Twitter data we have gathered, which includes a significant number of tweets and instant messages released every day. When it comes to analyzing massive volumes of data, certain AI calculations like the SPC computation and direct logarithmic Factor Model techniques are incredibly compelling and helpful. They also aid in grouping the data into pertinent categories. A different type of AI calculation called a backing vector machine is frequently used to extract crucial information from Twitter and create a structured inlay of data values that can be used in our analysis and arithmetic algorithms to help approach our target, or aim, which is to predict the time and/or location in which a specific type of crime will occur. Our approach is based on sentiment analysis. Twitter is used extensively in the United States as well as globally, creating many opportunities to augment decision support systems with Twitter-driven predictive analytics. Twitter is an ideal data source for decision support: its users, who number in the millions, publicly discuss events, emotions, and innumerable other topics; its content is authored and distributed in real time at no charge; and individual messages (also known as tweets) are often tagged with precise spatial and temporal coordinates.

VIII. FUTURE SCOPE

This research can be expanded upon to create profiles of significant influence in particular areas of crime using multiple psychological frameworks. The future goal is to foresee when and where a particular kind of crime will occur. Our strategy is grounded in (dedicated to) sentiment analysis. The main topics of this research's future work include a more in-depth analysis of message content, temporal modelling, and the inclusion of auxiliary data sources. We may provide the data to state-level analysts to help them lower crime rates in their individual cities thanks to the effort that helps us identify crime rates, and create an open-source platform for all to access it. As a result of this work's effective crime detection capabilities, crime could be decreased in a number of cities. Future programs or procedures led by the government can implement our analysis methods and software, etc. to aid them in preparation of crimes that may occur and information analysis and collection of data that may make it easier for them to tackle it.

IX. REFERENCES

1) >Agarwal, S., Godbole, S., Punjani, D., & Roy, S.(2007, October). How much noise is too much: A study in automatic text classification. In Seventh IEEE International Conference on Data Mining (ICDM 2007) (pp. 3-12). IEEE.
2) >Ahmad, K., Han, J., Hutson, E., Kearney, C., & Liu, S. (2016). Media-expressed negative tone and firmlevel stock returns. Journal of Corporate Finance, 37, 152-172.
3) >Baid, P., Gupta, A., & Chaplot, N. (2017). Sentiment analysis of movie reviews using machine learning techniques. International Journal of Computer Applications, 179(7), 45-49. >Bollen, J., Mao, H., & Pepe, A. (2011, July).Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. In Proceedings of the international AAAI conference on web and social media (Vol. 5, No. 1).
4) >Choudhury, M., Saraf, R., Jain, V., Mukherjee, A., Sarkar, S., & Basu, A. (2007). Investigation and modeling of the structure of texting language. International Journal of Document Analysis and Recognition (IJDAR), 10(3), 157-174.
5) >Diakopoulos, N. A., & Shamma, D. A. (2010, April). Characterizing debate performance via aggregated twitter sentiment. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1195-1198).
6) >Gamon, M. (2004). Sentiment classification on customer feedback data: noisy data, large feature vectors, and the role of linguistic analysis. In COLING 2004: Proceedings of the 20th International Conference on Computational Linguistics (pp. 841-847).
7) >Kabir, M., Kabir, M. M. J., Xu, S., & Badhon, B. (2019). An empirical research on sentiment analysis using machine learning approaches. International Journal of Computers and Applications, 1-9.
8) >Kolchyna, O., Souza, T. T., Treleaven, P., & Aste, T. (2015). Twitter sentiment analysis: Lexicon method, machine learning method and their combination. arXiv preprint arXiv:1507.00955.
9) >Matsumoto, S., Takamura, H., & Okumura, M. (2005,May). Sentiment classification using word subsequences and dependency sub-trees. In Pacific-Asia conference on knowledge discovery and data mining (pp. 301-311). Springer, Berlin, Heidelberg. Medhat, W., Hassan, A., & Korashy, H. (2014). Sentiment analysis algorithms and applications: A survey.
10) >Ain Shams engineering journal, 5(4), 1093-1113. O'Hare, N., Davy, M., Bermingham, A., Ferguson, P., Sheridan, P., Gurrin, C., & Smeaton, A. F. (2009, November). Topic-dependent sentiment analysis of financial blogs. In Proceedings of the 1st international CIKM workshop on Topic-sentiment analysis for mass opinion (pp. 9-16).
11) >Pang, B., & Lee, L. (2009). Opinion mining and sentiment analysis. Comput. Linguist, 35(2), 311-312. Renault, T. (2020). Sentiment analysis and machine learning in finance: a comparison of methods and models on one million messages. Digital Finance, 2(1), 1-13.
12) >Sarlan, A., Nadam, C., & Basri, S. (2014, November).Twitter sentiment analysis. In Proceedings of the 6th International conference on Information Technology and Multimedia (pp. 212-216). IEEE.
13) >Siddiqui, S., Darbari, M., & Yagyasen, D. (2020). Modelling and Simulation of Queuing Models through the concept of Petri Nets. ADCAIJ: Advances in Distributed Computing and Artificial Intelligence Journal, 9(3), 17-28.
14) >Wang, S. I., & Manning, C. D. (2012, July). Baselines and bigrams: Simple, good sentiment and topic
15) >And multiple open-sourced data and information aided in creating a path to success for our undertaken project here.