Deep Neural Network-Based Approach to Classification the Crime Related News Posts

S.P.C.W Sandagiri
Department of Computing and Information Systems,
Sabaragamuwa University of Sri Lanka
Belihuloya, Sri Lanka
chamithsandagiri@gmail.com

***Abstract*— Crime is a major problem faced today by society. Crimes have affected the quality of life and economic growth badly. We can identify the crime patterns and predict the crimes by detecting and analyzing the historical data. However, some crimes are unregistered and unsolved due to a lack of evidence. Researchers used different sources to get crimes related data to generate the prediction model. But, some crimes are unregistered. In this paper, we used online news posts to detect crimes. crimes. As the first step, we fetch the news posts using predefined keywords relating to the crimes. Then, we proposed the Long short-term memory (LSTM) approach to the classification of crime types and non-crime related posts. Our approach outperformed the existing approaches by obtaining 88.4% accuracy.**

Keywords— Crime detection, Online news, LSTM, GloVe

I. INTRODUCTION

Security is a very necessary aspect of life. Unless we are safe, our most important needs cannot be met. Security is therefore a requirement in human life that helps us to achieve our goals collectively or individually [1]. Crimes are a social problem, which costs our society deeply in many aspects. The ability to identify unsafe areas for crime and identify the most recent crime in a particular location has become a growing concern for both local authorities and residents. On the other hand, people are always interested in improving safety and make reliable relationships with neighbors when living in a busy society. The prevalence of crime is one of the greatest challenges for societies around the world, particularly in metropolitan areas [2]. The definition of daily crime applies to three factors: (i) driven criminals, (ii) the availability of an appropriate target, and (iii) the absence of competent guardians. [3].

Society struggles everywhere with the illegal acts of people or syndicates. Crime from minor theft to organized gangs is a threat to humanity and a huge problem for cities across the globe. In today's society, there is most certainly no area that will not have to deal with the threat of violence and the consequences. Many people have been subject to different types of violence or have been directly attacked. In our cultures, violence happens in multiple ways and has different implications at various times [2].

Previous work on predicting criminal incidents has primarily relied on the history of crime and various geospatial and demographic sources of information. While promising, the rich and quickly growing social media environment around events of interest are not included in these models. [4].

and has numerous repercussions. Studies have been undertaken over the years to understand the nature of criminal behavior, to identify individuals by their racial or cultural context, or to establish ways of detecting, anticipating, and prosecuting them effectively [5]. Governments around the world expend vast sums on crime prevention, compliance, and citizens' knowledge. In the last few years, people have noted the rising number of CCTV cameras looking into the streets of all big cities. Crime information is not easily available to the public [6].

There are more researches regarding crimes in the world, but using news media, there are few types of research about crimes and their behavior. Therefore, the paper aims in presenting a prediction model (algorithm) by using the deep learning technique, which is meant to possess a strong capability to predict crimes by factors of news media dataset using the Data Mining concept. Our main data source is the Online news media. The main goal is to identify each hidden data source and predict results.

The remainder of the paper is organized as follows: Section 2 discusses literature review and related works while section 3 presents the proposed methodology. Section 4 illustrates the results and discussion. Finally, section 5 concludes the paper with the conclusion and future work.

II. LITERATURE REVIEW AND RELATED WORKS

Crime behavior in many areas has historically been investigated, such as sociology [7], psychology [8], and economics [9], disaster management [10-11] as an unavoidable and important social problem in our societies. In addition to recent advances in big data analytical techniques, we can thoroughly analyze and understand crime Concerning current developments in big data processing methods, we can track and understand crime thoroughly [5].

Advances in the discovery of knowledge and the development of IT help the police to analyze past data and to predict future events. Data from three categories are used in crime analytics, technology, and applications: historical activities of criminals, Spatio-temporal information, and recently social media [12].

Conventional models of analysis of crime depend on social statistical indices and on demographic details obtained from localized crime zones known as maps of hot spots to forecast the proportions of a crime of various kinds. However, there is some debate as to whether the concentration of all types of crime is indicated on these maps [13]. For example, taxicab robberies occur in various locations that are not always indicative of areas of high crime [12].

In a detailed summary Bontcheva and Rout [14] identify core research problems and concerns in social media mining semantics. They have a list of semantic technologies and processes. While it is difficult to derive accurate informants from noisy, arbitrary social information, research indicates that sentiment analysis is an effective tool for extracting value added in the various fields [15 - 17].

As the aforementioned literature indicates, in recent years, the predictive potential of social media and in particular Twitter feeds has been confirmed. Social networking can also be used in several industries to produce added value. Nearly all of the above methods use volumes or sentiment analyses to forecast or describe specific events and situations, but mostly without any other details, such as fine-grained temporal or spatial detail.

Another approach to study is to compare two values from the same source; chosen crowds on aggressive twitters at two football stadiums in England [18]. Such findings do not provide more information for certain variables. Even the analysis of the same datasets needs recognition of another aspect when it comes to crime prediction. Besides, this study gathers basic data from the week. It is not more reliable than collecting data every day.

The better allocation of finite capital is supposed to maximize crime prediction. However, modern crime prediction methods have some drawbacks in predicting crime events since the predictive model does not include a crime predictor [19].

III. Proposed Approach

Fig. 1 shows the proposed deep learning-based crime detection approach. First, news posts that relate to the crimes are extracted. Then, preprocessing techniques are applied to clean the data set. Then, news posts are transformed into vectors to generate the feature vectors in the data preparation step. Finally, the neural network model is constructed to classify the data set.

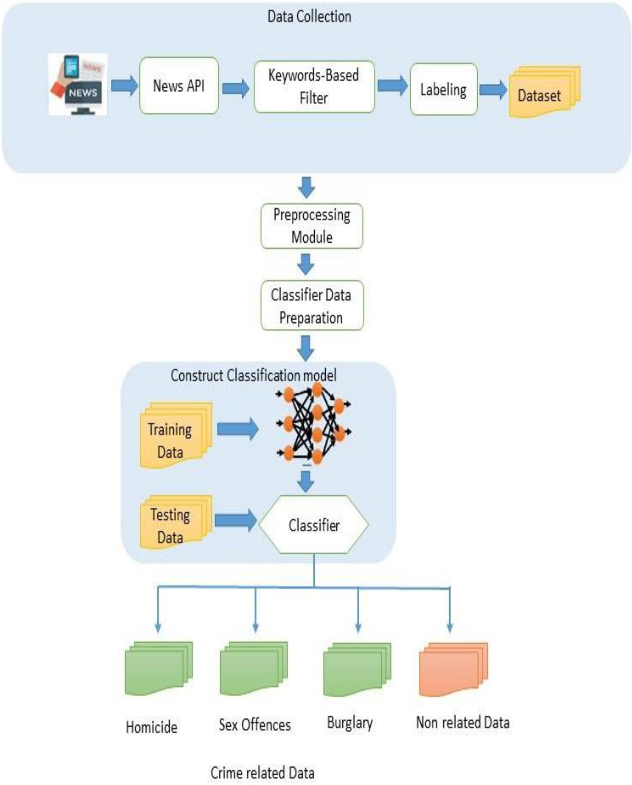

Fig. 1. Proposed methodological framework

A. Data collection

News posts are collected through a News Search API (available at https://currentsapi.services/en). The search of the news posts must be based on a set of keywords that can be used to classify the crime situations. Table I shows some of the keywords used for selected crime categories. Then, news posts are labeled based on the contents used to create the training set. The collected data set consists of more than 150,000 news posts from 2020 January 01 to 2020 January 31.

TABLE I. Crime Categories and Keywords Used

Crime Type	Keywords
Burglary	robbery, kidnapping, thief, steal, hijacking
Homicide	assassin, murder, death, suicide, shoot
Sex Offences	rape, abuse, sex offenses

B. Data preprocessing

The next move was to clean the data and only provide useful information for further analysis after receiving the required dataset. There were several issues with the dataset obtained in phase 1. So, as the next step, pre-processing techniques are very important for the extracted data. Since, in the Twitter post, there can be typos, inappropriate content like URLs, and stop words.

Data obtained from Twitter are therefore highly unstructured and noisy. Because of this, there will be a huge over-load for further processing the Twitter posts. Pre-processing techniques are used to produce clean tweet data for the next step. Further, it makes the text more digestible to improve the performance of machine-learning algorithms. The following data pre-processing steps were followed in our approach;

1) *Stopwords removal*

Stopwords are very normal words. In NLP tasks like sentimental analysis or text classification, words such as “we”, “our”, “have” etc. probably won't help at all. We, therefore, removed those stopwords to save time and effort to process large volumes of text.

2) *Unwanted content removal*

The following parts in the text corpus can also lead to an overload in the text analysis. Therefore we removed the following unwanted contents from the posts as our second step in the data preprocessing stage.

- HTML Tags
- URLs
- Extra whitespaces
- Emoji
- Quotes
- Hashtags
- Special characters
- Numbers

3) *Convert Accented Characters*

An accent (also called a diacritic) is a letter-added glyph. The word is derived from ancient Greek. Diacritic is mainly an adjective, although sometimes used as a noun, whereas diacritical is only an adjective. Accented terms such as “hôtel” and “café” can be translated into “hotel” and “cafe”

4) *Normalizing*

The purpose of normalizing is to equilibrate all text, e.g.: converting all characters to lowercase.

5) *Tokenizing*

Tokenization involves dividing text strings into smaller pieces, or "tokens." Paragraphs can be tagged to phrases, and phrases can be tagged to words.

6) Lemmitizing

Lemmatization is the method of transforming a term into its root, e.g. "shooting" into "shoot". Another way to get a word's basic form is stemming. Since it has some cons, we haven't used it in our preprocessing. For example, stemming "raping" will result in "rap"

We used Natural Language Toolkit (NLTK) to conduct the above-mentioned preprocessing steps. NLTK is a suite of libraries and programs for the processing of conceptual and computational in natural languages using Python.

C. Data Preparation

After completing the pre-processing, news posts are transformed into vectors to generate the feature vectors. The vectors are used in the learning phase for machine learning algorithms. Here, we used word embedding techniques to create the vector.

In word embedding, every word is represented as an N-dimensional dense vector with numerical representations. The technique allows words with similar meaning to have a similar representation. In machine learning, we can train our embedding model with the given corpus or we can use any pre-trained word embedding technique to perform NLP tasks.

There are several types of pre-trained word embedding techniques such as Word2Vec (https://code.google.com/archive/p/word2vec). Word2Vec is a statistical method for efficiently learning a standalone word embedding from a text corpus. In this study, we used Global Vectors for Word Representation (https://nlp.stanford.edu/projects/glove) word embedding technique. It is an extension of the Word2Vec model of learning word vectors effectively GloVe consists of 100-dimensional vectors and was generated from 2 billion words, which is a far larger training dataset than the one collected from our own. The following are the main steps that we followed here.

1) Sentence vectors

After completing the pre-processing, news posts are converted into sentence vectors assigning a unique integer to each element in vocabulary. After having experienced different lengths of sentences, we updated all sentences length to 100. If sentences have more words, the extra words have been removed. When they were less than 100 words in sentences, we used an alternative method of zero paddings to fill missing values. It iterated through each news post until the expected size was achieved.

2) Dense vector

The next step is to load the GloVe word embedding and then create dense vectors that contain the words in our corpus and their corresponding values from GloVe embedding. We create a dense vector-matrix the same as Sentence vectors.

D. Constructing the LSTM Base Model

The proposed model is combined with an MLP with a set of layers. The input layer (i.e. embedding layer) is created based on a word embedding method. There are two layers used as a hidden layer. The first one is the LSTM layer with a Bidirectional wrapper. This propagates the input forwards and backward through the LSTM layer and then concatenates the outputs. This helps LSTM to learn long term dependencies. The second one is a Dense Layer with 16 nodes. Since this is a multi-class classification, the output layer is defined with a dense layer with 5 nodes. Classification classes are Homicide, Sex Offences, Burglary, and non-crime.

The training of the neural network is performed using as training data, the integer sequence list represented by the features contained in the 2-D matrix (X), and the vector that contains the class (Y). Configuration parameters are presented in Table II. The classification is the output generated by the neural network, which, through a training phase, allows classify the class label for a new instance, from a set of input features of an instance.

TABLE II. CONFIGURATION PARAMETERS OF THE NETWORK

Parameter	Value
epochs	50
LSTM Unit	100
batch size	32
optimizer	Adam
loss	sparse_categorical_crossentropy
activation function (output layer)	softmax

IV. RESULTS AND DISCUSSION

The tests were performed with an Intel Core i5-3770, 2.70 GHz CPU, and 6 GB RAM on a computer running Microsoft Windows 10. Python was used for neural network implementation. Tweets linked to crime are obtained from the Twitter API. Each tweet is classified according to the information for training.

A. Performance Metrics

We used several standard criteria such as accuracy and loss to evaluate the performance of the LSTM-based classification approach: Accuracy is the most important measure for evaluating the model.

Accuracy represents the ratio of tuples that are correctly classified by the model. It is calculated using the following equation:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{N}}$$

The recall is the fraction of the total amount of relevant instances that were retrieved.

$$\text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}}$$

Precision is the fraction of relevant instances among the retrieved instances:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}}$$

Where TP is true positive news posts, FN is a false negative news post, TN is true negative and FP is false positive posts. F1 score is the harmonic mean of precision and recall:

$$\text{F1} = 2 \times \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Presicion}}$$

There are many different classifiers, and their performance depends on the problem. To evaluate the proposed LSTM Base classifier with the following classification model.

(i) ANN + TF-IDF: As we mentioned, news posts are transformed into vectors to generate the feature vectors. Here we used the TF-IDF method to generate the vectors.

(ii) ANN + GloVe: Here we used the GloVe method to generate the vectors and neural network with one hidden layer.

B. News posts detection

Given below are some samples of News post headings correctly identified by our approach.

"A New Hampshire man accused of sexually assaulting two girls and manufacturing child sexual abuse images was sentenced to prison Thursday."

"A 20-year-old Chicago woman faces murder charges after police say she killed two of her young sons, leaving one in a bathtub and throwing another from an 11th-floor window."

"Walter Johnson, 35, was charged with first-degree murder in the slaying of Devell Hill, 26, at the corner of Madison Street and Karlov Avenue around 2:30 p.m., Chicago police said."

"Baltimore police say a man who claimed to have been set on fire during a robbery was, in fact, not set on fire during a robbery but was set on fire during illegal drug activity"

From the above tweets, it is straight forward that the above News were related to crime. The first one is related to Sex Offences, second and third Tweets are related to Homicide. The fourth post is related to Burglary.

C. Performance: Training data vs Testing Data

Neural network training is the process of fine-tuning the weights and biases from the input data. The optimization process updates the weight parameters to minimize the loss function and calculate the model error. The loss function gives a brief understanding of a model erroneous with a numerical representation. It helps to improve better models. Fig 2 shows the variation of loss for training and testing data. The final neural network model achieves an average loss function of 0.384 on the testing dataset and 0.302 on the training dataset.

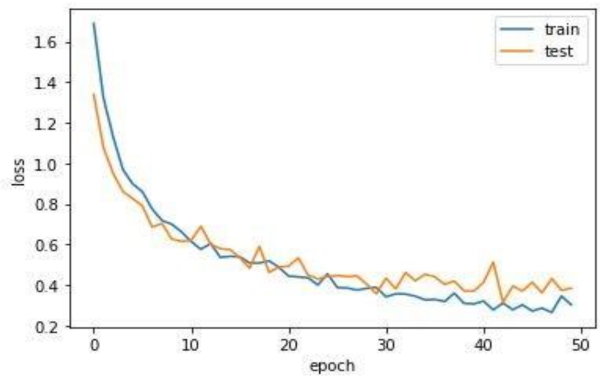

Fig. 2. Loss function in training and validation phases by epoch

The performance of the model is evaluated through accuracy as well. Fig 3 shows the variation of accuracy for training and testing data. The final model achieves an average accuracy of 0.884 on the testing dataset and 0.91 on the training dataset. Fig 3 is exhibiting the test data curve is aligned with the training data curve. This means our final model works well with new data.

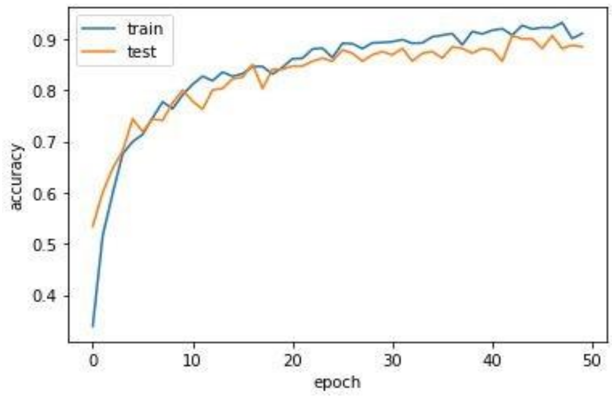

Fig. 3. Accuracy in training and validation phases by epoch

D. Performance Comparison

We compared our crime classification approach with the ANN + GloVe approach and ANN + TF-IDF approach. Fig. 4 and Fig. 5 show the variation of accuracy and loss for LSTM + GloVe, ANN + GloVe, and ANN + TF-IDF approaches respectively.

Table III shows the results for various evaluation metrics vs for LSTM + GloVe, ANN + GloVe, and ANN + TF-IDF approaches. From Table III, it can be seen clearly that the proposed approach for classifying the crime outperformed the other approaches.

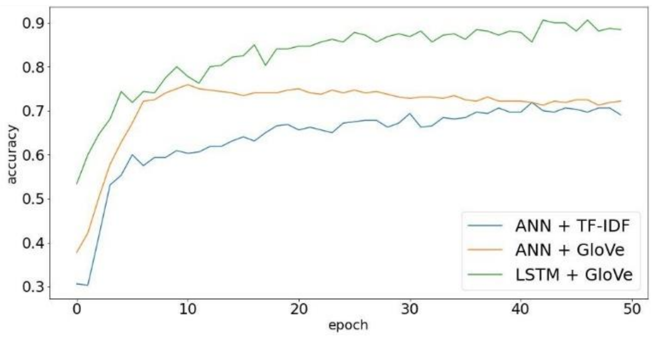

Fig. 4. The variation of accuracy for LSTM + GloVe, ANN + GloVe and ANN + TF-IDF for crime classification

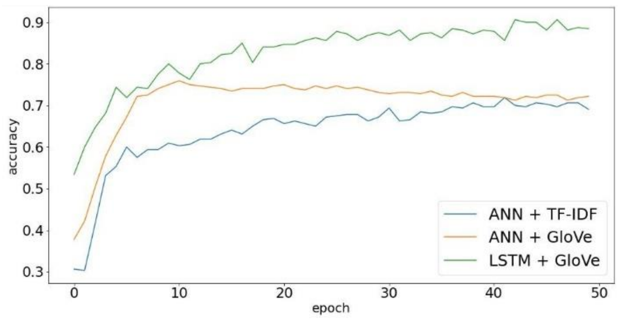

Fig. 5. The variation of loss for LSTM + GloVe, ANN + GloVe and ANN + TF-IDF for crime classification

TABLE III. RESULTS COMPARISON

Models	Accuracy	Precision	Recall	F1-score	Loss
ANN + TF-IDF	0.690	0.68	0.7	0.69	1.03
ANN + GloVe	0.721	0.75	0.72	0.70	0.75
LSTM + GloVe	**0.884**	**0.90**	**0.88**	**0.88**	**0.38**

V. Conclusion and Future Works

In this paper, we proposed an approach to improving the accuracy of the classification of crime types and non-crime related posts from online news platforms. First, we fetched the News using API, and then we used NLP techniques to clean the news data. Next, we applied the LSTM based classification model to filter out the crime nonrelated news.

We evaluated our research by comparing it to the ANN + GloVe approach and ANN + TF-IDF approach. The empirical study of our prototyping system has proved the effectiveness of our approach. It obtained 88.4% accuracy and 90% Precision. We strongly believe that our approach outperformed well in removing the crime non-related news and classify crimes according to type. This removal of crime non-related news can pave the way for future researches such as prediction, hotspot analysis, crime analysis, etc. in high accuracy.

Our future work falls in two folds. Firstly, we hoped integral Social Media data to provide a more sophisticated rating scale. Finally, we also planned to implement the crime prediction approach using a deep neural network approach.

References

[1] A.-S. Hissah and H. Al-Dossari, "Detecting and Classifying Crimes from Arabic Twitter Posts using Text Mining Techniques," International Journal of Advanced Computer Science and Applications (IJACSA), vol. 9, 2018.

[2] J. Bendler, A. Ratku, and D. Neumann, "Crime mapping through geospatial social media activity," 2014.

[3] L. E. Cohen and M. Felson, "Social change and crime rate trends: A routine activity approach," American sociological review, pp. 588-608, 1979.

[4] X. Wang, M. S. Gerber, and D. E. Brown, "Automatic crime prediction using events extracted from Twitter posts," in International conference on social computing, behavioral cultural modeling, and prediction, 2012, pp. 231-238.

[5] D. Yang, T. Heaney, A. Tonon, L. Wang, and P. Cudré-Mauroux, "CrimeTelescope: crime hotspot prediction based on urban and social media data fusion," World Wide Web, vol. 21, pp. 1323-1347, 2018.

[6] J. Bendler, A. Ratku, and D. Neumann, "Crime mapping through geospatial social media activity," 2014.

[7] D. Garland, "Governmentality'and the problem of crime: Foucault, criminology, sociology," Theoretical Criminology, vol. 1, pp. 173-214, 1997.

[8] S. Z. Levine and C. J. Jackson, "Eysenck's theory of crime revisited: Factors or primary scales?," Legal and Criminological Psychology, vol. 9, pp. 135-152, 2004.

[9] A. K. Lynch and D. W. Rasmussen, "Measuring the impact of crime on house prices," Applied Economics, vol. 33, pp. 1981-1989, 2001.

[10] K. Banujan, B. Kumara, and I. Paik, "Strengthening Post-Disaster Management Activities by Rating Social Media Corpus," International Journal of Systems Service-Oriented Engineering, vol. 10, pp. 34-50, 2020.

[11] K. Banujan, T. B. Kumara, and I. Paik, "Twitter and Online News analytics for Enhancing Post-Natural Disaster Management Activities," in 2018 9th International Conference on Awareness Science and Technology (iCAST), 2018, pp. 302-307.

[12] S. Aghababaei and M. Makrehchi, "Mining Twitter data for crime trend prediction," Intelligent Data Analysis, vol. 22, pp. 117-141, 2018.

[13] J. Eck, S. Chainey, J. Cameron, and R. Wilson, "Mapping crime: Understanding hotspots," 2005.

[14] K. Bontcheva and D. Rout, "Making sense of social media streams through semantics: a survey," Semantic Web, vol. 5, pp. 373-403, 2014.

[15] F. Neri, C. Aliprandi, F. Capeci, M. Cuadros, and T. By, "Sentiment analysis on social media," in 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, 2012, pp. 919-926.

[16] W. Chamlertwat, P. Bhattarakosol, T. Rungkasiri, and C. Haruechaiyasak, "Discovering Consumer Insight from Twitter via Sentiment Analysis," J. UCS, vol. 18, pp. 973-992, 2012.

[17] D. Choi and P. Kim, "Sentiment analysis for tracking breaking events: a case study on Twitter," in Asian Conference on Intelligent Information and Database Systems, 2013, pp. 285-294.

[18] A. Ristea, C. Langford, and M. Leitner, "Relationships between crime and Twitter activity around stadiums," in 2017 25th International Conference on Geoinformatics, 2017, pp. 1-5.

[19] X. Chen, Y. Cho, and S. Y. Jang, "Crime prediction using Twitter sentiment and weather," in 2015 Systems and Information Engineering Design Symposium, 2015, pp. 63-68.