The Nexus Between Big Data Analytics and the Proliferation of Fake News as a Precursor to Online and Offline Criminal Activities

Spyridon Evangelatos[‡,§], Thanasis Papadakis[‡], Nikolaos Gousetis[‡], Christos Nikolopoulos[§], Petrina Troulitaki[⋆], Nikos Dimakopoulos[⋆], George Bravos[⋆], Michael V. Lo Giudice[†], Ali Shadman Yazdi[†], Alberto Aziani[†]

[‡] *Research & Innovation Development, Netcompany-Intrasoft S.A.*, Luxembourg, Luxembourg

[⋆] *Research & Development Department, Information Technology for Market Leadership*, Athens, Greece

[†] *Università Cattolica del Sacro Cuore and Transcrime*, Milan, Italy

[§] *Department of Electronic Engineering, Hellenic Mediterranean University*, Crete, Greece

Emails: {Spyros.Evangelatos, Thanasis.Papadakis, Nikolaos.Gousetis}@netcompany.com, cnikolo@hmu.gr, {PTroulitaki, NDimakopoulos, GeBravos}@itml.gr, {Michaelvictor.Logiudice, Ali.Shadma, Alberto.Aziani}@unicatt.it

Abstract—This paper presents a novel framework for the thorough analysis of fake news and disinformation campaigns, which have the potential to result in both offline and online criminal activities. With a primary focus on the analysing the spread of disinformation across social media and online platforms, it aims to uncover the underlying dynamics and mechanisms driving the dissemination of false information. The developed framework integrates state-of-the-art natural language processing (NLP) techniques, for sentiment analysis, deep learning (DL) algorithms, for the prediction of criminal activities related to the disinformation spread, and graph analysis, in identifying the key actors and propagation pathways. To address the emerging challenges of disinformation, that transcend the online realm and have tangible real-world consequences, this framework extends its analysis to potential offline actions incited by disinformation, such as acts of violence and public unrest or the disruption of public health efforts, especially in case of pandemics. By exploring the complex interconnections between disinformation and crime, our research aims to contribute to a deeper understanding of the societal implications of false information and provide actionable insights for policymakers, security practitioners and the broader public.

Index Terms—Big Data Analytics, Artificial Intelligence, Fake News and Disinformation Analysis, Online and Offline Crimes

1 Introduction

The advent of social media has ushered in a profound transformation in the landscape of human communication and information sharing. The digital revolution has not only exponentially increased the volume of generated data but has also given rise to a burgeoning challenge in the realm of disinformation and fake news (D&FN), with implications extending towards criminal activities [1]. Within this context, the domain of data analysis assumes a pivotal role, as researchers grapple with the intricate facets of D&FN among the vast expanse of social media data. D&FN, characterized by its deceptive and often false information, poses an inherent threat capable of inflicting considerable harm [2]. This perilous propagation has the capacity to exert a formidable influence over public opinion, engendering a state of confusion and eroding trust in established sources of information [3]. The dynamic interplay of D&FN with the analysis of big social media data is pivotal in comprehending its far-reaching consequences. The virulent influence wielded by D&FN, on public opinion, is conspicuously observable [4]. False information disseminates with unprecedented velocity within the confines of social media platforms, primarily due to the prevalent tendency of individuals to uncritically embrace online content, often eschewing the rigorous scrutiny of its veracity.

In light of the escalating magnitude of the D&FN predicament and its palpable repercussions on public opinion, the European Union (EU) has adopted a comprehensive approach to mitigate its dissemination. Within said approach, the EU initiated the FERMI project with the primary objective to develop a holistic framework comprising interconnected analytical tools that analyse fake news and various disinformation campaigns that may lead to offline and online crimes. Furthermore, the project aspires to delineate and recommend contextually relevant security countermeasures tailored to distinct geographic locations and segments of society, thereby mitigating the pernicious influence of D&FN. In summary, the symbiotic relationship between the burgeoning realm of big data analytics and the vexing conundrum of D&FN on social media is undeniably intricate. The multifaceted dimensions of this issue necessitate innovative approaches, such as the FERMI project, to not only elucidate the mechanics of D&FN dissemination but also to fortify our societal resilience against its deleteri-

This work has received funding from the European Union's Horizon Europe research and innovation programme through FERMI project under grant agreement No. 101073980.

ous effects.

2 Related Work

The fight against disinformation campaigns and fake news through big data analytics has witnessed significant growth and innovation in recent years. Researchers and security practitioners have explored various approaches to identify sources of disinformation and understand, detect and combat the spread of false information in the digital age [5]. Existing literature reveals a multifaceted landscape, characterized by diverse methodologies and strategies [6]. Many studies have focused on the development of machine learning (ML) models that analyze textual and multimedia content to identify disinformation patterns and malicious actors [7]. Others have investigated network-based approaches, such as graph analysis, community detection [8], to unveil the intricate web of disinformation propagation, and game theoretic approaches, to model the effect of disinformation propagation on opinion dynamics [9]. Additionally, efforts have been made to incorporate user behavior and credibility assessment [10] into the analysis, recognizing the pivotal role of human interactions in disinformation dynamics [11]. Collaborative initiatives between academia, industry, and government agencies have led to the creation of large-scale datasets [12] and platforms [13] for disinformation research, fostering a rich ecosystem of tools and resources. Nevertheless, challenges persist, including the adaptability of disinformation tactics, the need for real-time detection and the ethical implications of data collection as well as surveillance [14], [15]. In this paper, we provide an in-depth review and synthesis of the current state of research in the fight against disinformation, through big data analytics, highlighting both the advancements and ongoing challenges in this critical field.

3 The Overall Architecture

In the pursuit of enhancing ML methodologies for the prediction of offline and online criminal activities associated with D&FN, FERMI has conceived and executed a novel swarm learning framework. This framework has been tailored to cater to the specific requisites of law enforcement agencies (LEAs). Notably, FERMI's framework employs a fully decentralized paradigm, as illustrated in Figure 1, which ensures adherence to extant data protection regulations and minimizes the attack surface. At the same time, it facilitates the dynamic and agile collaboration between multiple LEAs throughout Europe, since the role of a central entity will not be needed. The framework employs a permissioned blockchain network for the inclusion of nodes or agents participating in the framework and for the secure dissemination of insights. It is imperative to note that this framework accommodates popular DL frameworks such as TensorFlow, PyTorch, and Caffe, in addition to state-of-the-art NLP libraries like Hugging Face. This level of integration enables the processing of textual content from websites and social media channels. The ensuing sections provide a comprehensive examination of the three principal components of the overarching platform, along with a discussion of key challenges and results.

The platform's input is derived from posts on social media platforms (in our case Twitter) that have already been flagged as disinformation by independent fact-checking entities. Given said social media post, the platform, through crawling mechanisms, creates a graphical representation of the nodes linked to it, where each node is an additional tweet. This resultant graph, encompassing retweets and quotations of the original post, serves as the fundamental input to be further analysed by the proceeding technological components.

In order to retrieve the most influential nodes inside the graph the PageRank algorithm was selected [17]. The algorithm measures the importance of each node within the graph based on the number of incoming relationships and their importance, which may be taken into account during the analysis depending on the actual examination process. In case further tuning is deemed necessary, the overall decision can incorporate different weights for the relationships. The proposed PageRank algorithm can be mathematically articulated as follows:

$$PR(\alpha) = \frac{(1-d)}{N} + d \sum_{x \in \mathcal{N}(\alpha)} \frac{PR(x)}{C(x)} \tag{1}$$

where N is the number of nodes, $\mathcal{N}(x)$ denotes the set of neighbouring nodes with links to node α, $C(x)$ is the number of outgoing links in node α, and d is the damping factor. The contribution of $PR(x)$ from a neighboring page x is divided by $C(x)$ assuming each link has an equal chance to be selected. The damping factor d can be set to any value between 0 (inclusive) and 1 (exclusive) but is usually set to 0.85. This equation is used to iteratively update a candidate solution repeating some calculations until convergence. It is noteworthy that the user originating the initial post attains the highest PageRank score, with subsequent nodes receiving progressively lower scores. In summary, the PageRank algorithm demonstrates a remarkable efficacy in identifying nodes with the highest influence over the graph and serves as an initial mechanism for identifying the most influential graph nodes within the network.

4 Sentiment Analysis

4.1 Overview of the Technology

FERMI's Sentiment Analysis Tool emerges as a valuable asset in addressing the prevailing issue of fight against D&FN. Rooted in NLP and ML, the tool is designed to assess the emotional disposition of a social media post's author, with respect to crucial antecedents. Through doing so, the Sentiment Analysis Tool provides significant aid in facilitating the identification of potential linkages between online and offline criminal behaviours. This application assumes particular importance in the context of combating the proliferation of D&FN, given the common use of deceptive information in eliciting emotional responses from readers. By scrutinizing sentiment patterns embedded in social media content, linguistic nuances employed by individuals involved in the dissemination of D&FN can be revealed. Consequently, this analytical methodology constitutes a significant stride in the endeavour to combat criminal activities and counter-terrorism, while simultaneously addressing the

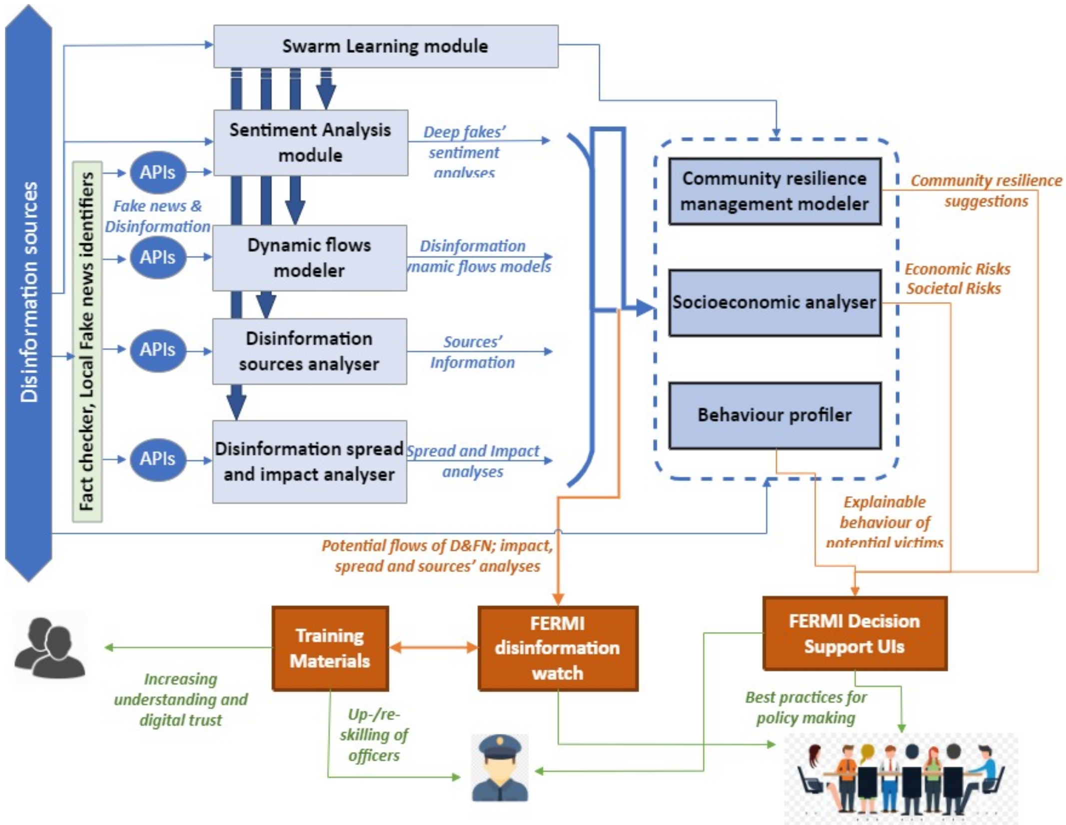

Fig. 1. The FERMI conceptual architecture with its core components (a) the sentiment analysis module, the crime predictor and the disinformation spread and impact analyser. The platform receives as input a fake news item identified by an independent authority such as fact checkers and proceeds with its analysis.

impact D&FN has on public sentiment and discourse. The tool harnesses the power of the cutting-edge bidirectional encoder representations from Transformers (BERT) language model [16] to delve into the vast realm of social media data (e.g., twitter posts data graph with highly influential nodes spreading disinformation). However, its capabilities could extend beyond just Twitter, encompassing a wide array of big data text sources such as online comments, chat logs, and witness statements, just to name a few. Through a seamless two-phase process, namely the *training phase* and the *inference phase*, the tool unveils the sentiments embedded within these textual treasures.

4.2 Training Phase and Relevant Datasets

As will be elaborated with the steps taken during the training phase of the fine-tuned model, the complex nature of social media data, particularly those related to D&FN, the proposed model evidently needed different aspects to accurately predict content's sentiment polarity. For the purposes of model training and benchmarking, the TweetEval dataset was selected as the corpus of choice. This dataset encompasses textual content extracted from tweets, categorized into three distinct sentiment classes: 0 denoting negative sentiment, 1 representing neutral sentiment and 2 indicating positive sentiment. The dataset is meticulously partitioned into three subsets: a training set comprising 45,615 data points, a validation set consisting of 2,000 data points and a testing set encompassing 12,284 data points. It is noteworthy that this dataset exhibits a high level of completeness, as there are no missing values present and the incidence of duplicate data points is exceptionally low, a mere 0.06%. A discernible class imbalance is observed within the dataset, with the neutral class demonstrating over-representation, while the negative class displays an underrepresented proportion. Within the scope of the conducted ML experiments, the issue of class imbalance in the training dataset was duly recognized and addressed. The primary objective was to evaluate the potential enhancement of the model's performance through the implementation of various strategies. Three distinct approaches have been systematically investigated and are listed below:

- **Oversampling the Underrepresented Classes**: This technique involves the generation of duplicate instances within the underrepresented classes to rectify the imbalance.
- **Weighted Loss During Training**: During the model's training process, weighted loss functions were employed. These functions assigned increased significance to the underrepresented classes, thereby mitigating the impact of class imbalance.
- **Supplementing with Additional Data**: To effectively tackle the class imbalance issue, supplementary data from the underrepresented classes was introduced. This supplementary data is sourced from an external dataset, specifically, the t4sa dataset.

In selecting the appropriate model for each step of the training phase, the following two fundamental methodologies were examined, namely (i) feature extraction, which leverages a pre-trained model as a feature extraction mechanism and (ii) fine-tuning , where architectural modifications to the model have been undertaken by appending additional layers to the pre-trained model's structure. The initial step involved the identification of approaches and architectural configurations including the base models' selection, encoding, fine-tuning and feature extraction deemed worthy of examination. Within the context of this experiment, two distinct strategies have been chosen to evaluate. The first architectural configuration involved the utilization of a pre-trained base model, either RoBERTa or BERT, augmented by a straightforward classifier positioned atop it. This top classifier underwent fine-tuning while leaving the layers of the pre-trained base model trainable. In contrast, the second selected model employed a pre-trained model as a feature extractor, with a long short-term memory (LSTM) classifier applied on top of it. In this latter architecture, token embeddings were fed into the LSTM classifier. The experimental procedure encompassed a comprehensive evaluation of both methods, followed by a comparative analysis to discern the optimal choice for the Sentiment Analysis Tool.

The base models' selection: The chosen models, considered state-of-the-art in text classification, for evaluation in the experiments include BERT, RoBERTa, DistilBERT, and ALBERT . In the conducted experiments, the emphasis was primarily placed on architectural configurations, specifically RoBERTa and BERT, which were identified as being pre-trained on sentiment analysis tasks. For the fine-tuning methodology, the chosen approach involved utilizing a pre-trained RoBERTa model initially pre-trained on the SST3 dataset and subsequently fine-tuned with the TweetEval dataset. In the case of the feature-extraction approach, an array of models was assessed, including RoBERTa pre-trained on SST3, RoBERTa base, and RoBERTa pre-trained on TweetEval, among others.

The fine-tuning method: In the fine-tuning mode, all layers of the pre-trained base model are made trainable, permitting them to engage in the learning process throughout training. Notably, the upper layers tend to specialize in the specific task being addressed. However, as the training advances, the model gradually tends to forget previously acquired knowledge. In the course of the experiments, the alternative of freezing specific layers was also explored. The optimization of hyperparameters was conducted using the training split of the TweetEval dataset, in conjunction with the validation split of TweetEval. Hyperparameter tuning was specifically tailored to optimize the average recall metric. The hyperparameters selected for tuning encompassed the Learning Rate, Weight Decay, Per Device Train Batch Size, Per Device Eval Batch Size, Num Train Epochs, and the choice of Optimizer, which is the algorithm employed to update the model's weights during the training process.

In addition to the aforementioned procedures, the number of trainable layers was also fine-tuned. Following the selection of optimal hyperparameters, an exhaustive search was conducted to identify the most suitable seed for data sampling and model initialization.

4.3 Results and Challenges

In the experimental setup, two distinct methodologies were implemented to address the challenges intrinsic to 3-class sentiment analysis. Firstly, the fine-tuning approach was adopted, involving the fine-tuning of a pre-trained RoBERTa model on SST3, using the TweetEval benchmark dataset for training. Simultaneously, an architecture was constructed following the feature extraction approach, wherein various pre-trained models acted as feature extractors, subsequently training an LSTM classifier using the TweetEval dataset. Both approaches underwent meticulous hyperparameter tuning and evaluation metric assessment on the test dataset. The results from the experiments indicated that the fine-tuning approach yielded the most promising outcomes, showcasing the highest average recall metric. In the benchmark rankings, the trained model secured the third position, with an average recall metric of 72.2% on the TweetEval test set. Subsequently, different pre-processing pipelines with this model were explored to investigate whether extensive data cleaning would result in performance improvements. Surprisingly, the outcomes indicated that actions such as the removal of stop words, punctuation marks, and lemmatization led to only marginal decreases in evaluation metrics. Additionally, the class imbalance issue was addressed within our dataset, through three distinct strategies applied to the training dataset. These strategies encompassed oversampling within the same dataset, the application of weighted loss techniques to assign higher weights to underrepresented classes, and oversampling utilizing an external t4sa dataset with categorized tweets spanning positive, negative, and neutral sentiments. Remarkably, the results highlighted that the most significant enhancement in evaluation metrics was achieved through the second approach, which involved the utilization of weighted loss techniques. This model exhibited an average recall metric of 72.4% when assessed on the TweetEval test dataset. Consequently, elevating our model to the second position in the benchmark rankings.

Twitter poses a unique set of challenges for NLP, primarily attributable to several distinctive characteristics inherent to the platform. A key challenge stems from the brevity of tweets, necessitating the use of concise and novel language specific to Twitter. This succinct mode of communication often incorporates slang and acronyms as well as being further constrained by the platform's character limit. Consequently, Twitter users develop a rapidly evolving and unique vocabulary presenting a formidable challenge for analysis. Moreover, selecting a three-class sentiment analysis approach, instead of a two-class one, introduces several technical challenges, including heightened complexity, data imbalance issues, semantic ambiguity, labeling complexities and the necessity of selecting suitable evaluation metrics. Lastly, Twitter operates as a continuous stream of dynamically generated tweets, introducing a temporal dimension. This temporal aspect necessitates the deployment of methodologies capable of handling real-time data influx. In summation, these factors collectively underscore the intricacies associated with the analysis and interpretation of content on the Twitter platform.

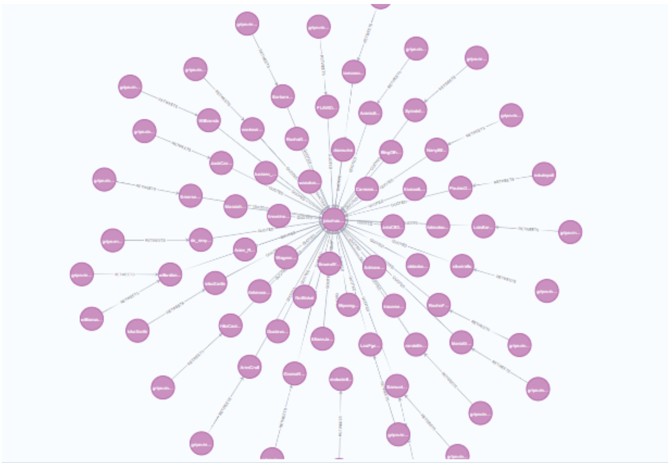

Fig. 2. The graphical representation of the nodes created from tweets crawled from the Twitter API. The graph contains all the retweets and quotes of a specific post that was provided for further analysis.

5 Crime Predictor

5.1 Introduction to the Technology

Contemporary research has turned to social media –and online content in general– as a new means of predicting offline events, from election results to stock market fluctuations [18], [19]. Employing online content to understand and predict crime is certainly an emerging trend [20]–[23], one which we advance through the construction of a DL algorithm trained on the intensity of D&FN, at a macro-level, and past crime occurrences, at the incident-level. This crime prediction mechanism, utilizing big data, NLP and DL, generates informed, accurate predictions of crimes, occurring offline, following the spread of D&FN online.

Our DL algorithm can be applied to forecast the impact of multiple disinformation topics on offline crimes. In this paper, for demonstration purposes, we will focus on COVID-19 D&FN as it exhibits well the algorithms's overall capabilities. The United States (US) were chosen as the geopolitical area to train the model due to the pertinent role of D&FN in the American political landscape, with one-third of Americans regularly encountering false content online and two-thirds believing said content cause significant levels of public confusion [24]. Moreover, the US is characterized by the high availability of both incident level crime data and D&FN datasets.

Capable of making forecasts for multiple types of crime, our DL algorithm employs multiple architectures - and ensembles of them - identifying which worked best for each type of crime. The best performing architecture for each type of crime is then selected for each crime's prediction. In other words, depending on the crime type the prediction is requested for, the DL algorithm will utilize the architecture (or ensemble of architectures) identified as being most accurate during the training phase. The results for two crime types (assault and vandalism) will be presented, with the latter predicted using transformers and the former using an ensemble of 1-dimensional convolutional neural networks (CNN) and Transformers.

5.2 Data & Pre-processing

5.2.1 *Crime Data*

Crime data referring to 11 types of crimes were collected from 31 American cities (see Fig. 3). The inclusion criteria for cities required that their local institutions publicly publish crime data at the incident level, including the date and type of each crime occurrence. The dataset had to span the entire period from 2020 to 2022, with additional data collected for 2023, if available, to provide further unseen observations for model testing. The choice of the 2020–2022 time-frame was deliberate, aiming to encompass the impact of COVID-19 and associated D&FN campaigns.

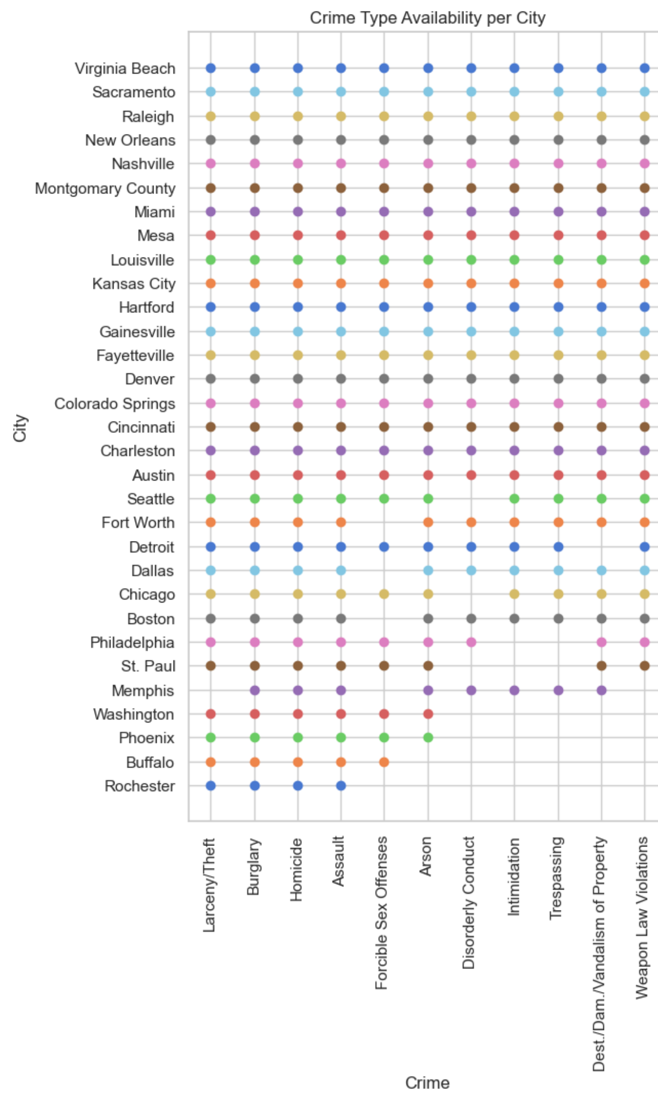

Fig. 3. Crime types by cities (data for Montgomery refers to the Maryland county, whereas data for the other 30 cities refer to the municipality.)

The data underwent standardization to align crime types with the Federal Bureau of Investigation's universal crime reporting system's categories. However, due to data privacy policies in some municipalities, not all models had access to crime instances from all 31 American cities, resulting in varying sample sizes for each set of models as seen in Figure 3. Furthermore, the study focused on 11 specific crime types deemed theoretically relevant to online disinformation spread, excluding less obviously connected crime types like gambling offenses.

In total, 3,123,893 crime incidents were included, a collection that captures the vast majority of crime in the sample cities during the years of interest. Just as well, it represents a shift away from small samples when testing predictive models, toward a big data approach that provides artificial intelligence-driven models with nearly all relevant observations as it can be readily seen in Figure 4.

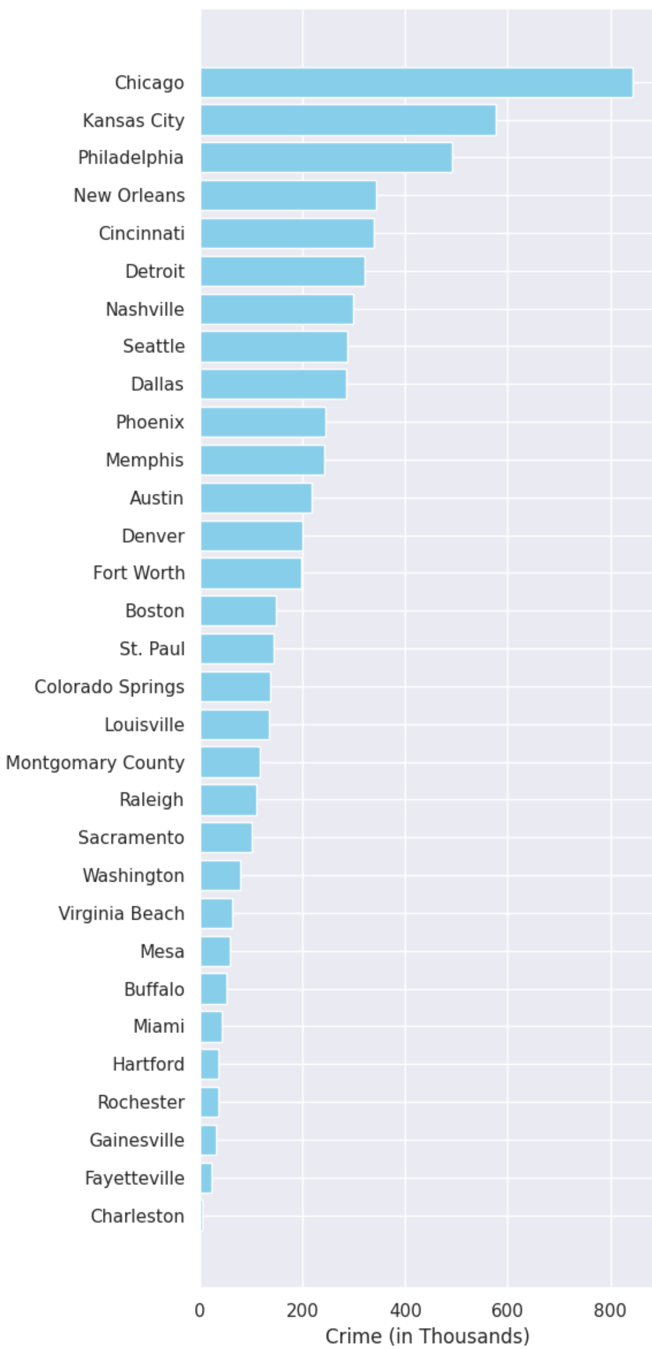

Fig. 4. Crime Count by City.

5.2.2 *Disinformation Intensity*

D&FN intensity was collected to train the device. NELA-GT datasets were chosen due to their comprehensive coverage of the years of interest and their use throughout state-of-the-art literature on disinformation [25]–[30]. In total, the NELA-GT datasets, for the years 2020, 2021 and 2022, comprised nearly 5.4 million news articles with date of publication and a label for their reliability [31]–[34]. From these datasets, daily measures of intensity for COVID-19 disinformation were extracted and transformed into a time-series through keyword matching the plain text of the articles with NELA-GT provided COVID-19 keywords.

5.2.3 *Other Predictors*

Socio-economic controls, as well as spatial mobility data, were used to ensure the models understood the commonly agreed upon contextual factors that influence changes in crime occurrence [35]–[38]. The city's population, GDP per capita, gender demographics, age structure, unemployment rate, educational attainment level, law enforcement employment per capita, and daily spatial mobility data were collected from the US Census Bureau, Department of Labor, and Google's mobility reports.

5.2.4 *Data Structure*

Various pre-processing steps were undertaken before feeding the data into the DL models. To provide a more extensive view of long-term trends and minimize the impact of daily fluctuations, the collected daily data was aggregated into weekly observations. Additionally, all continuous variables underwent a logarithmic transformation ($\log(x + 1)$), and ratio variables were standardized to fall within a range of 0 to 1.

Our crime-specific models were developed using a windowing technique, which involved dividing the data into 12-week intervals. This approach captures seasonal variations and monthly effects while emphasizing recent developments, typically within the three months leading up to the point from which we aim to predict future crimes. Each model's input matrix comprises a 12-week data sequence containing information on crime category, disinformation intensity, mobility, macroeconomic controls and city-specific details.

To capture seasonality within each input window, season and month-based dummy variables were introduced. For the model to capture diverse crime data scales, all windows were scaled to a consistent range. The windows were split into training and testing sets, with an 80% allocation for training data and a 20% allocation for testing data. Thus, the output of each model consists of 12 values, each representing the forecasted incidences of the specific crime category for the subsequent 12 weeks. This comprehensive approach ensures a more robust analysis of crime trends, while accounting for seasonal variations and monthly influences.

5.3 Deep-Learning Architecture

5.3.1 *Convolutional Neural Networks*

CNNs are effective for time-series data prediction [39], [40]. They comprise two key components: the CNN, which extracts and filters relevant features and the fully connected layer, which uses these features for predictions. Rectified linear unit activation functions introduce non-linearity in convolutional layers, and four dense layers in the fully connected part further refine features and reduce dimensionality [41].

Our CNN design features three convolutional blocks with varying numbers of filters: 500 in the first set, 250 in the second, and 128 in the third. These filters enhance the network's pattern recognition capacity. We employ rectified linear unit activation in each convolutional layer and train the model to minimize mean squared error loss. During training, the model continually changed its learning rate.

5.3.2 *Transformers*

Another, more novel, DL model for time-series forecasting, transformers, can be best characterized by the addition of self-attention mechanisms, often found in NLP [42], [43]. Through self-attention mechanisms, the model focuses on different parts of the input sequence when making predictions. For a sequence of N elements, in our case 12, denoted $X = [x_1, x_2, \ldots, x_{12}]$. These mechanisms then compute a new sequence, often referred to as the contextual or weighted sequence, denoted as $Z = [z_1, z_2, \ldots, z_{12}]$. For each variable it is provided, three sets of vectors are computed: (1) query vectors, (2) key vectors, and (3) value vectors. The first represents a given variable's importance, what the model needs to pay attention to, the second, how much other variables will affect the given one, and the third, represents the content of said variable [44]. These vectors are computed as linear transformations of the input sequence X using the learned weight matrices. For a specific variable, x_i, the query, key, and value vectors can be computed as follows:

$$\text{Query Vector: } q_i = W_q \cdot x_i \tag{2}$$

$$\text{Key Vector: } k_i = W_k \cdot x_i \tag{3}$$

$$\text{Value Vector: } v_i = W_v \cdot x_i \tag{4}$$

where W_q, W_k, and W_v are the learned weight matrices for a given variable in the self-attention mechanism. The self-attention mechanism computes attention weights for each pair of variables in the input sequence, which are computed through a similarity function, often the dot or scaled dot product:

$$\text{Attention}(q_i, k_j) = \frac{q_i \cdot k_j}{\sqrt{d_k}} \tag{5}$$

where q_i and k_j are the query and key vectors of variables x_i and x_j respectively, and d_k is the dimension of the key vectors.

After calculating the attention weights, the mechanism generates the weighted sum of the value vectors to obtain an output for each element. This weighted sum incorporates information from all the variables in the input sequence, where the importance of each is determined by the attention weights. Thus, the self-attention mechanism allows the model to focus on different parts of the input sequence when making a prediction, making it a powerful tool for capturing long-range dependencies and context in sequences relevant to accurate forecasting.

The output for a specific variable, denoted as z_i, is computed as:

$$z_i = \sum_{j=1}^{N} \text{Attention}(q_i, k_j) \cdot v_j \tag{6}$$

where q_i and k_j represent the query and key vectors of variables x_i and x_j, respectively, and v_j represents the value vector for variable x_j.

Our architecture combines convolutional layers for feature extraction with transformer layers, for capturing temporal dependencies in time series data. Transformer layers are stacked to enhance feature representation. Dropout and layer normalization improve model robustness, followed by dense layers leading to the output layer. The model is trained to minimize mean squared error loss and use mean absolute error for evaluation. This design exhibits the use of transformers in time-series forecasting, as they enable precise predictions through appreciating nuanced temporal patterns.

5.3.3 *Ensemble*

To leverage the strengths of both the CNN and transformers' architecture, an ensemble method was applied. An ensemble takes advantage of the strengths and capacity possessed by each independent architecture [45], [46], using either a voting between or averaging of outputs to produce a final output. For our device, averaging was chosen as the ideal ensemble method. The average function takes the element-wise average of the predictions produced by the models. The ensemble model is then trained to minimize the mean squared error loss, and mean absolute error is used for evaluation, as is the case for the transformers model.

$$\text{Ensemble Pred.} = \frac{\text{CNN Pred.} + \text{Transformers Pred.}}{2} \tag{7}$$

5.4 Results

After being trained on the complete time-series data, the model underwent testing using a randomly selected 24-week window. This window was divided into two distinct periods: the first 12 weeks served as the "seen" data, while the subsequent 12 weeks were labeled as the "unseen" data. The "seen" data was subsequently input into the deep-learning model that had exhibited the best performance for the specific crime type under consideration. As a result, this model generated 12 predictions, each corresponding to one of the "unseen" weeks. In this section, we present the results for assault and vandalism, crimes with plausible connections to the dissemination of disinformation online [47]. Figure 5 showcases an example of the 12 "unseen" weeks, illustrating both the actual values for assault (above) and vandalism (below), as well as the corresponding predicted values.

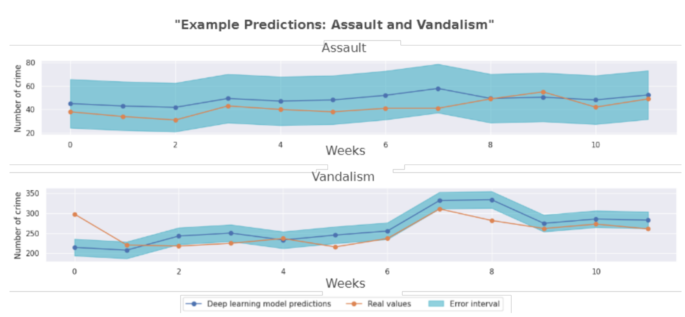

Fig. 5. Twelve-weeks forecasts for assault and vandalism.

The most suitable model for assault incorporated an ensemble of transformers and CNN, which yielded a mean absolute error (MAE) of 21.9. In contrast, the optimal model for vandalism exclusively relied on the transformers' architecture, resulting in a MAE of 20.9. In order to gauge the efficacy of our model, we introduced an autoregressive integrated moving average (ARIMA) ML model to produce predictions for the same time frame, utilizing historical crime data. Interestingly, the ARIMA model demonstrated higher MAEs not only for assault and vandalism but also across all 11 crime types (Figure 6).

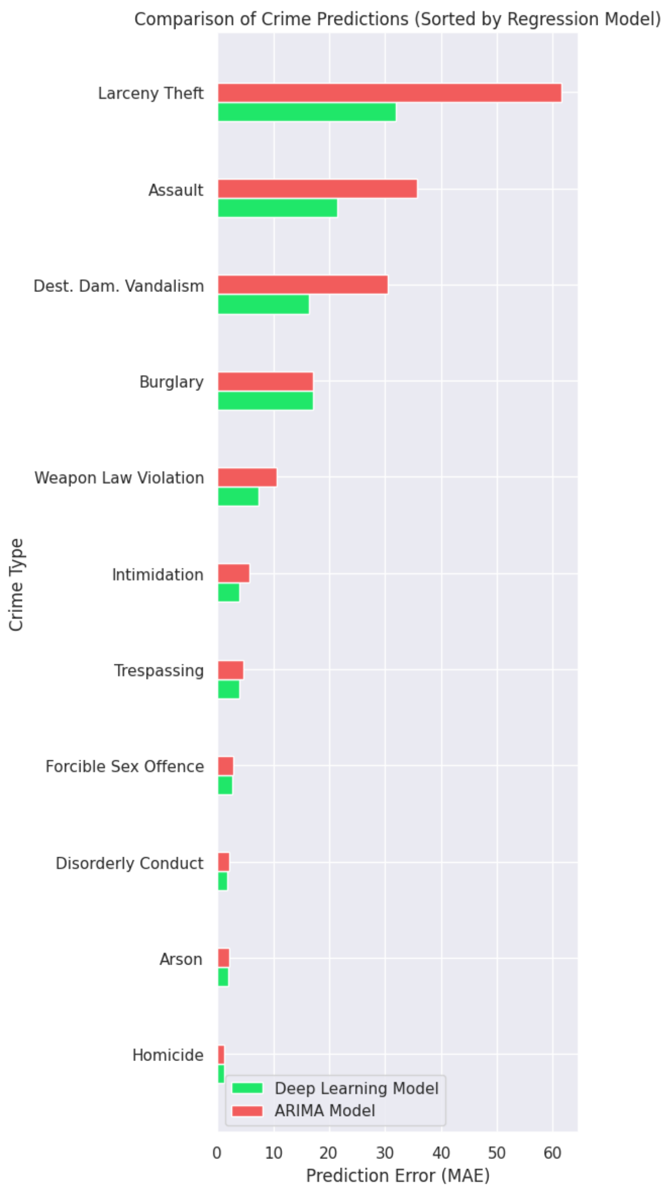

Fig. 6. Mean absolute error across crime types,

6 Conclusions

In this paper, a comprehensive framework has been introduced, aimed at combating the issue of disinformation spread, particularly as it relates to offline and online criminal activities, leveraging the capabilities of big data analytics. The focus is squarely on understanding the intricate dynamics of disinformation propagation and its potential real-world consequences. The analysis conducted in this study sheds light on the complex nature of disinformation campaigns, highlighting the critical role that interdisciplinary approaches play in comprehending and mitigating this problem. Through the integration of advanced ML techniques and network analysis our framework stands as a potent resource for LEAs in their quest to identify disinformation patterns and anticipate criminal activities stemming from D&FN. However, it is manifest that the battle against D&FN remains a dynamic and continuously evolving challenge. Subsequent research endeavors must be undertaken to refine and adapt these methodologies in order to stay ahead of the ever-evolving strategies employed by disinformation campaigns. Moreover, it is imperative to underscore the significance of ethical considerations surrounding data privacy and algorithmic bias must remain at the forefront of technological developments in this domain, along with the emerging issue of limited access to data from major social media platforms necessitates further exploration and resolution in order to maintain the efficacy of disinformation mitigation strategies.

References

[1] David Blanco-Herrero and Carlos Arcila Calderón, ”Spread and reception of fake news promoting hate speech against migrants and refugees in social media: Research Plan for the Doctoral Programme Education in the Knowledge Society,” in Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM'19), Association for Computing Machinery, New York, NY, USA, 949–955.

[2] E. Huber, B. Pospisil and W. Haidegger, ”Modus Operandi in Fake News : Invited Paper,” 2021 IEEE Conference on Cognitive and Computational Aspects of Situation Management (CogSIMA), Tallinn, Estonia, 2021, pp. 127-132.

[3] J. F. DeFranco, N. Kshetri, R. Sharma and D. Rojas-Torres, ”Mitigating Disinformation and Building Trust in Social Media,” in IEEE IT Professional, vol. 23, no. 6, pp. 62-66, 1 Nov.-Dec. 2021

[4] W. Ahmed *et al.*, ”Developing Techniques to Support Technological Solutions to Disinformation by Analyzing Four Conspiracy Networks During COVID-19,” in IEEE Transactions on Engineering Management, vol. 70, pp 1-18, 2023.

[5] K. Demestichas, K. Remoundou and E. Adamopoulou, ”Food for Thought: Fighting Fake News and Online Disinformation,” in IEEE IT Professional, vol. 22, no. 2, pp. 28-34, 1 March-April 2020.

[6] A. Altheneyan and A. Alhadlaq, ”Big Data ML-Based Fake News Detection Using Distributed Learning,” in IEEE Access, vol. 11, pp. 29447-29463, 2023.

[7] R. Kozik, *et al.*, ”A Meta-Analysis of State-of-the-Art Automated Fake News Detection Methods,” in IEEE Transactions on Computational Social Systems, vol. 10, pp. 1-10, 2023.

[8] B. Rath, A. Salecha and J. Srivastava, ”Detecting Fake News Spreaders in Social Networks using Inductive Representation Learning,” 2020 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM), The Hague, Netherlands, 2020, pp. 182-189.

[9] Z. Guo, J. Valinejad and J. -H. Cho, ”Effect of Disinformation Propagation on Opinion Dynamics: A Game Theoretic Approach,” in IEEE Transactions on Network Science and Engineering, vol. 9, no. 5, pp. 3775-3790, 1 Sept.-Oct. 2022.

[10] Md Momen Bhuiyan, Amy X. Zhang, Connie Moon Sehat, and Tanushree Mitra, ”Investigating Differences in Crowdsourced News Credibility Assessment: Raters, Tasks, and Expert Criteria,” in Proc. ACM Hum.-Comput. Interact. 4, CSCW2, Article 93, October 2020.

[11] A. Gupta *et al.*, ”Combating Fake News: Stakeholder Interventions and Potential Solutions,” in IEEE Access, vol. 10, pp. 78268-78289, 2022.

[12] S. S. Reddy, *et al.*, ”A Novel Approach to Detect Fake News Using eXtreme Gradient Boosting,” 2022 10th International Symposium on Digital Forensics and Security (ISDFS), Istanbul, Turkey, 2022, pp. 1-4.
[13] G. Ramachandran *et al.*, ”WhistleBlower: Towards A Decentralized and Open Platform for Spotting Fake News,” 2020 IEEE International Conference on Blockchain (Blockchain), Rhodes, Greece, 2020, pp. 154-161.
[14] Xinyi Zhou, Reza Zafarani, Kai Shu, and Huan Liu. 2019. Fake News: Fundamental Theories, Detection Strategies and Challenges. In Proceedings of the Twelfth ACM International Conference on Web Search and Data Mining (WSDM ’19). Association for Computing Machinery, New York, NY, USA, 836–837.
[15] Xinyi Zhou and Reza Zafarani, ”A Survey of Fake News: Fundamental Theories, Detection Methods, and Opportunities,” ACM Comput. Surv., Vol. 53, no. 5, Article 109, September 2021.
[16] M. Heidari *et al.*, ”BERT Model for Fake News Detection Based on Social Bot Activities in the COVID-19 Pandemic,” 2021 IEEE 12th Annual Ubiquitous Computing, Electronics & Mobile Communication Conference (UEMCON), New York, NY, USA, 2021, pp. 0103-0109.
[17] F. Chung, ”A Brief Survey of PageRank Algorithms,” in IEEE Transactions on Network Science and Engineering, vol. 1, no. 1, pp. 38-42, 1 Jan.-June 2014.
[18] A. Bermingham and A. F. Smeaton, ”On using Twitter to Monitor Political Sentiment and Predict Election Results,” in International Joint Conference for Natural Language Processing, Chiang Mai, Thailand, 2011.
[19] J. Bollen, H. Mao, and X. Zeng, ”Twitter Mood Predicts the Stock Market,” Journal of Computational Science, vol. 2, no. 1, 2011.
[20] M. Chen, Q. Xu, A. Zeng, and L. Zhang, ”Are Transformers Effective for Time Series Forecasting?” arXiv preprint, arXiv:2205.13504, 2022.
[21] K. Muller and C. Schwarz, ”Fanning the Flames of Hate: Social Media and Hate Crime,” Journal of European Economic Association, vol. 19, no. 4, 2021, pp. 2131–2167.
[22] X. Wang, M. S. Gerber, and D. E. Brown, ”Automatic Crime Prediction Using Events Extracted from Twitter Posts,” in Social Computing, Behavioral – Cultural Modeling and Prediction, vol. 7227, 2012, pp. 231–238. Springer: Berlin, Heidelberg.
[23] M. L. Williams and P. Burnap, ”Cyberhate on Social Media in the Aftermath of Woolwich: a Case Study in Computational Criminology and Big Data,” British Journal of Criminology, vol. 56, no. 2, 2016, pp. 211–238.
[24] M. Barthel, J. Holcomb, and A. Mitchell, ”Many Americans Believe Fake News is Sowing Confusion,” Pew Research Center, 2016.
[25] M. A. Alonso, C. Gomez-Rodriguez, D. Vilares, and J. Vilares, ”Sentiment Analysis for Fake News Detection,” Electronics, vol. 10, no. 11, 2021.
[26] M. Amini, Y. Akbari, J. A. Godarzi, and K. Sharifani, ”Operating Machine Learning Across Natural Language Processing Techniques for Improvement of Fabricating News Models,” International Journal of Science System Research, vol. 12, no. 9, 2022, pp. 25–44.
[27] [1] R. Baly, G. Da San Martino, J. Glass, and P. Nakov, ”We Can Detect Your Bias: Predicting the Political Ideology of News Articles,” in Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), Nov. 2020, pp. 4982-4991.
[28] E. Choi, Y. Choi, S. Gabriel, S. Hallinan, P. Nguyen, F. Roesner, and M. Sap, ”Misinfo Reaction Frames: Reasoning about Readers’ Reactions to News Headlines,” arXiv preprint, arXiv:2104.08790, 2021.
[29] A. Das, V. Kovatchev, M. Lease, and H. Liu, ”The State of Human-Centred NLP technology for Fact-Checking,” Information Processing & Management, vol. 60, no. 2, 2023.
[30] C. Ding and S. Raza, ”Fake News Detection Based on News Content and Social Contexts: a Transformer-Based Approach,” International Journal of Data Science and Analytics, vol. 13, 2022, pp. 335–362.
[31] M. Gruppi, S. Adali, and B. D. Horne, ”NELA-GT-2019: a Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles,” arXiv preprint, arXiv:2003.08444, 2020.
[32] M. Gruppi, S. Adali, and B. D. Horne, ”NELA-GT-2020: a Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles,” arXiv preprint, arXiv:2102.04567, 2021.
[33] M. Gruppi, S. Adali, and B. D. Horne, ”NELA-GT-2021: a Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles,” arXiv preprint, arXiv:2203.05659, 2022.
[34] M. Gruppi, S. Adali, and B. D. Horne, ”NELA-GT-2022: a Large Multi-Labelled News Dataset for the Study of Misinformation in News Articles,” arXiv preprint, arXiv:2203.05659, 2023.
[35] P. J. Brantingham and P. L. Brantingham, Environmental Criminology. Beverly Hills, CA: Sage Publications, 1981.
[36] I. Ehrlich, ”On the Relationship between Education and Crime,” in Education, Income, and Human Behavior, 1975, pp. 313–338. NBER: Cambridge.
[37] R. Freeman, ”The Economics of Crime,” in Handbook of Labor Economics, vol. 3, 1999, pp. 3529–3571. Elsevier: Amsterdam.
[38] C. R. Hermann, A. R. Maroko, and T. A. Taniguchi, ”Subway Station Closures and Robbery Hot Spots in New York City – Understanding Mobility Factors and Crime Reduction,” European Journal of Criminal Policy and Research, vol. 27, no. 3, 2017, pp. 415–432.
[39] S. Guessoum et al., ”The Short-Term Prediction of Length of Day Using 1D Convolutional Neural Networks (1D CNN),” Sensors, vol. 22, no. 23, 2022, pp. 9517.
[40] W. Li, Y. Li, W. Liu, L. Meng, and Y. Xing, ”Operation Energy Consumption Estimation Method of Electric Bus Based on CNN Time Series Prediction,” Mathematical Problems in Engineering, vol. 2022, 2022.
[41] S. Kiranyaz et al., ”1D Convolutional Neural Networks and Applications: A Survey,” Mechanical Systems and Signal Processing, vol. 151, 2021, p. 107398.
[42] X. Chen, Y. Cho, and S. Y. Jang, ”Crime Prediction using Twitter Sentiment and Weather,” 2015 Systems and Information Engineering Design Symposium, Charlottesville, United States of America, 2015, pp. 63–68.
[43] K. Rasul, N. Rogge, and E. Simhayev, ”Yes, Transformers are Effective for Time Series Forecasting (+ Autoformer),” Hugging Face, 2023.
[44] A. N. Gomez et al., ”Attention is all You Need,” arXiv preprint, arXiv:1706.03762, 2017.
[45] A. F. Atiya, ”Why does Forecast Combination Work So Well?” International Journal of Forecasting, vol. 36, no. 1, 2020, pp. 197–200.
[46] V. Murali, ”Everything you Need to Know about Ensemble Learning,” Medium, 2021.
[47] K. Muller and C. Schwarz, ”Fanning the Flames of Hate: Social Media and Hate Crime,” Journal of European Economic Association, vol. 19, no. 4, 2021, pp. 2131–2167.