# From News to Knowledge: Predicting Hate Crime Trends through Event Extraction from Media Content

Jiangwei Liu, Xiangzhen Jia, You Wu, Jingshu Zhang, and Xiaohong Huang

*Abstract*—**Social media platforms have emerged as fertile ground for the proliferation of hate speech, which can exacerbate the dissemination of hate crimes. The Federal Bureau of Investigation UCR Program gathers data on hate crimes and disseminates annual reports to identify national patterns and inform law enforcement agencies and policymakers, these reports often fail to keep pace with urgent demands. Real-time monitoring and predictive analysis of hate crime trends are imperative for more effective prevention and response efforts. This paper presents a framework that leverages information extraction techniques to extract incidents from articles published in The New York Times, enabling accurate prediction of hate crime trends at both the federal and state levels. Experimental findings demonstrate the superiority of our approach compared to other traditional methods. By expanding forecasting approaches for federal and state levels' hate crime trends, this framework offers valuable insights for law enforcement agencies and policymakers.**

*Index Terms*—**hate crime prediction, event extraction, media content analysis, text mining, predictive modeling**

## I. Introduction

THe prevalence of social media has facilitated individuals in freely expressing their thoughts and disseminating information through consumer-generated content (CGC) on various platforms, such as Facebook, blogs, and online forums. While the majority of CGC interactions are advantageous, there has been a surge in the utilization of derogatory and offensive language on digital platforms. The rapid escalation of divisive rhetoric in online platforms can give rise to localized surges and potentially instigate acts of hatred. For instance, during the political campaign, Donald Trump's inflammatory discourse exacerbated an escalation in hate crimes targeting Muslims and immigrants. The report titled 'Trump effect' led to hate crime surge can be accessed at the following URL: https://www.bbc.com/news/world-us-canada-38149406. Karsten Müller and Carlo Schwarz conducted a study investigating the potential amplification of antiminority sentiment through social media, with a particular emphasis on the political ascent of Donald Trump. The findings indicate that the content on social media platforms has the potential to influence real-life outcomes [1].

According to the FBI's UCR Program, a hate crime is classified as a criminal act driven by the perpetrator's prejudice against a specific race, ethnicity, gender, sexual orientation, disability, religion, or gender identity. For additional information, please visit the FBI's website. This paper presents a comprehensive analysis of the existing literature on hate crime research, delineating two primary domains: sociology (including econometrics) and computer science.

Sociology primarily focuses on comprehending the underlying causes and repercussions of hate crimes, while computer science concentrates on detecting and preventing instances of hate speech and cybercrimes. Although studies have explored associations between hate crimes and factors such as depression, impulsivity, or post-traumatic stress disorder (PTSD), most have predominantly focused on analyzing only one specific type of hate crime. For instance, Robert J. Cramer et al. investigated the link between depressionivity or PTSD with victimization among sexual orientation minority adults [2]. This study laid the risks faced by sexual orientation minorities and suicide prevention efforts. Karsten Müller and Carlo Schwarz use case study method to investigate the relationship between media consumption and violence. The findings presented in their research underscore the significant role that social media plays in the propagation of violence targeted towards minority groups [1]. However, limited research exists regarding predicting trends in local or national levels of hate crimes.

Conversely, within computer science-based research, various methods have been developed to detect hate speech, including keyword-driven approaches and advanced deep learning models. Njagi Dennis Gitari et al. constructed a lexicon to identify hateful language in online discussions using specific features [3] while Mozafari Farakhsh et al. proposed a neural network based on a pre-trained language model (BERT), to identify offensive content [4]. However, the application of computer science technology to forecast hate crime patterns at local or national levels remains largely unexplored.

Hate crime is a significant concern with profound impacts on individuals and communities. While previous studies have primarily focused on analyzing and detecting hate crimes, there is a need to enhance the accuracy of predicting hate

Manuscript received July 24, 2023; revised February 10, 2024.

This work was supported in part by Research Project on Industry-Education Integration of the Henan Provincial Department of Education for the year 2023.

Jiangwei Liu is an assistant professor in the School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046, China (e-mail: majorliujw@gmail.com).

Xiangzhen Jia is an undergraduate student in the School of Computer and Information Engineering, Henan University of Economics and Law, Zhengzhou 450046, China (e-mail: jiaxiangzhen666@gmail.com).

You Wu is a PhD candidate in the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: wuy1997@163.com).

Jingshu Zhang is a PhD candidate in the School of Information Management and Engineering, Shanghai University of Finance and Economics, Shanghai 200433, China (e-mail: zhangjingshu@163.sufe.edu.cn).

Xiaohong Huang is an assistant professor in the School of Logistics and Management Engineering, Yunnan University of Finance and Economics, Kunming 650221, China (e-mail: huangxiaohong@163.sufe.edu.cn).

crime trends in order to mitigate their detrimental effects [5], [6]. This paper aims to address this issue by proposing an innovative approach for forecasting hate crime trends. Firstly, we provide a concise overview of the existing literature on hate crimes and identify areas that require further investigation. Next, we describe our data sources and processing methods, which involve utilizing deep learning techniques to conduct event extraction. Subsequently, we propose a framework that incorporates factors related to these incidents into econometric models for predicting hate crime trends. Our experimental findings demonstrate the significant enhancement in model performance achieved by this approach.

The described process is as follows: In contrast to previous studies, our investigation focuses on conducting event extraction from news report content. Subsequently, we identify and integrate three predictive factors related to events into our framework. Our research draws inspiration from a prior study [7]. We employ this event extraction strategy by training our model on the Patch corpus and applying it to the New York Times corpus for identifying instances of hate crime. By utilizing time series and regression models as our foundational models, we estimate parameters through Least Square (LS) and Maximum Likelihood Estimation (MLE) methods, and forecast hate crime by incorporating event-related factors into our regression model, thereby showcasing a significant enhancement in model performance resulting from these factors. Our findings have significant implications for policymakers and law enforcement agencies, enabling them to develop evidence-based strategies aimed at preventing and addressing hate crimes.

The subsequent sections of this paper are organized as follows: Section II offers a thorough review of the pertinent literature, while Section III offers an overview of the dataset and the methods employed for data processing. Section IV introduces our framework and its constituent modules. In the subsequent part, Section V showcases the experimental results, examines limitations, outlines future work, and Section VI concludes the paper.

## II. Related Work

### A. Hate Speech Detection

The identification and detection of hate speech present a critical challenge for social media platforms, despite their establishment of guidelines and rules to identify and remove such content [8], [9]. Given the substantial volume of posts and responses, automated systems are indispensable for this task. The existing literature related to the detection of hate speech using automated methods can be categorized into two primary domains: approaches based on keyword analysis and approaches based on machine learning techniques.

The keyword based techniques rely on dictionaries or databases of speech expressions to identify posts that contain hateful phrases [3]. These unsupervised methods are effective and straightforward to implement; however, they do have limitations. For instance, they may not be capable of recognizing intricate expressions or contents that include hate speech but do not employ specific terms from the dictionary.

The prevalence of machine learning approaches has made them the dominant methodology for automatic hate speech identification [10], [11]. These approaches typically employ text classification algorithms to address the problem of identifying hate speech.These approaches typically employ text classification algorithms to address the problem of identifying hate speech. By annotating a corpus to indicate the presence or absence of hate speech, conventional machine learning algorithms for text classification can be employed. For instance, Mozafari Farakhsh et al. used transfer learning techniques to automatically identify hate speech in online media text [4].

### B. Event Extraction

The task of Event Extraction (EE) involves the identification of textual content and relations between entities, often following several preliminary NLP processes, rendering it a sophisticated form of Information Extraction (IE) [12]. Extracting events from news content serves as a preliminary step for numerous tasks, such as news summarization generation and news aggregation [13]. Pattern recognition and machine learning methodologies are commonly employed in the realm of closed-domain event extraction.

Pattern matching methods rely on predefined patterns or schemas, making them less data-intensive but challenging to define and maintain. For instance, Xu and Liu put forth a hybrid collaborative filtering methodology designed for event extraction and recommendation in event-centric social networks, demonstrating high accuracy rates in their experimental evaluations [14]. Hamborg et al. enhanced the universal event extraction model, Giveme5W1H, to identify key events in an article [13]. This method utilizes syntactic parsing techniques and specific regulations to automatically extract pertinent expressions that capture the essential elements of the event.

The advantage of machine learning and deep learning techniques lies in their capacity to autonomously and efficiently extract salient features from textual data [15]. For example, Davani and his colleagues employed a deep learning model for event extraction, specifically identifying hate crime instances in news content [7]. They conducted an analysis of experimental results, providing a lower-bound estimate of hate crime incidence, particularly in cities not covered by the FBI data. Building upon this approach, our study adopts a deep learning-based event extraction method to identify hate crime events, aiming to provide more precise and comprehensive predictions of hate crime trends.

### C. Empirical Study of Hate Crimes

This empirical study provides a comprehensive analysis of hate crimes, revealing their patterns, impacts, and strategies for addressing them. A multitude of literature focuses on the impact of specific hate crime events [16]. To illustrate, Herek et al. conducted a study examining the psychological impact of being targeted by hate attacks on individuals with diverse sexual orientations [17]. The findings revealed that individuals who survived hate crimes targeting their sexual orientation experienced significantly higher levels of depression, hostility, worry, and post-traumatic stress symptoms. These results emphasize the crucial importance of recognizing and addressing the unique needs of hate crime survivors within clinical practice and public governance.

One recent instance that stands out is the surge in racial animosity directed towards Asian Americans during the Covid-19 pandemic in America. Lu and Sheng conducted a study utilizing Google search data and Twitter posts containing instances of hate speech targeting Asians to assess the impact of the global health crisis on racial animosity. [18]. Their conclusion highlights that diminishing the focus on connecting the disease to a particular ethnic group or race can significantly alleviate racial animosity. These results emphasize the necessity of implementing targeted interventions to address hate speech and hate crimes in the context of public health crises.

## III. Data Introduction and Processing

### A. Corpus for Event Extractor

Mostafazadeh Davani and his colleagues gathered news articles and annotated whether a news report involved hate crimes [7]. We use this dataset to train our event extraction model. The corpus was annotated with two labels: one indicating the presence of a hate crime in the news story, and the other specifying its corresponding category based on hate crime characteristics. The corpus consists of a total of 5,171 samples, including 3,192 negative samples and 1,979 positive samples.This corpus can be obtained from https://github.com/aiida-/HateCrime.

### B. Dataset for Extracting Event Factors

Historical news reports usually can be acquired from large news websites. The New York Times provides an API that allows users to acquire monthly historical news stories, including titles, abstracts, keywords, URLs, and publication dates. The API can be accessed at https://developer.nytimes.com/apis. However, the API does not provide the full content of the news articles. Therefore, we employed web crawler technology to collect the news content associated with each URL from January 2007 to December 2020.

After filtering out irrelevant categories and news reports lacking substantial information, a total of 165,913 relevant news reports remained. Subsequently, we categorized these reports into quarters and calculated their overall count in each group as a time series denoted as $news\_num$, averaging at 2,962 reports per quarter. We then applied the trained event extraction model to assess whether a given news report describes an event of hate crime. The resulting predictions were tallied and named as $event\_detect\_num$.

### C. FBI's Hate Crime Statistics

In our study, we collected a comprehensive dataset of hate crimes reported to FBI in America spanning the period from 2007 to 2019. This hate crime statistics encompass a wide range of valuable information, including the frequency of hate crime incidents categorized by bias motivation and quarter. The released statistics can be obtained at FBI's UCR official website https://cde.ucr.cjis.gov. These statistics are further broken down by state, federal agency, and local law enforcement entities. Moreover, it also includes information about the population of each respective state, federal agency, or locality. To gain deeper insights into hate crime trends, we recommend analyzing both quarterly statistics and annual hate crime figures. Consequently, we plotted the original time series data using a yearly cycle divided into four quarters. Figure 1 offers a visual representation of the fluctuations and patterns observed in hate crime incidents over the study period.

The cyclical patterns observed in the quarterly hate crime time series, as depicted in Figure 1, suggest the presence of a potential seasonal trend. In order to assess the stationarity of the time series, we employ the Augmented Dickey-Fuller Test (ADF test). The statistical results indicate the presence of a unit root, indicating that the time series is indeed nonstationary. To extract meaningful insights from this time series, we utilize the moving average approach to decompose it into its constituent components: trend, seasonal, and irregular. The decomposition is illustrated in Figure 2, providing a visual representation of the individual components and their contributions to the overall time series.

We conduct Ljung-Box Tests on the irregular components of the decomposed time series and observe a significant $p\_value$ of 0.005. This result suggests that the irregular components exhibit characteristics of white noise, indicating a satisfactory decomposition. Subsequently, we proceed to construct models for the trended data. To achieve this, we estimate the model parameters using the data spanning from 2007 Q1 to 2018 Q4. By utilizing the estimated model, we are able to make forecasts for hate crime trends beyond the fourth quarter of 2018.

## IV. Methodology

Our predictive model incorporates two distinct types of models: time series models and regression models. The conceptual framework of the entire model is depicted in Figure 3. A time series model can be utilized as a standalone approach to directly forecast hate trends or serve as a baseline model for comparison and integration into a hybrid framework. A regression model can utilize multiple predictive factors to directly forecast hate trends, but the inclusion of event factors in the regression model leads to better performance. This improvement is primarily attributed to the event features extracted by the event extraction model, which plays a crucial role in our framework by autonomously extracting event types and attributes from news reports. These extracted elements are then utilized to construct event-related factors, which capture the contextual information surrounding hate crime incidents. Finally, This integration is achieved through the utilization of regressive methods. By incorporating these diverse factors, we aim to enhance the accuracy of our predictions for FBI hate crime trends.

With the intention of offering a comprehensive grasp of our approach, we have organized this section into three distinct subsections: "Time Series Module", "Event Extraction Module", and "Regression Module". Each subsection delves into the detailed explanation of the respective component, elucidating their functionalities and illustrating how they synergistically integrate to predict hate crime trends.

### A. Time Series Module

Currently, many modeling methods are only applicable to stationary time series. Therefore, it is essential to transform

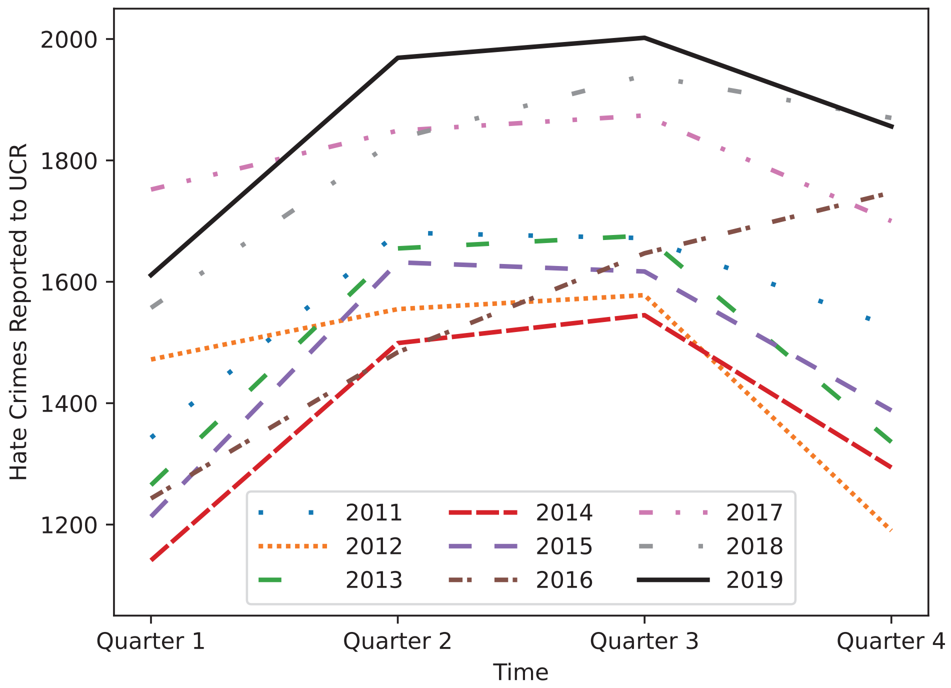


Fig. 1. The number of hate crimes released by FBI is displayed annually in a collapsed format.

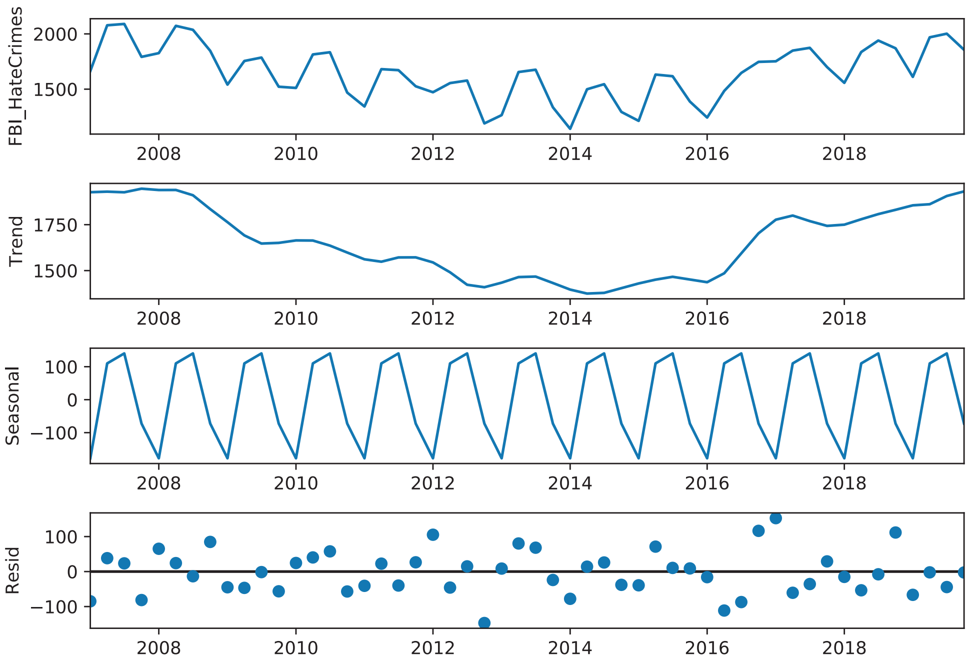


Fig. 2. Quarterly hate crime time series and its constituent components: trend, seasonal, and irregular.

the time series into a stationary form before employing these methods. Differencing and logarithmic transformation are two commonly used methods to convert time series into stationary series. Autoregressive Integrated Moving Average (ARIMA) is a commonly used method for modeling stationary time series [19]. In the ARIMA model, a nonseasonal time series is denoted as $ARIMA(p, d, q)$, where the integer parameters have specific roles. The parameter $p$ denotes the autoregressive (AR) component's order, $d$ signifies the degree of differencing necessary to transform the non-stationary time series into a stationary one, and $q$ represents the moving average (MA) component's order.

For instance, given a stationary and invertible $ARMA(p, q)$ model, if a time series $y_t$ meets a transformation $c_t = \Delta y_t = y_t - y_{t-1} = (1 - L)y_t$, then $y_t$ is considered an $ARIMA(p, 1, q)$ process. Here, $L$ shifts the observed values of a time series backward by one unit to obtain previous observations.

The general formulation of an $ARMA(p, q)$ model is expressed as follows:

$$c_t = c + \sum_{i=1}^{p} \alpha_i c_{t-i} + \varepsilon_t + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} \tag{1}$$

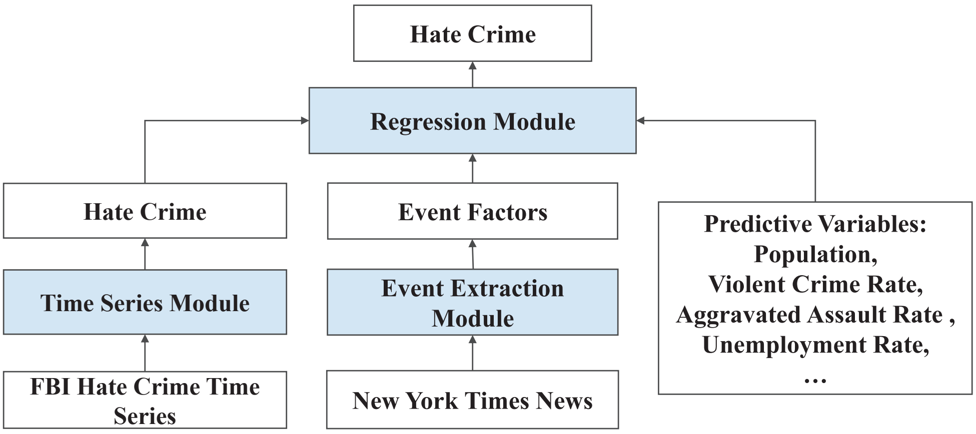


Fig. 3. Hate crime prediction framework.

Here, $\varepsilon_t$ represents a white noise process with variance $\sigma^2$, which signifies the random and uncorrelated errors in the model. The term $c$ represents the intercept term, accounting for the constant component in the time series. The coefficients $\alpha_i$ and $\theta_i$ correspond to the AR and MA processes, respectively. The AR process establishes a dependency connection between the present value of a time series and its preceding values, while the MA process accounts for the influence of past error terms on the current value.

In our study, we initially evaluate the need for seasonal adjustment in a time series to eliminate cyclical or periodic patterns. To achieve this, we utilize R programming language functions to decompose the hate crime series $fbi\_num$ into trend components, seasonal effect component, and residual component.

After performing the decomposition, our primary objective is to model the trend component $fbi\_num\_noseas$. By eliminating the seasonality and capturing the underlying trend, we aim to create a time series that exhibits a constant mean and variance over time. By utilizing the ARIMA model, we can effectively model the stationary behavior of the transformed trend component and gain insights into the underlying patterns and trends in the hate crime data.

### B. Event Extraction Module

The Event Extraction Module assumes a pivotal role within our hate crime prediction framework, as it undertakes the critical task of extracting event features from news reports. These extracted elements are subsequently utilized to construct event-related predictive factors, which are essential for enhancing the performance of hate crime trends prediction.

In pursuit of this objective, we leverage prior research [7] and customize the Multi-Instance Learning technique [20] to suit our requirements. This method has proven to be effective in extracting hate crime events from textual sources. The Event Extraction Module consists of three primary components, as illustrated in Figure 4. To clearly delineate the individual components, we have highlighted their separation using red dashed lines.

The first component of our framework focuses on generating sentence embeddings that capture the local features of an article. To capture the sequential information inherent in the input words, we employ a Bidirectional LSTM layer. This layer effectively considers both past and future contexts when generating the embeddings, enhancing the representation of the local features [21], [22].

In the upper left region, as depicted in Figure 4, Convolutional Neural Network (CNN) and Pooling layers are employed to capture the features of a specific news context. This fusion enables the scanning of sentence embeddings, producing a condensed representation that captures the comprehensive characteristics of the article. Specifically, the CNN layer is adept at capturing local patterns and features, while the Pooling layer aggregates the information to create a more condensed and representative global feature embedding [23].

The third component takes on the responsibility of extracting hate crime events, including both event types and attributes. We utilize classification algorithms to tackle these information extraction challenges. The task of differentiating between hate crime events and non-hate crime events is framed as a binary classification problem. In contrast, we tackle the identification of event attributes as a task of classifying multiple labels, enabling the assignment of multiple attributes to a given hate crime event.

We train two separate models using the Patch Hate Crime dataset to extract event-related factors. These event-related factors are then integrated into the Regression Module, which analyzes the relationship between these factors and FBI hate crime trends. This approach enables us to make accurate predictions and forecasts regarding future hate crime trends based on the extracted information.

### C. Regression Module

Previous research has indicated that hate crimes may be associated with a range of criminological and social-psychological theories, such as Strain Theory Criminology and Criminological Theories. In our study, we aim to expand upon these existing theories by considering additional, explainable factors that have not been extensively explored before.

To achieve this, we begin by conducting data exploration and performing collinearity tests to ensure the dependability and accuracy of the predictors. Through this process, we identify 14 alternative predictors that demonstrate acceptable levels of collinearity and are deemed relevant to hate crimes. These 14 alternative predictors, which encompass a wide

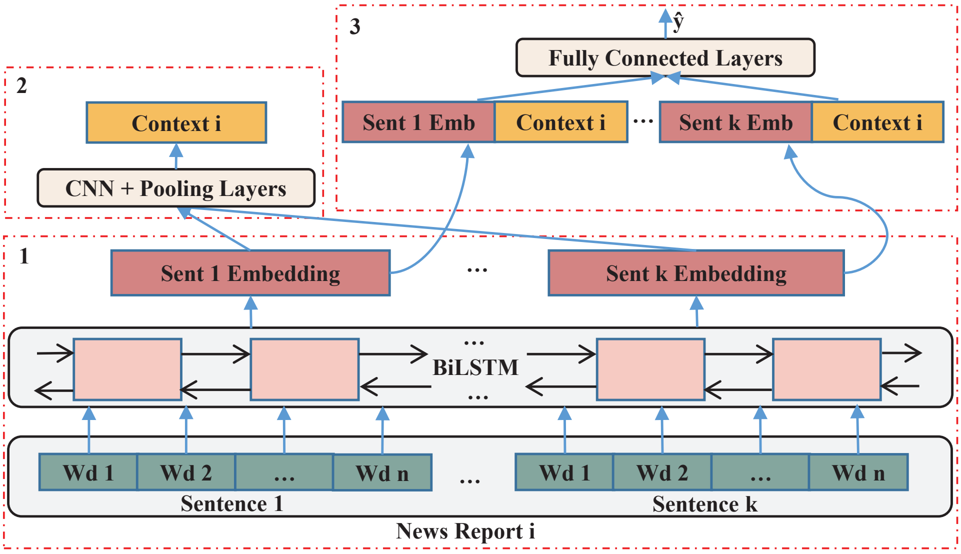


Fig. 4. Design and structure of Event Extraction Module.

range of factors, are summarized in Table I. By including these additional variables in our analysis, we aim to provide a holistic overview of the factors associated with hate crimes.

There is often a significant disparity between materialistic desires and the means of achieving them. Criminological strain theory explores the potential correlation between the disparity between materialistic aspirations and the available means to fulfill them, and the occurrence of hate crimes or violent crimes [33]. In our study, we take into consideration the influence of strain theory and other criminological theories by including specific variables. To predict the detrended $fbi\_num\_noseas$ values, we formulate our prediction task as follows:

$$\begin{aligned}
&\mathrm{fbi_num_noseasonnal_t} = \\
&+\alpha_0+\alpha_1*\mathrm{aggravated_assault_rate_{t-1}} \\
&+\alpha_2*\mathrm{arrest_drug_abuse_violation_{t-1}} \\
&+\alpha_3*\mathrm{arrest_weapon_{t-1}}+\alpha_4*\mathrm{burglary_rate_{t-1}} \\
&+\alpha_5*\mathrm{homicide_victim_black_{t-1}} \\
&+\alpha_6*\mathrm{murder_nonneg_manslaughter_rate_{t-1}} \\
&+\alpha_7*\mathrm{population_{t-1}}+\alpha_8*\mathrm{rape_rate_{t-1}} \\
&+\alpha_9*\mathrm{robbery_rate_{t-1}} \\
&+\alpha_{10}*\mathrm{law_enforce_employee_{t-1}} \\
&+\alpha_{11}*\mathrm{uner_quar_{t-1}}+\varepsilon_2
\end{aligned} \tag{2}$$

Here, $\varepsilon_2$ denotes the residuals that captures the unaccounted variability in the model. These variables encompass the rate of unemployment, rate of homicide and non-negligent manslaughter, incidence of rape, incidence of robbery, incidence of aggravated assault, incidence of burglary, Victims of homicides among the Black population, the number of arrests for carrying or possessing weapons, the number of arrests for drug abuse, and the total number of law enforcement employees.

In line with the social psychological theory, media coverage, particularly sensationalist reporting of hate crime events, can contribute to the proliferation of such crimes [24]. To investigate this correlation, we employ event extraction techniques to specifically identify hate crime incidents from news articles. From this extracted data, we construct two social-psychological variables: the number of The New York Times articles related to hate crime events ($news\_num$) and the number of hate crime events detected from the news ($event\_detect\_num$).

It is important to acknowledge that the hate crime events identified from The New York Times may differ from those reported by the FBI. The sources may identify different incidents, and there may be variations in statistical granularity and coverage. However, in our analysis, we employ the hate crime incidents reported in the media as predictors, which is not incongruous with using the hate crime incident indicators released by the FBI as the dependent variable. These two types of data are interconnected and complementary, rather than being interchangeable substitutes for one another.

To incorporate these two event-related variables into our analysis, we update the formula 2 as follows:

$$\begin{aligned}
&\mathrm{fbi_num_noseasonnal_t} = \\
&+\alpha_0+\alpha_1*\mathrm{aggravated_assault_rate_{t-1}} \\
&+\alpha_2*\mathrm{arrest_drug_abuse_violation_{t-1}} \\
&+\alpha_3*\mathrm{arrest_weapon_{t-1}}+\alpha_4*\mathrm{burglary_rate_{t-1}} \\
&+\alpha_5*\mathrm{homicide_victim_black_{t-1}} \\
&+\alpha_6*\mathrm{murder_nonneg_manslaughter_rate_{t-1}} \\
&+\alpha_7*\mathrm{population_{t-1}}+\alpha_8*\mathrm{rape_rate_{t-1}} \\
&+\alpha_9*\mathrm{robbery_rate_{t-1}} \\
&+\alpha_{10}*\mathrm{law_enforce_employee_{t-1}} \\
&+\alpha_{11}*\mathrm{uner_quar_{t-1}}+\alpha_{12}*\mathrm{event_detect_num} \\
&+\alpha_{13}*\mathrm{news_num}+\varepsilon_3
\end{aligned} \tag{3}$$

In the updated model (equation 3), we introduce both $news\_num$ and $event\_detect\_num$ as independent variables to examine their impacts on hate crime incidents. The coefficient associated with $news\_num$ represents the influence of the number of New York Times articles pertaining to hate crime events at time $t$, while the coefficient associated with $event\_detect\_num$ signifies the impact of the number of hate crime events extracted from the news at time $t$. Positive coefficients imply that more media coverage corresponds to a rise in hate crime incidents. Conversely, negative coefficients indicate an inverse connection between media coverage and hate crimes, meaning that increased media coverage is associated with a decrease in hate crime incidents.

TABLE I
SUMMARY OF POTENTIAL PREDICTIVE FACTORS FOR HATE CRIME IDENTIFIED IN MULTIPLE THEORETICAL APPROACHES.

| Theoretical Approaches | Potential Predictive Factors | Whether Measurable | Whether Accepted |
|---|---|---|---|
| Strain Theory Criminology [24]–[26] | population | Yes | Yes |
| | employment rate | Yes | No |
| | unemployment rate | Yes | Yes |
| | economic downturn | Yes | No |
| | poverty rate | Yes | No |
| Criminological Theories [27]–[30] | rate of violent criminal incidents | Yes | No |
| | incidence rate of murder and nonnegligent manslaughter | Yes | Yes |
| | rate of sexual assault incidents | Yes | Yes |
| | rate of robbery incidents | Yes | Yes |
| | rate of aggravated assault incidents | Yes | Yes |
| | rate of property crime incidents | Yes | No |
| | rate of burglary incidents | Yes | Yes |
| | rate of larceny theft incidents | Yes | No |
| | white individuals killed in homicides | Yes | No |
| | black individuals killed in homicides | Yes | Yes |
| | arrests related to acts of violence | Yes | No |
| | arrests for carrying weapons | Yes | Yes |
| | arrests for offenses involving drug abuse | Yes | Yes |
| | arrests related to theft | Yes | No |
| | total complement of law enforcement personnel | Yes | Yes |
| | number of agencies | Yes | No |
| Social-Psychological Theories [24], [31], [32] | quantity of news coverage | Yes | Yes |
| | quantity of news coverage involving hate crimes | Yes | Yes |
| | hate exposure level | Yes | Yes |
| | prejudice | No | No |

Nevertheless, using only the raw counts of $news\_num$ and $event\_detect\_num$ may not fully capture the magnitude of hate crime reports circulating on social media during a given period. To address this limitation, we develop a relative indicator, referred to as the $hate\_exposure\_index$, to portray the prevailing level of hate crime dissemination at a specific time. The calculation of the $hate\_exposure\_index$ is as follows:

$$\text{hate\_exposure\_index} = \frac{\text{event\_detect\_num}}{\text{news\_num}} \tag{4}$$

In equation 4, the $hate\_exposure\_index$ is introduced as a relative indicator derived from $news\_num$ and $event\_detect\_num$ to depict the present extent of hate crime dissemination. By utilizing this relative indicator, we obtain a more precise representation of the diffusion of hate crime events on social media platforms.

Subsequently, we consolidate factors associated with strain theory criminology, factors linked to criminological theories, and variables pertaining to social psychology theories into a unified predictive model. Our objective is to forecast hate crimes utilizing the following equation:

$$\begin{aligned}
&\mathrm{fbi_num_noseasonnal_t} = \\
&\alpha_0 + \alpha_1 * \mathrm{aggravated_assault_rate_{t-1}} \\
&+\alpha_2 * \mathrm{arrest_drug_abuse_violation_{t-1}} \\
&+\alpha_3 * \mathrm{arrest_weapon_{t-1}} + \alpha_4 * \mathrm{burglary_rate_{t-1}} \\
&+\alpha_5 * \mathrm{homicide_victim_black_{t-1}} \\
&+\alpha_6 * \mathrm{murder_nonneg_manslaughter_rate_{t-1}} \\
&+\alpha_7 * \mathrm{population_{t-1}} + \alpha_8 * \mathrm{rape_rate_{t-1}} \\
&+\alpha_9 * \mathrm{robbery_rate_{t-1}} \\
&+\alpha_{10} * \mathrm{law_enforce_employee_{t-1}} \\
&+\alpha_{11} * \mathrm{uner_quar_{t-1}} + \alpha_{12} * \mathrm{event_detect_num} \\
&+\alpha_{13} * \mathrm{news_num} + \alpha_{14} * \mathrm{hate_exposure_index} + \varepsilon_4
\end{aligned} \tag{5}$$

This integrated model enables us to explore the collective impacts of diverse factors on the prediction of hate crimes. By incorporating variables from various criminological and social psychological theories, we attain a more comprehensive comprehension of the intricate elements that contribute to the manifestation of hate crimes. This approach allows for a deeper understanding of the multifaceted nature of hate crimes and facilitates a more holistic analysis of their determinants.

Lastly, we unified the hate crime prediction task, encompassing the ARIMA model, factors associated with strain theory criminology, factors linked to criminological theories, and variables pertaining to social psychology hate crime theory, into a comprehensive model. This integrated framework captures the combined influences of diverse factors on the prediction of hate crimes, while considering the temporal dependencies present in the hate crime time series. The representation of this model is expressed by the following regression equation:

$$\begin{aligned}
&\mathrm{fbi_num_noseasonnal_t} = \alpha_0 + f(ARIMA(p,d,q)) \\
&+\alpha_1 * \mathrm{aggravated_assault_rate_{t-1}} \\
&+\alpha_2 * \mathrm{arrest_drug_abuse_violation_{t-1}} \\
&+\alpha_3 * \mathrm{arrest_weapon_{t-1}} + \alpha_4 * \mathrm{burglary_rate_{t-1}} \\
&+\alpha_5 * \mathrm{homicide_victim_black_{t-1}} \\
&+\alpha_6 * \mathrm{murder_nonneg_manslaughter_rate_{t-1}} \\
&+\alpha_7 * \mathrm{population_{t-1}} + \alpha_8 * \mathrm{rape_rate_{t-1}} \\
&+\alpha_9 * \mathrm{robbery_rate_{t-1}} \\
&+\alpha_{10} * \mathrm{law_enforce_employee_{t-1}} \\
&+\alpha_{11} * \mathrm{uner_quar_{t-1}} + \alpha_{12} * \mathrm{event_detect_num} \\
&+\alpha_{13} * \mathrm{news_num} + \alpha_{14} * \mathrm{hate_exposure_index} + \varepsilon_5
\end{aligned} \tag{6}$$

In order to assess the validity of the hate crime events, we excluded three event-related factors from equation 6. The modified equation is presented as follows:

$$\begin{aligned}
&\mathrm{fbi_num_noseasonnal_t} = \alpha_0 + f(ARIMA(p,d,q)) \\
&+\alpha_1 * \mathrm{aggravated_assault_rate_{t-1}} \\
&+\alpha_2 * \mathrm{arrest_drug_abuse_violation_{t-1}} \\
&+\alpha_3 * \mathrm{arrest_weapon_{t-1}} + \alpha_4 * \mathrm{burglary_rate_{t-1}} \\
&+\alpha_5 * \mathrm{homicide_victim_black_{t-1}} \\
&+\alpha_6 * \mathrm{murder_nonneg_manslaughter_rate_{t-1}} \\
&+\alpha_7 * \mathrm{population_{t-1}} + \alpha_8 * \mathrm{rape_rate_{t-1}} \\
&+\alpha_9 * \mathrm{robbery_rate_{t-1}} \\
&+\alpha_{10} * \mathrm{law_enforce_employee_{t-1}} \\
&+\alpha_{11} * \mathrm{uner_quar_{t-1}} + \varepsilon_6
\end{aligned} \tag{7}$$

By removing the event-related factors, we can explore the effect of various factors on the occurrence of hate crime that extend beyond the influence of recent specific events. This helps evaluate the predictive capability of our model in forecasting hate crime trends even in the absence of recent events, thereby testing its validity and robustness.

In order to differentiate among the various models, we have assigned distinct names to each of them. Employing specific names for each model enables convenient referencing during discussions and comparisons. Furthermore, this naming approach facilitates the evaluation of the models' predictive performance in hate crime forecasting, a crucial step in assessing the efficacy of our methodology. To be more precise, we assign the following labels to the equations: Model 1 refers to equation 1, Model 2 refers to equation 2, Model 3 refers to equation 3, Model 4 refers to equation 5, Model 5 refers to equation 6, and Model 6 refers to equation 7. An overview of all the models is summarized in Table II.

In our analysis, we employ the commonly used method, namely Maximum Likelihood Estimation (MLE), to estimate the parameters for the different models. By utilizing these established estimation techniques, we aim to obtain reliable parameter estimates for each model, ensuring the accuracy and validity of our statistical analyses.

## V. Experiments and Results

In this section, we present the experimental results, which include the findings from event extraction, the predictive outcomes of federal hate crime trends using linear regression technique, and the prediction results of subnational hate crime trends using longitudinal data regression method. The models are trained using a dataset spanning from 2007 to 2018 to forecast hate crime trends beyond 2018. Additionally, we assess the accuracy and reliability of these predictions by comparing them with the FBI's statistical reports.

To evaluate the event extraction method, we utilize Recall, Precision, and F-measure as statistical performance metrics. These measurements are widely used to evaluate classification models and provide a clear understanding of their classification capabilities. True positives and true negatives represent the correctly forecasted values, while false positives and false negatives indicate instances where the forecasted values contradicts the actual values. Recall, Precision, and F-measure are widely recognized as the three primary statistical assessment metrics for event extraction. Recall, Precision, and F-measure are calculated using the following formulas:

$$Recall = \frac{True\_Positives}{True\_Positives + False\_Negatives} \tag{8}$$

$$Precision = \frac{True\_Positives}{True\_Positives + False\_Positives} \tag{9}$$

$$F1\ score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{10}$$

Furthermore, we employ a range of other evaluation metrics to analyze the performance of the predictive framework. These evaluation metrics include Log-Likelihood, Adjusted R-Squared, Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE). RMSE, in particular, is a widely adopted accuracy metric that quantifies the deviation between the forecasted values and the true observations:

$$RMSE = \sqrt{\frac{1}{N}\sum\nolimits_{\mathrm{t}=1}^{N}(y_t - \hat{y}_t)^2} \tag{11}$$

However, RMSE is influenced by the scale of the data, making it inappropriate for direct comparison between models that employ different datasets. To address this limitation, we additionally utilize MAPE as an evaluation metric. MAPE is expressed as a percentage and remains unaffected by variations in data size and dimensionality, allowing for a reliable and meaningful comparison of the accuracy achieved by different models:

$$MAPE = \frac{1}{N}\sum\nolimits_{\mathrm{t}=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| \tag{12}$$

Here, $y_t$ represents the true observation while $\hat{y}_t$ stands for the estimated value. $N$ symbolizes the overall count of observations within the dataset.

### A. Event Extraction Performance

Patch Hate Crime corpus is partitioned into a training set (70% of the observations), a validation set (10% of the observations), and a testing set (20% of the observations). The sample size of each partition is shown in Table III. For consistency, we adopt the identical parameter configurations as [7] and conduct the experiment in 10 iterations. The mean Recall, Precision, and F1 score obtained from these repeated experiments are presented in Table IV.

The results of our repeated experiments correspond to the results presented in [7]. Subsequently, we preserve the trained model for the purpose of detecting hate crime events in news articles sourced from The New York Times. We proceed by quantifying the news articles and hate crime events identified by event extraction module, categorizing these measured values on a quarterly basis. This process yields two distinct time series: the count of news articles from The New York Times ($news\_num$) and the count of detected hate crime events ($event\_detect\_num$).

### B. Time Series Prediction Results

After eliminating the seasonal effects from the FBI's original statistics ($fbi\_num$), we obtain the modified time series referred to as $fbi\_num\_noseas$. To assess the stationarity of this time series, we utilize the Augmented Dickey-Fuller (ADF) test. The test results indicate the presence of a nonlinear trend, confirming that the sequence is not stationary. In order to convert the non-stationary sequence into a stationary one, we perform the differencing operation, resulting in the transformed time series $d\_fbi\_num\_noseas$.

TABLE II
SUMMARY OF MODELS

| Models | Equations | Descriptions |
|---|---|---|
| Model 1 | Equation (1) | ARIMA model. |
| Model 2 | Equation (2) | Equation including predictors of strain theory and criminological theories. |
| Model 3 | Equation (3) | Equation3 is updated by adding two event-related variables into formula 2. |
| Model 4 | Equation (5) | Equation5 is updated by adding another event-related variable $hate\_exposure\_index$ into formula 3. |
| Model 5 | Equation (6) | Equation6 is an integrated model, including the ARIMA factors, stress criminology theory factors, general criminology theory factors, and social psychology hate crime theory factors. |
| Model 6 | Equation (7) | To examine the efficiency of the event factors, three event extracted factors are removed from Equation6. |

TABLE III
PATCH HATE CRIME DATASET PARTITION.

| dataset partition | sample size | Events Categories |
|---|---|---|
| training set | 3619 | e.g. personal attacks, arson, |
| validation set | 513 | intentional destruction, and hate |
| testing set | 1039 | demonstrations. |

TABLE IV
PERFORMANCE EVALUATION OF THE EVENT EXTRACTION MODULE.

| Corpus | Recall | Precision | F1 Score |
|---|---|---|---|
| Patch Hate Crime Corpus | 0.8325 | 0.8164 | 0.8244 |

To determine the optimal order of the AR and MA terms in the ARIMA model, we utilize the Partial Autocorrelation Function (PACF) and Autocorrelation Function (ACF) tests. By analyzing the PACF test statistics, we find that the optimal order for the AR term is 1, indicating that the current observation is influenced by the immediately preceding observation. The MA term, on the other hand, has an order of 0, suggesting that there is no significant influence from the lagged moving average terms. Based on these findings, we proceed to fit an ARIMA(1,1,0) model to the stationary data $d\_fbi\_num\_noseas$.

The fitted model's performance is assessed using several statistical measures, including the Log-Likelihood, Adjusted R-Squared, and D-W statistical magnitude. The Log-Likelihood value of -283.39 signifies the goodness of fit, with higher values indicating a better fit. With an Adjusted R-Squared value of 0.7415, the model accounts for approximately 74.15% of the variance in dependent variables, demonstrating a substantial explanation of its variability. Additionally, the D-W statistic of 1.65 suggests the absence of significant autocorrelation in the residuals. We conduct dynamic predictions of hate crime trends using the fitted model and assess its performance using RMSE and MAPE, which yield values of 93.11 and 4.60, respectively. The statistical evaluation results are summarized in Table V, providing an summary of the performance of the baseline model (Model 1). The subsequent models will be compared against the results obtained from this baseline model for comparative analysis.

TABLE V
PERFORMANCE OF ARIMA MODEL ON THE NATIONAL HATE CRIME PREDICTION.

| R-Squared | Log Likelihood | Durbin-Watson | RMSE | MAPE |
|---|---|---|---|---|
| 0.7415 | -283.3863 | 1.65 | 93.1066 | 4.5966 |

### C. National Hate Crime Trend Prediction Results

We conduct a thorough evaluation of multiple models, including Model 2, Model 3, Model 4, Model 5, and Model 6, to forecast national hate crime trends. Table VI serves as a comprehensive summary of the forecast performance, allowing for a straightforward comparison of the performance of different models. Table VI showcases a tabular format where each column provides a specific measure for assessing the model's performance, while every row demonstrates a unique model's performance.

Upon examining the findings depicted in Table VI, five noteworthy observations come to light.

*1) The regression models outperformed the ARIMA model on nearly all evaluation metrics, suggesting their superiority:* This can be attributed to the regression models' incorporation of theoretical predictors of hate crime motivation, which goes beyond the ARIMA model's consideration of solely the properties of the sequential data. This emphasizes the significance of these predictors in improving the predictive performance of the models.

*2) The substantial predictive value of directly extracted event factors:* Among the regression models, those incorporating event-related variables (Model 3 and Model 4) demonstrated superior performance compared to the model without such variables (Model 2). Additionally, among the models that consider time series factors, Model 5 exhibited superior performance compared to Model 6. These results indicate the substantial predictive value of event factors derived from event extraction module.

*3) The substantial predictive value of the supplementary event factor $hate\_exposure\_index$:* Among the regression models, Model 4 (RMSE = 55.8069 and MAPE = 2.1358) exhibited the highest performance. This finding confirms our hypothesis that the event factors obtained from event extraction technique are inadequate. By constructing a relative measure indicator to express hate crime exposure level, the model performance has improved. This indicates the usefulness of the constructed relative measure indicator $hate\_exposure\_index$, highlighting its value as a beneficial supplement for hate crime trends prediction.

*4) Hybrid models are not necessarily optimal, particularly when the time series experiences sudden changes in trend:* In our study, we observed that Model 1 attained the most elevated level of Adjusted R-Square compared to the other models analyzed. To investigate the potential of a hybrid model, we constructed Model 5, which integrates both the ARIMA and regressive models. However, as shown in Table VI, while the fitting evaluation metrics outperformed other models, it did not match the forecast evaluation metrics of Model 4. This discrepancy can be attributed to the steep

TABLE VI
PERFORMANCE OF DIFFERENT MODELS ON THE NATIONAL HATE CRIME PREDICTION TASK

| Models | Evaluation Metrics | | | |
|---|---|---|---|---|
| | Log Likelihood | Adjusted R-Squared | RMSE | MAPE |
| Model 1 | -283.3863 | 0.7415 | 93.1066 | 4.5966 |
| Model 2 | -279.5407 | 0.7177 | 62.8915 | 2.6049 |
| Model 3 | -278.6715 | 0.7119 | 86.3024 | 3.5789 |
| **Model 4** | **-278.1900** | **0.7258** | **55.8069** | **2.1358** |
| Model 5 | -280.6132 | 0.7145 | 57.6689 | 2.2772 |
| Model 6 | -282.3604 | 0.7110 | 74.4394 | 3.9717 |

decline in hate crime numbers during the first quarter of 2019. The inertia of the ARIMA component likely contributed to increased errors, resulting in Model 5 exhibiting suboptimal performance on the test set. In summary, the inherent characteristics of the ARIMA model lead to an amplification of discrepancies, making Model 6 inferior to Model 2, and Model 5 inferior to Model 4. Furthermore, Model 5's performance on the test set is deemed suboptimal.

In summary, the experimental results substantiate the effectiveness of event-extracted variables in enhancing hate crime prediction performance, thereby underscoring the efficacy of the event extraction module within our framework.

*5) National Hate Crime Trend Prediction Visualization :* The performance of each model can be visualized to provide a more intuitive understanding. Figure 5 illustrates the detrended hate crimes and forecast values for each model. The figure clearly demonstrates that the model incorporating event extraction factors outperforms the model without such factors, highlighting the added value of including event-related information in the predictive modeling process.

However, there is a significant disparity between the predicted and actual hate crime numbers in 2016 and 2017. Further investigation has revealed that this phenomenon can be attributed to the ”Trump Effect”. According to the research conducted by Stephen and Griffin [34], the increase in hate crimes cannot be solely attributed to a specific individual's provocative language during a particular political period. Instead, it is argued that the successful election as president may have conferred legitimacy in the minds of the hate crime perpetrators, thereby exacerbating the increase in hate crimes.

In conclusion, our experimental results provide compelling evidence supporting the effectiveness of our proposed framework for predicting national hate crime trends. The regressive models incorporating event-related variables and the hybrid model all outperform the ARIMA model, underscoring the significance of incorporating theoretical predictors and event-related variables in hate crime prediction. The event extraction module also proves to be a valuable tool in predicting hate crime trends. Nonetheless, the notable disparity between the predicted and actual hate crime numbers in 2016 and 2017 implies that there may be external factors, not accounted for in our framework, that influence hate crime trends. Future research should prioritize investigating these factors and developing more advanced models to achieve accurate predictions of national hate crime trends.

### *D. Panel Data Prediction Results and Analysis*

The panel data analysis approach has gained widespread usage in the literature for modeling longitudinal data that encompasses both time-invariant and time-varying variables. It proves particularly valuable when dealing with data that comprises both individual and time-series components. Our methodology can be easily customized to forecast hate crime patterns at the state level using the panel data analysis approach. By incorporating state-level variables and control variables, our framework facilitates a comprehensive understanding of the factors that influence hate crime trends at the state level. This approach offers a primary benefit by enhancing the degrees of freedom within the data and reducing collinearity among the explanatory variables. Furthermore, by simultaneously considering the unique characteristics of each state as well as general factors and control variables, this framework has the potential to improve the accuracy of individual predictions. Furthermore, this approach allows for the simultaneous estimation of hate crime trends across states, eliminating the requirement for individualized models for each state.

The procedure of state-level hate crime trends prediction consists of five key steps: (a) Identify the state where hate crime incidents occurred based on the textual information from news reports. (b) Organize the data into a panel structure. (c) Perform diagnostic tests, such as the Hausman test or the Breusch-Pagan Lagrange Multiplier test, to assess whether fixed effects or random effects models are more appropriate for the panel data analysis. (d) Compute the unknown coefficients that characterize the relationships between the variables within the panel data framework. (e) Predict the trajectory of hate crime patterns at the state level.

*1) State Identification and Event Factor Construction:* Firstly, we conduct information extraction technique, e.g. NER, to determine in which state the hate crime incidents refer to. To enhance the performance of recognition, we assembled a dataset comprising 730,000 names of American places or institutions and associated them with their respective states. Subsequently, 500 news articles are randomly selected and annotated with their corresponding sate labels. The annotations were performed by two annotators, and our NER program achieved a Cohen's Kappa agreement of 0.79, indicating a high level of reliability and suitability for preparing panel data variables. The descriptive statistics of automatically mapping location names mentioned in the news reports to their corresponding states using NER are presented in Table VII.

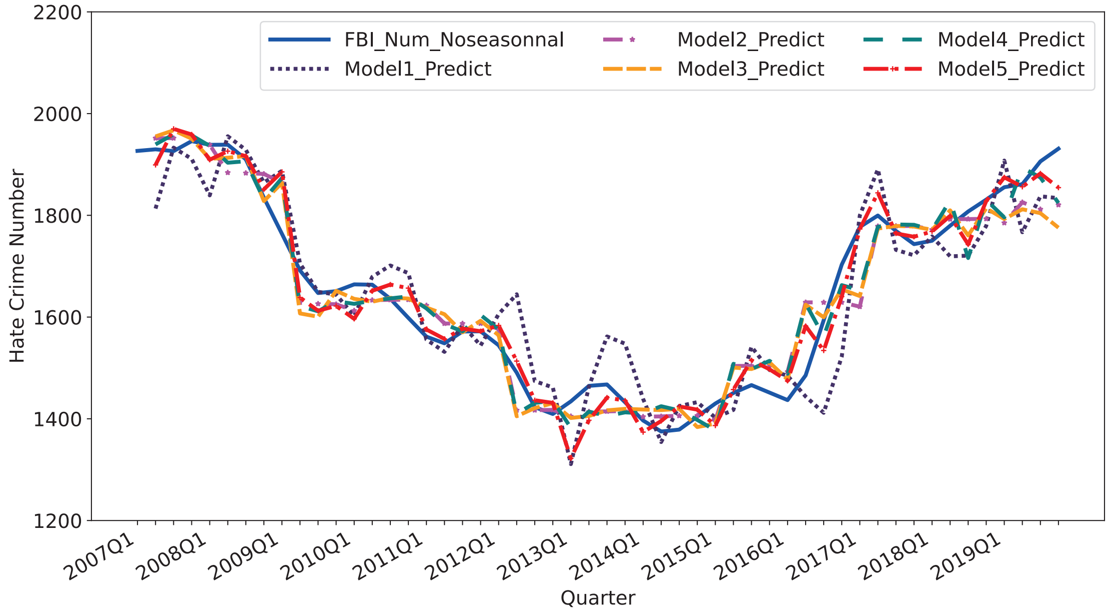


Fig. 5. Visualization of the performance of different approaches in predicting national hate crime incidents.

TABLE VII
THE DESCRIPTIVE STATISTICS FOR NER DATASET.

| Sample size | Annotators | Manually annotated samples | Cohen's Kappa statistics |
|---|---|---|---|
| 730,000 | 2 | 500 | 0.79 |

TABLE VIII
THE TATE-LEVEL PREDICTION RESULTS.

| Models | Evaluation Metrics | | | |
|---|---|---|---|---|
| | Log Likelihood | R-Squared | RMSE | MAPE |
| Model 7 | -4844.8233 | 0.6177 | 16.7422 | 39.8561 |
| Model 8 | -4800.4589 | 0.6223 | 14.2305 | 37.1116 |

The analysis considered only common public predictor variables and state-specific event factors, while disregarding state-specific special factors related to each state. The emphasis is on demonstrating the utility of event factors.

Following the methodology employed in the previous national prediction task, we utilize the NER method to automatically map location names mentioned in the news reports to their corresponding states. Subsequently, we employ a quarterly approach, wherein we tally the occurrences of news articles and identified hate crime events, organizing them according to state. Consequently, each state is linked with a pair of time series variables: the number of news articles involving that state ($*\_news\_num$) and the predicted number of hate crime incidents for that state ($*\_event\_detect\_num$), with the asterisk (*) symbol indicating any given state. We then transform these variables into a panel data format, incorporating event-related and other factors.

*2) Panel Data Preparation and Effects Testing:* Due to the sparsity of statistical data provided by the FBI and the incompleteness of extracted hate crime incidents from news sources, we have excluded samples from 4 states. This led to the creation of a panel data set comprising 47 states and 99.4% of the sample data. This dataset was used for the final analysis and modeling of state-level hate crime prediction.

Hausman test can be utilized to ascertain the optimal model form between fixed effects (FE) and random effects (RE). It helps researchers decide whether to include individual-specific effects (fixed effects) or assume that these effects are random and uncorrelated with the regressors (random effects). The results of the Hausman test indicated that utilizing the equation with fixed effects is a superior choice. This implies that each state possesses its own distinct intercept, thereby accounting for state-specific factors that contribute to hate crime trends.

*3) Parameter Estimation and Experimental Results:* We employed the Least Squares (LS) method to compute the coefficients using the dataset spanning from January 2007 to December 2018. In alignment with the national hate crime prediction, the test dataset consists of samples spanning from January 2019 to December 2019. Similar to the national hate crime trends prediction, we forecasted hate crime trends for the period between 2019 quarter 1 and 2019 quarter 4.

The experimental results of Model 7, which excludes event-related factors, and Model 8, which integrates event-related factors, are presented in Table VIII. The results demonstrate that incorporating event-related variables in Model 8 enhances its performance in terms of RMSE (2.5) and MAPE (2.7) compared to Model 7. This suggests that event-related variables offer valuable information and contribute to the improved accuracy of state-level hate crime predictions.

Although the RMSE and MAPE values are slightly higher than desired, the emphasis is on demonstrating the utility of event factors and the generalization of our framework. The analysis considered only common public predictor variables and state-specific event factors, while disregarding state-specific special factors related to each state. The experimental results confirm this observation, demonstrating the effectiveness of event-related factors in predicting state-level hate crime trends.

TABLE IX
THE LEVENE TEST AND PAIRED SAMPLES T-TEST RESULTS FOR ASSESSING THE IMPACT OF FACTORS RELATED TO EVENTS.

| Test | Statistics | P-value | Conclusion |
|---|---|---|---|
| Levene test | 9.49e-06 | 0.9975 | Variance equales. |
| Paired t-test | -7.931 | 2.090e-13 (***) | Average not equals: the actual: 38.6957 Model 7: 34.7538 Model 8: 35.1390 |

*4) Significance Test of the Effectiveness of Event Factors:* To examine the significance of the effectiveness of event-related factors, we conduct a significance test using appropriate statistical methods. We conduct a Levene test and paired samples t-test comparing Model 7, which excludes event-related factors, and Model 8, which integrates event-related factors, to assess the performance in predicting state-level hate crime trends. The Levene test and paired samples t-test results are shown in Table IX.

The Levene Test for equal variances indicates comparable variances between the two prediction series, with a Levene value of 9.49e-06 and a p-value greater than 0.05, suggesting no significant difference in variances. However, the paired samples t-test (with statistic -7.931 following its p-value 2.090e-13) demonstrates that the two forecasting series have different means.

The actual data points in the test dataset have an average value of 38.6957. Model 7 has a mean of 34.7538, while Model 8 has a mean of 35.1390. The mean value of Model 8, which incorporates event factors, is more closer to the actual mean value. These results indicate a significant superiority of Model 8 over Model 7. The incorporation of event factors in Model 8 substantially enhances the prediction of state-level hate crimes. These findings provide empirical evidence supporting the effectiveness of the proposed framework for hate crime trends prediction based on event extraction.

In conclusion, the results of the paired samples t-test present compelling evidence supporting the efficacy and scalability of the event extraction module. This module not only proves valuable in predicting hate crime trends at the national level but also demonstrates its utility in forecasting state-level trends. These findings establish the efficacy of our proposed framework for hate crime trends prediction, which relies on event extraction, and highlight its applicability across various levels of analysis.

### E. Shortcomings and Future Work

While various theories, such as prejudice theory, criminological theory, and social-psychological theory, offer insights into the causes of hate crime, limited research has systematically explored the underlying factors driving these crimes. Our primary emphasis was on integrating specific variables from the fields of criminology and social psychology as factors for prediction, aiming to demonstrate the potential of event-related factors in enhancing prediction accuracy. Consequently, we acknowledge the omission of other factors from theories such as prejudice theory for the purpose of comparative analysis. Future research endeavors will encompass a more comprehensive examination of factors identified in existing literature, seeking to validate their efficacy in hate crime prediction.

Moreover, we intend to expand our data collection efforts and perform analyses to explore potential shifts or changes in patterns to examine the ”Trump effect.” By employing a two-stage model, we will evaluate its impact on enhancing predictive performance. Furthermore, we are actively developing novel models, including Copula-based models and neural network models, which hold the potential to improve the accuracy and robustness of our hate crime prediction framework.

## VI. CONCLUSION

Hate crimes present a significant global challenge, underscoring the need for timely identification of national and local trends to enable effective responses by policymakers and law enforcement agencies. In this study, we propose an innovative framework that integrates information extraction technologies, particularly event extraction, for predicting hate crime trends.

Our framework offers several advantages over existing approaches. Firstly, it can not only predict hate crime trends at the national level, but can also be readily adjusted to predict state-level hate crime patterns by incorporating panel data analysis techniques, facilitating a comprehensive understanding of hate crime dynamics across different levels. Secondly, by leveraging event extraction technologies, our framework captures the impact of significant events on hate crime trends, thereby improving the accuracy and timeliness of predictions. Moreover, our framework demonstrates flexibility and can be modified to predict trends in other types of crimes.

Experimental results unequivocally demonstrate the remarkable performance enhancement achieved by our framework in predicting hate crime trends. Specifically, the event-extracted factors significantly improves the prediction accuracy, providing valuable insights into the underlying causes of hate crimes. Furthermore, our framework offers holistic and strategic insights for policymakers and law enforcement agencies, enabling them to develop effective responses to hate crimes and proactively justify specific legislation. We firmly believe that our work will inspire further research in this domain, contributing to the reduction of hate crime incidents, fostering social harmony, and advancing a more inclusive society.

## REFERENCES

[1] K. Müller and C. Schwarz, “From hashtag to hate crime: Twitter and antiminority sentiment,” *American Economic Journal: Applied Economics*, vol. 15, no. 3, pp. 270–312, 2023.
[2] R. J. Cramer, S. Wright, M. M. Long, N. D. Kapusta, M. R. Nobles, T. M. Gemberling, and H. J. Wechsler, “On hate crime victimization: Rates, types, and links with suicide risk among sexual orientation minority special interest group members,” *Journal of Trauma & Dissociation*, vol. 19, no. 4, pp. 476–489, 2018.
[3] N. D. Gitari, Z. Zuping, H. Damien, and J. Long, “A lexicon-based approach for hate speech detection,” *International Journal of Multimedia and Ubiquitous Engineering*, vol. 10, no. 4, pp. 215–230, 2015.
[4] M. Mozafari, R. Farahbakhsh, and N. Crespi, “A bert-based transfer learning approach for hate speech detection in online social media,” in *Complex Networks and Their Applications VIII*, ser. Studies in Computational Intelligence. Springer, 2020, pp. 928–940.
[5] S. Han, H. Huang, J. Liu, and S. Xiao, “American hate crime trends prediction with event extraction,” *arXiv preprint arXiv:2111.04951*, 2021.
[6] R. Rawat, “Logical concept mapping and social media analytics relating to cyber criminal activities for ontology creation,” *International Journal of Information Technology*, vol. 15, no. 2, pp. 893–903, 2023.

[7] A. Mostafazadeh Davani, L. Yeh, M. Atari, B. Kennedy, G. Portillo-Wightman, E. Gonzalez, N. Delong, R. Bhatia, A. Mirinjian, and X. Ren, "Reporting the unreported: Event extraction for analyzing the local representation of hate crimes," *arXiv preprint arXiv: 1909.02126*, 2019.

[8] L. Silva, M. Mondal, D. Correa, F. Benevenuto, and I. Weber, "Analyzing the targets of hate in online social media," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 10, no. 1, 2016, pp. 687–690.

[9] W. Yin and A. Zubiaga, "Towards generalisable hate speech detection: a review on obstacles and solutions," *PeerJ Computer Science*, vol. 7, p. e598, 2021.

[10] A. Pawar, P. Gawali, M. Gite, M. A. Jawale, and P. William, "Challenges for hate speech recognition system: Approach based on solution," in *2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS)*, 2022, pp. 699–704.

[11] A. Chhabra and D. K. Vishwakarma, "A literature survey on multimodal and multilingual automatic hate speech identification," *Multimedia Systems*, vol. 29, pp. 1203–1230, 2023.

[12] F. Hogenboom, F. Frasincar, U. Kaymak, F. De Jong, and E. Caron, "A survey of event extraction methods from text for decision support systems," *Decision Support Systems*, vol. 85, pp. 12–22, 2016.

[13] F. Hamborg, C. Breitinger, and B. Gipp, "Giveme5w1h: A universal system for extracting main events from news articles," *arXiv preprint arXiv:1909.02766*, 2019.

[14] M. Xu and S. Liu, "Semantic-enhanced and context-aware hybrid collaborative filtering for event recommendation in event-based social networks," *IEEE Access*, vol. 7, pp. 17 493–17 502, 2019.

[15] M. Li, R. Xu, S. Wang, L. Zhou, X. Lin, C. Zhu, M. Zeng, H. Ji, and S.-F. Chang, "Clip-event: Connecting text and images with event structures," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 16 420–16 429.

[16] K. Relia, Z. Li, S. H. Cook, and R. Chunara, "Race, ethnicity and national origin-based discrimination in social media and hate crimes across 100 us cities," in *Proceedings of the International AAAI Conference on Web and Social Media*, vol. 13, no. 1, 2019, pp. 417–427.

[17] G. M. Herek, J. R. Gillis, and J. C. Cogan, "Psychological sequelae of hate-crime victimization among lesbian, gay, and bisexual adults," *Journal of Consulting and Clinical Psychology*, vol. 67, no. 6, pp. 945–951, 1999.

[18] R. Lu and Y. Sheng, "From fear to hate: How the covid-19 pandemic sparks racial animus in the united states," *arXiv preprint arXiv:2007.01448*, 2020.

[19] I. M. Sumertajaya, E. Rohaeti, A. H. Wigena, and K. Sadik, "Vector autoregressive-moving average imputation algorithm for handling missing data in multivariate time series," *IAENG International Journal of Computer Science*, vol. 50, no. 2, pp. 727–735, 2023.

[20] W. Wang, Y. Ning, H. Rangwala, and N. Ramakrishnan, "A multiple instance learning framework for identifying key sentences and detecting events," in *Proceedings of the 25th ACM International on Conference on Information and Knowledge Management*, 2016, pp. 509–518.

[21] L. Dongping, Y. Yingchun, S. Shikai, H. Jun, S. Haoru, Y. Qiang, H. Sunyan, and D. Fei, "Research on deep learning model of multimodal heterogeneous data based on lstm," *IAENG International Journal of Computer Science*, vol. 49, no. 4, pp. 1016–1022, 2022.

[22] Y. Wang, X. Cheng, and X. Meng, "Sentiment analysis with an integrated model of bert and bi-lstm based on multi-head attention mechanism," *IAENG International Journal of Computer Science*, vol. 50, no. 1, pp. 255–262, 2023.

[23] Aripin, A. Wisnu, and H. Solichul, "Multichannel convolutional neural network model to improve compound emotional text classification performance," *IAENG International Journal of Computer Science*, vol. 50, no. 3, pp. 866–874, 2023.

[24] D. P. Green, L. H. McFalls, and J. K. Smith, "Hate crime: An emergent research agenda," *Annual Review of Sociology*, vol. 27, no. 1, pp. 479–504, 2001.

[25] N. Hall, A. Corb, P. Giannasi, and G. John, "Understanding hate crimes: Sociological and criminological perspectives," in *The Routledge International Handbook on Hate Crime*. Routledge, 2014, pp. 69–80.

[26] M. A. Walters, "A general theories of hate crime? strain, doing difference and self control," *Critical Criminology*, vol. 19, pp. 313–330, 2011.

[27] C. E. Mills, "A common target: Anti-jewish hate crime in new york city communities, 1995-2010," *Journal of Research in Crime and Delinquency*, vol. 57, no. 6, pp. 643–692, 2020.

[28] C. E. Mills, J. D. Freilich, S. M. Chermak, T. J. Holt, and G. LaFree, "Social learning and social control in the off-and online pathways to hate crime and terrorist violence," *Studies in Conflict & Terrorism*, vol. 44, no. 9, pp. 701–729, 2021.

[29] B. Perry, *In the name of hate: Understanding hate crimes*. Routledge, 2002.

[30] S. Karstedt, H. Nyseth Brehm, and L. C. Frizzell, "Genocide, mass atrocity, and theories of crime: Unlocking criminology's potential," *Annual Review of Criminology*, vol. 4, pp. 75–97, 2021.

[31] T. F. Pettigrew, "Social psychological perspectives on trump supporters," *Journal of Social and Political Psychology*, vol. 5, no. 1, pp. 107–116, 2017.

[32] R. Agnew, "The contribution of social-psychological strain theory to the explanation of crime and delinquency," in *The Legacy of Anomie Theory*. Routledge, 1999, pp. 113–137.

[33] N. Hall, A. Corb, P. Giannasi, and G. John, "Understanding hate crimes: Perspectives from the wider social sciences," in *The Routledge International Handbook on Hate Crime*. Routledge, 2014, pp. 81–94.

[34] G. S. Edwards and S. Rushin, "The effect of president trump's election on hate crimes," *Available at SSRN 3102652*, 2018.