# Estimating Crime Rates Using Jumping Finite Automata on Tweets

Stephen Obare, Abejide Ade-Ibijola, George Okeyo and Kennedy Ogada

***Abstract*—In the Fourth Industrial Revolution, crime is hardly reported to the Police, or other law enforcement agencies. Most victims prefer to go to Social Media and vent, as this medium is easier for them to access and requires no paperwork or interrogations. This trend leaves policy makers and the law enforcers with skewed dataset, due to unreported crimes. Hence, it is paramount that one finds a way to "mine" the crime data reported on social media. In this paper, we have attempted to estimate crime rates, using Twitter as a data source. To do this, we have used a formal technique — Jumping Finite Automata (JFA), for the abstraction of a corpus of crime-related words and used shuffle algorithms to establish semantic relationships between these words. The JFA was implemented in a tool called "Crime-Ripper". Crime-Ripper uses tweets retrieved from crime hashtags on Twitter to estimate crime rates and produce reports that are map annotations, showing areas of a city and their respective estimated crime-rates. Crime-Ripper is expected to find applications in law enforcement, policy making and public safety sensitization.**

***Index Terms*—Information Extraction, Tweet Comprehension, Crime Estimation, Jumping Finite Automata Applications, parsing**

## I. Introduction

CRIME, in any society, is notorious for hampering the quality of life, distorting the growth of economies, promoting underdevelopment, and making governance of these societies expensive and difficult [1, 2, 3]. A victimization survey in Mexico shows that in that year, victims of crime suffered loses worth US$ 12.9 billion, with associated health expenses of US$ 619 million [4]. Other ills associated with high crime rates include unemployment, high cost of security personnel, and decrease in economic growth [4, 5].

In South Africa, high crime rate is known as one major deterrent of both local and international tourists, and this remains a barrier to the growth of tourism in the country [6, 7]. The trend of crime in Johannesburg (between 2005 and 2022) as reported by the government of South Africa, shows that crimes such as aggravated robbery, and drug use are consistently on the rise [8, 9]. The entire stack of crime in Johannesburg include: sexual offences, murder, assaults (grievous bodily harm (GBH)), car-jacking, bank robbery, gang-related violence, aggravated robbery, and so on [8, 10].

One major problem is that crimes are mostly unreported to the police, and about 10% of social media users would rather tweet/post stories/videos involving crimes [1, 11, 12]. It is a known fact that many crimes are not reported to law enforcement agencies in South Africa [13, 14]. The categories of crimes that are mostly unreported are commercial and white-colar crimes [14]. These crimes result in dropped charges, and fractionally few of these cases are successfully prosecuted and captured [14]. Budhram and Geldenhuys [14, 15] suggested that bribery is one of the reasons for some reported cases not getting added to the total numbers, and that the population does not trust the detection units of the police for these reasons.

Under-reporting of crimes is also an issue in other countries [16]. The United States homicide cases are under-reported for reasons such as misclassification; an example is classifying police shootings as *excessive use of force* or *account of lethal force* and not homicide [17]. There is also a huge disparity between the international homicide rates reported by the World Health Organisation (WHO), the Interpol, and the United Nations with WHO reporting the lowest — these are due to data availability, communication and regulation gaps among these authorities [18]. In South Korea, crime numbers are not released to the public at all, making it impossible for researchers and other bodies to quantify, estimate, or mitigate crime [19]. Crimes such as sexual harassment and rape are also under-reported [20]. A study conducted at the University of Texas showed that, even though there is low cases of reported sexual harassment, a survey showed that about 15% of female students claimed they had been raped [21] — these issues around under-reporting of crimes leaves the police and law enforcement agencies with skewed data [14].

However, research has proven that social media is a place for people to vent their anger [22, 23] and express several other emotions [24], mobilise protests [25] and when they are victims of crimes, publicize about their experiences [22]. We have evidences that suggests that it is not uncommon for people to rather tweet about their experience of crimes committed against them, rather than report it to the police [22]. There are also evidences to show that the police have been using social media as a source of intelligence and evidence gathering, monitoring riots, and analysing tension in neighbourhoods [26, 25].

A study showed that from the 32 million tweets retrieved from Twitter, 501,057 were deemed to be crime-related [22]. The Queensland Police Service (QPS) in the USA was able to convert up to 100 posts or comments made on social media pages into intelligence reports [22]. Hence, we know the data generated by users on these social media platforms contain a lot of information that is not in flat-police-files. In South Africa, a clear statement put on the South African police's Twitter handle (as retrieved on 16th of October 2020) encourages the citizens to report crimes via social media

Manuscript received September 26, 2022; revised May 22, 2023.

Stephen Obare is a PhD Candidate at Jomo Kenyatta University Agriculture and Technology, Kenya, e-mail: smobareo@gmail.com

Professor Abejide Ade-Ibijola is a Professor of Artificial Intelligence and Applications; Director: Postgraduate Diploma in Business Administration (with a Digital Transformation Lens) and Acting Academic Director at Johannesburg Business School, South Africa. e-mail: abejideai@uj.ac.za

Professor George Okeyo is an Assistant Teaching Professor at Carnegie Mellon University Africa, Rwanda. e-mail: gokeyo@andrew.cmu.edu

Dr. Kennedy Ogada is an Information Technology Consultant at Jomo Kenyatta University of Agriculture and Technology, Kenya. e-mail: kenogada@gmail.com

— an example of this can be seen in Figure 1. By mere inspection, one can literality read about crimes on social media. A screenshot of a citizen reporting crime activities in Johannesburg on the 16th of October 2020 can be seen in Figure 2.


Fig. 1: Twitter Handle of the South African Police

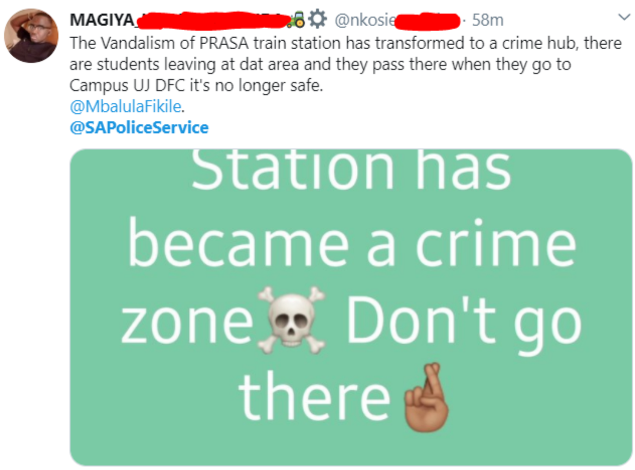


Fig. 2: A citizen reporting a hot crime zone in Johannesburg on 16th October 2020.

Indeed, social media has been instrumental in gathering data for the estimation of many different types of sentiments and trends [27, 28, 29, 30]. The method used for these estimations vary largely; from basic statistics methods to mathematical modelling, sentiment analysis algorithms and machine learning models, to formal methods.

One formal method (which has not seen many applications to real life problems) that can be used to parse unstructured texts such as tweets is representing a group of texts as a Jumping Finite Automaton (JFA). If this can be done, the question: "*does this tweet report a crime?*" can be reduced to a *decision problem* that can be answered by a JFA. A JFA is similar to a classical Finite Automaton (FA) (often represented as a deterministic or non-deterministic), with the exception that its transitions can jump around the input string, until the input is completely consumed and the JFA is in an accepting state. In this paper, we have designed several JFAs for the parsing of tweets and classified each parse as a *crime tweet* or otherwise. This classification is expected to provide an aggregated crime count that, in combination with reported crimes, should give a clearer picture of crime rates and identify crime hotspots in a society. Johannesburg tweets were used as a test case in this paper. The following are contributions of this paper. We have:

1) introduced JFAs as a new approach to parsing tweets,
2) designed many JFAs for the parsing of tweets and identification of crime tweets,
3) implemented the JFAs designed in (2) above, and presented results of crime classification using these JFAs, and
4) presented an evaluation of this approach, and argued that JFAs are can be used in classifying crime tweets.

The rest of this paper is organised as follows. Section 2 presents the background to this study. Section 3 presents the review of relevant literature. Section 4 presents the design of the JFAs used for parsing the tweets. Section 5 presents implementations of the JFA algorithms. Section 6 presents the evaluation and the last section presents the conclusions and future work.

## II. Background Information and Related Works

This section introduces the problem solved in this paper, and highlights the motivations behind this work. First, we start with a statement of problem addressed in this paper.

### A. *Problem Statement and Motivation*

We know that law enforcement agencies around the world are sitting with skewed crime statistics [12, 14, 17], due to unreported crimes. We also have evidences to show that crimes are often tweeted on social media by victims, using hashtag of law enforcements [22, 23, 24] and that the police are now using social media as a source of intelligence for monitoring criminal activities [26, 25]. It will be useful to *device techniques for the extraction of crime reports from social media*, thereby giving law enforcement a more complete report of crime incidences. The problem addressed in this work is how this can be done. We have presented a novel application of JFA to this task — *devised JFAs for estimating crimes, using Johannesburg (South Africa) tweets as a case study*. Next, we make a case by discussing the current crime numbers as reported in Johannesburg.

### B. *Crime in Johannesburg, South Africa and Around the World*

According to ISSAfrica in 2019, approximately half a million crimes (precisely 421,683) were reported (ranging from robbery, carjacking, and so on) in Gauteng province of South Africa, most happening in and around Johannesburg, the major city of this province [8]. In Johannesburg common crimes (according to Shaw and Haysom[31]) include: xenophobic hate crime [32] (mob beatings, riots, looting and burning of foreigned owned businesses), mafia and gang related organised crime (extortion, drug trading, and illegal weapons trade), housebreaking and burglary, rape [33], and gender-based violence [34]. These crime numbers are not particularly accurate because crimes are known to be underreported in South Africa [14] and in the world at large [18]. Violent crimes are so prevalent in South Africa that almost a third of all South Africa has recorded a crime classified as violent [35].

Crime affects many countries and cities around the world. Even though Johannesburg is regarded as a very dangerous city, a 2019 ranking of the *most crime infected countries and cities around the world* reveals that Caracas, Venezuela is the most dangerous city in the world, recording 119.87 homicides per 100,000 residents [36]. In 2020, Venezuela was reported to have a crime index of 84.86, the highest of any country in the world [37, 38]. The U.S. Department of State issued a Level 4 travel advisory for Venezuela, indicating that it is not safe to travel to the country and travellers should not travel there [37]. Another ranking places Los Cabos, Mexico as both the most dangerous city of the country and the world's most violent city after a rise in its homicide rate, as per Citizens' Council for Public Safety and Criminal Justice report [36].

### C. Why Crimes Are Not Reported to the Police?

A number of reasons has been cited in literature as to why people do not report crimes as follows:

1) social media users will rather tweet the crime [12],
2) sexually assaulted victims do not like to report because of the associated stigma [20],
3) most people often believe that the police will not do anything about the crime, especially when they consider the crime as minor [39], and
4) disbelieve in the police [40].

We know that many crimes are not reported but tweeted, hence, this work focuses on designing JFAs (a formal structure that has not seen any application in this domain) for the annotation of tweets as either a report of criminal activity or not. In the following section, we discuss why we have chosen JFAs for this task.

### D. Why JFAs?

Meduna and Zemek [41] expressed a reason why JFAs are useful for describing/parsing context-sensitive languages, and this stated in Remark 1.

*Remark 1 (**Why JFAs**? Meduna and Zemek [41]):* "*Traditional computer science methods were developed for information processing that was continuous. These methods included finite automata that represented information in a continuous manner from left-to-right, symbol-by-symbol way. However, modern information methods process information differently. A computational step can be executed in the middle or at the far end of the information, this would mean that the process will need to **jump** over information to get to the desired position. This is why jumping finite automata is preferred.*"

### E. Building Blocks and Definition of Terms

A formal background and the terms used in this paper are presented in this section, starting with the definition of a Jumping Finite Automaton (or JFA) and supported with an example. A JFA is a type of Finite Automaton. A finite automata is a very simple intelligent machine that is used to identify patterns from an input. The function of a finite automata is to reject or accept input depending if the defined pattern of the finite automata occurs within the input [42, 43]. A JFA works the same except they can read input data discontinuously – after reading a symbol, they can jump over other symbols within the words [41]. Hence, a JFA shares most attributes with a classical an FA with the additional ability of being able to read input symbols anywhere within the input string.

*Remark 2 (General Jumping Finite Automaton (GJFA)):* A **GJFA** is the most generic definition of jumping finite automata, as it presents an infinite class of JFAs, based on the number of symbols deleted in each step of computation [44]. Hence, a **GJFA** of *degree one* is referred to as a **JFA** — see Remark 3 for a formal description of a **JFA**. This paper is focused on GJFAs of degree of one, i.e. **JFA** only.

We proceed to present a formal definition of a General Jumping Finite Automaton (or GJFA). [A General Jumping

Finite Automaton or GJFA [41]] A GJFA is a quintuple such that $M = (Q, \Sigma, R, s, F)$ where:

1) $Q$ is a finite set of states,
2) $\Sigma$ is the finite set of input alphabets,
3) $R$ is the finite set of rules, where $py \longrightarrow q$ ($p$,$q \in Q$, $y {\in} \Sigma$),
4) $s {\in} Q$ is the start state, and
5) $F {\subseteq} Q$ is a set of one or more final states.

*Remark 3 (When is GJFA $\equiv$ JFA?):* Let $M$ be a GJFA. If all transition rules $py \longrightarrow q \in R$ satisfy $\mid y \mid \leq 1$, then $M$ is a Jumping Finite Automaton (JFA). The language accepted by such a JFA is denoted as $\mathcal{L}(M)$ [44].

There is a particular property of a JFA that makes it suitable for parsing crime tweets and using the result of the parsing for the estimation of crime.

*Remark 4 (Important Properties of a JFA):* As proven in Meduna and Zemek [41] and refined (with disproves of certain properties) in Vorel [44], the properties of JFAs that are generally acceptable/agreeable are:

1) **Union:** A JFA and a GJFA are both closed under the Union operation.
2) **Intersection:** A JFA is closed under the intersection operation, while a GJFA is not.
3) **Complement:** A JFA is closed under the complement operation, while a GJFA is not.
4) **Shuffle:** A JFA is closed under the shuffle operation. This is agreed on by both Meduna and Zemek [41] and Vorel [44] — even though Vorel [44] disputed other proofs presented by Meduna and Zemek [41] on Homomorphism and other few properties.

The most important property of a JFA to this work is the *shuffle* property — *referred to as the ability to read letters anywhere inside the input string* [45] or discontinous way method of reading data [46]. This is because it, after abstracting a context-sensitive language (such as a language of a crime tweets) into a JFA, becomes possible to shuffle and pick certain keywords (or their respective synomyms) from tweets and parse these tweets as crime tweets. An example of this process is presented in Example 1.

*Example 1 (Parsing a Simple Crime Tweet Using a JFA): In this example, we begin by abstracting certain keywords to symbols of an alphabet. Let $\Sigma = \{a_i, b_j, c_k\}$ be an alphabet, representing a three clusters of keywords. Where $\overline{a_i} = \{a_1, a_2, \ldots, a_i | i \in [\mathbb{N}]\}$, represent a set of places (names of places such as sub-burbs, malls, townships, etc.), $\overline{b_j} = \{b_1, b_2, \ldots, b_j | j \in [\mathbb{N}]\}$ is a set of crime types (such as hijacking, robbery, rape, etc.), and $\overline{c_k} = \{c_1, c_2, \ldots, c_k | k \in [\mathbb{N}]\}$ is a set of crime verbs (such as stab[bing], shoot[ing], kill[ing], etc.). Then we can define a formal language that is a set of words such that when combined in a tweet, it most probably will indicate that this tweet is a crime tweet by identifying the place the crime happened, the type of crime and the kind of crime action. Let $\mathcal{L}(M) = \{w \in \{a, b, c\} : |w_a|, |w_b|, |w_c| \geq 1\}$ be the language of crime tweets that is accepted by the JFA $M$, then the tweet:*

"@saps We have just been mugged at gun-point by two street guys at the corner of biccard and wolmanrans. Our wallets and phones were stolen."

Is acceptable by the JFA in Figure 3.

$\mathbf{M} = (\{S_0, S_1, S_2, S_3\}, \{\overline{a_i}, \overline{b_j}, \overline{c_k}\}, R, S_0, \{S_3\})$

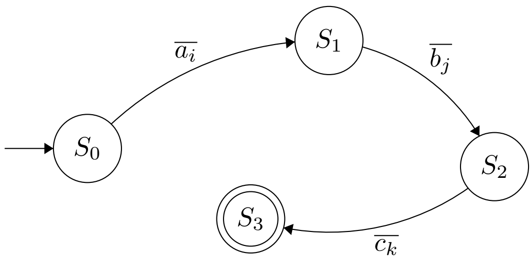


Fig. 3: JFA for the language of crime tweets that accept any combination for the language: $\mathcal{L}(M) = \{w \in \{a_i, b_j, c_k\}^* : |w_{a_i}|, |w_{b_j}|, |w_{c_k}| \geq 1\}$

The JFA $M$ accepts the $\mathcal{L}(M)$ because, given the sets: $a_i = \{biccard, wolmarans\}$, $b_j = \{mug[ged]\}$, and $c_k = \{stole[n]\}$; $M$ can reach an accepting state (i.e. $S_3$) on the string $b_2 a_1 c_2$ or a similar string $a_2 b_2 c_2$. The transitions from the start to accepting for $b_2 a_1 c_2$ are as follows:

$$\begin{aligned} b_2 S_0 a_1 c_2 &\curvearrowright S_1 b_2 c_2 && [S_0 a_1 \longrightarrow S_1] \\ &\curvearrowright S_2 c_2 && [S_1 b_2 \longrightarrow S_2] \\ &\curvearrowright S_3 && [S_2 c_2 \longrightarrow S_3] \end{aligned}$$

Example 1 shows how we have constructed JFAs to parse tweets and estimate if (or not) the tweet is a crime tweet or not; i.e. solve the decision problem, "*does this tweet talk about crime*?".

### *F. Data Source and Ethics*

The source of the dataset used in this paper is Twitter (www.twitter.com). This platform comes highly recommended by researchers that have carried out similar tasks such as crime estimation from social media data [47]. Twitter is the most used social media platform in the world with public access to its data, contributed by millions of users in real time [47]. As regards ethics, the dataset is available in the public domain and the computation that we have carried with JFAs have produced aggregated results — i.e. no person or individual was singled out in our results.

### *G. Summary of Background*

In this section we have presented the problem and justified why it is an important problem to solve, and why we think a JFA is appropriate for the solution. We have also presented the review of related work, the gap in existing solutions, an overview of our new technique, our source of data and we have addressed the ethics questions around the dataset used. Technical terms used are also defined here. In the next section, we present the design of the actual JFA used for parsing crime tweets.

## III. Related Work

In this section we present related works to the this work, starting by the research works that have attempted to extract sentiment of a population from the analysis of the tweets extracted from this population. In this case, social media is the population of interest, with hashtags used for demarcating topologies between localities, towns, and countries.

### *A. Extracting Sentiments from Social Media*

Social media has been instrumental in estimating many sentiments or trends. In China, social media data has been used to monitor air quality [48]. Malleson and Andresen [27] used a comparison of mobile referenced criminal events from residents with spatially referenced criminal events and reported that on both cases, the same geographical hotspots were outlined. The authors have also suggested that the use of social media data is very promising for checking the implications of regular activities and the theories of crime distribution, due to the explicit spatial and temporal nature of these activities [27].

A similar attempt to derive patterns from social media data was presented in the work of Hipp *et al* [28] where they predicted hit-and-run cases using semantic analysis to understand Twitter posts. The authors used Latent Dirichlet Allocation for dimensionality reduction in a linear model for prediction. Other indications that social media data can help was reported in the work of Aghababaei and Makrehchi [30], who specified that social media data and its associated context often produces behavioural signals for crime predictions. In the next section, we present attempts to estimate crime with respect to tools, techniques or algorithms.

### *B. Existing Tools and Techniques for Estimating Crimes*

Here we present two aspects (with focus on technological tools and techniques) to arriving at crime numbers, namely:

1) estimation of crimes based on reported crimes and information extraction from other sources (such as social media), and
2) prediction of crimes (using predictive models or algorithms).

In earlier subsections we have explained why it is important to have unskewed crime estimations, and how this paper presents an approach that uses JFAs to estimate unreported crimes using tweets. Another aspect that is included here is the prediction of crime. *Why do we have to predict crime?* A number of related works have shown that crimes are not randomly distributed in society and hence, the study of the patterns of crime, and development of prediction models are of paramount importance [49, 50, 51].

We hereby present a number of related work on crime predictive models or algorithms. Wang et al [49] proposed solutions to making inferences on big data to arrive at crime rate estimations. In the work of Ingilevich Varvara and Ivanov Sergey [50], crime rate was also predicted based on social factors such as employment rates, population density of cities, number of homeless people, and so on. Spatial analysis was applied in monitoring crime by ToppiReddy et al [51]. A company based in Cape Town, South Africa named the *Solution House Software* have developed an AI tool that outputs potential threats before the associated incidents occur and it is done through studying of where and when crime normally occurs. The tool uses aggregated data from multiple sources of information to determine the likelihood of criminal activities, including variables such as weather patterns for the determination of the crime hotspots [52].

Wang and Brown [29] approached crime modelling in a different fashion. They used all criminal incidents as training data within a predictive model. Geographic locations are denoted by a rich set of spatial and demographic features. This representation allowed them to predict crimes in previously unseen places. A tool was developed by ShotSpotter (www.shotspotter.com) (a USA company) that triangulates the location of a gunshot, after this company realised that only 20% of gunshots are reported to the police with inadequate information for the police to establish a lead. This tool reports gunshots to the police in real time with details such as where the gun fired (to 10 feet precision) and the type of gun that was fired. This tool is in use by law enforcement agencies in more than 90 cities around the world. In 2019, Cape Town, South Africa was added to the least of cities that use this tool [52].

Several other methods for predicting crime has been proposed. A method based on data mining was presented by Sathyadevan et al [53], a predictive policing technique was presented in Wang et al [54], a predictive crime mapping technique presented in Rosser et al [55], a hotspot maps centered approach presented by Gerber [47], an ensemble machine learning model was presented by Almaw and Kadam [56]. Hayward and Mass [57] made a list of other aspects of Artificial Intelligence that is increasingly finding application in crime prediction or analytics — this list includes:

1) video surveillance (i.e. facial recognition) — distinguishing micro-facial expression to determine and identify deception and emotion,
2) voice recognition (i.e. echo location), an example is the *Speech2Face* algorithm, an AI program that takes the voice audio of a person and attempts to estimate their facial images, age, gender and ethnicity,
3) recognise body language, and this has been used to detect possible shoplifters,
4) recognising unique body signatures based on laser technology. This is currently being used at the pentagon, USA to identify unique heartbeat signatures from about 200 metres range, and
5) anticipate and/or predict situations. This is currently being used by Axon Enterprise (an American company) to analyse US police body camera videos and produce reports relating to anticipations.

A number of similar AI techniques in crime prediction were presented in Khairuddin et al [58].

### C. Real Life Applications of JFAs

Due to the relatively new nature of JFAs, they have not been applied in many domains. In fact, they have only been applied in *one* real life problem. This application can be found in the work of Obare et al [59]. To the best of our search and knowledge, this is the second time JFAs will be applied in a real life problem.

### D. Gap and Proposed Solution

Given that the prediction of crime will continuously find applications in crime inflicted areas, such as the one we are looking at in this paper, there exist a need to keep finding ways of performing this task. In this paper, we have used JFAs as a new approach to comprehend tweets, and to classify tweets as either reporting a crime or not. This is expected to support law enforcement's estimation of the crime numbers, as well as allow for opportunity for intervention. In the next section, we present the design of JFAs for comprehending and estimating crime tweets.

## IV. JFA Design

In the previous sections we have made a case for the first application of JFAs in estimating crime rates. We have argued the continuous need for this task to be performed, even though it has been done with many different statistical or mathematical techniques in previous related works. In this section, we have presented the design of the actual JFAs use for the estimation of crimes, based on tweet data extracted from Twitter. First we describe the alphabets from which the symbols of the JFAs are picked, and the input strings of the language constructed. This is discussed in Section IV-A.

### A. Alphabet Look-up

In this section we present the alphabet for the JFAs designed in this work. Let $M$ be a JFA for *recognising a crime tweet*, then we take the each tweet to be parsed as a *string*, and each words in the tweet as a *symbol* in $\mathcal{L}(M)$. Furthermore, we granulate the alphabet (set of symbols for $\mathcal{L}(M)$) into four subsets: $\Sigma_a, \Sigma_b, \Sigma_c$, and $\Sigma_d$, where $\Sigma_M = \Sigma_a \cup \Sigma_b \cup \Sigma_c \cup \Sigma_d$. The description of the word composition of these subsets are as follows: $\Sigma_a$ is the set of places and landmarks in the Gauteng province of South Africa, $\Sigma_b$ is the set of crime keywords, $\Sigma_c$ is a set of objects involved in a crime, and $\Sigma_d$ is a set of crime hashtags. $\Sigma_d$ is an optional alphabet that may not be included in the design of a JFA — *examples are presented in this section.*

The symbols in these subsets are specified as follows: $\Sigma_a$ is the set shown in Table I, $\Sigma_b$ in Table II, $\Sigma_c$ in Table III, and $\Sigma_d$ in Table IV.

TABLE I: Places and landmarks, $\overline{a_i}$, in the Gauteng region of South Africa

| $\overline{a_i}$ | Place | $\overline{a_i}$ | Place | $\overline{a_i}$ | Place | $\overline{a_i}$ | Place | $\overline{a_i}$ | Place | $\overline{a_i}$ | Place | $\overline{a_i}$ | Place |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | Actonville | 21 | Cleveland | 41 | Ennerdale | 61 | Kagiso | 81 | Mamelodi East | 101 | Pretoria West | 121 | Sunnyside |
| 2 | Akasia | 22 | Crystalpark | 42 | Erasmia | 62 | Kameeldrift | 82 | Meadowlands | 102 | Primrose | 122 | Tarlton |
| 3 | Alberton | 23 | Cullinan | 43 | Etwatwa | 63 | Katlehong | 83 | Meyerton | 103 | Protea Glen | 123 | Temba |
| 4 | Alexander | 24 | Daveyton | 44 | Evaton | 64 | Katlehong North | 84 | Midrand | 104 | Putfontein | 124 | Tembisa |
| 5 | Atteridgeville | 25 | Dawn Park | 45 | Fairland | 65 | Kempton Park | 85 | Moffatview | 105 | Rabie Ridge | 125 | Tembisa South |
| 6 | Bedfordview | 26 | De Deur | 46 | Florida | 66 | Khutsong | 86 | Mondeor | 106 | Randburg | 126 | The Barrage |
| 7 | Bekkersdal | 27 | Devon | 47 | Fochville | 67 | Kliprivier | 87 | Moroka | 107 | Randfontein | 127 | Tokoza |
| 8 | Benoni | 28 | Diepkloof | 48 | Ga-Rankuwa | 68 | Kliptown | 88 | Mudlersdrif | 108 | Ratanda | 128 | Tsakane |
| 9 | Boipatong | 29 | Diepsloot | 49 | Garsfontein | 69 | Krugersdorp | 89 | Naledi | 109 | Roodepoort | 129 | Vaal Marina |
| 10 | Boksburg | 30 | Dobsonville | 50 | Germiston | 70 | Kwa-Thema | 90 | Nigel | 110 | Rosebank | 130 | Vanderbijlpark |
| 11 | Boksburg-North | 31 | Douglasdale | 51 | Hammanskraal | 71 | Langlaagte | 91 | Norkempark | 111 | Sandringham | 131 | Vereeniging |
| 12 | Booysens | 32 | Dube | 52 | Heidelberg | 72 | Laudium | 92 | Norwood | 112 | Sandton | 132 | Villieria |
| 13 | Boschkop | 33 | Duduza | 53 | Hekpoort | 73 | Lenasia | 93 | Olievenbosch | 113 | Sebenza | 133 | Vosloorus |
| 14 | Brackendowns | 34 | Dunnottar | 54 | Hercules | 74 | Lenasia South | 94 | OR Tambo International Airport | 114 | Sebokeng | 134 | Wedela |
| 15 | Brakpan | 35 | Edenpark | 55 | Hillbrow | 75 | Linden | 95 | Orange Farms | 115 | Sharpeville | 135 | Welbekend |
| 16 | Bramley | 36 | Edenvale | 56 | Honeydew | 76 | Loate | 96 | Orlando | 116 | Silverton | 136 | Westonaria |
| 17 | Brixton | 37 | Eersterust | 57 | Ivory Park | 77 | Lyttleton | 97 | Parkview | 117 | Sinoville | 137 | Wiedaburg |
| 18 | Bronkhorspruit | 38 | Ekangala | 58 | Jabulani | 78 | Mabopane | 98 | Pretoria Central | 118 | Sophia Town | 138 | Wonderboom-poort |
| 19 | Brooklyn | 39 | Eldorado Park | 59 | Jeppe | 79 | Magaliesberg | 99 | Pretoria Moot | 119 | Soshanguve | 139 | Yeoville |
| 20 | Carltonville | 40 | Elsburg | 60 | Johannesburg Central | 80 | Mamelodi | 100 | Pretoria North | 120 | Springs | 140 | Zonkizizwe |

TABLE II: Crime keywords, $\overline{b_j}, j \in [40]$.

| $\overline{b_j}$ | Crime Keyword | $\overline{b_j}$ | Crime Keyword | $\overline{b_j}$ | Crime Keyword | $\overline{b_j}$ | Crime Keyword |
|---|---|---|---|---|---|---|---|
| 1 | ammunition | 11 | fraudulent | 21 | mugging | 31 | steal |
| 2 | assassinate | 12 | gun | 22 | murder | 32 | stole |
| 3 | arrest | 13 | gunfire | 23 | person | 33 | stolen |
| 4 | bullet | 14 | hi-jack | 24 | riffle | 34 | suspect |
| 5 | bullets | 15 | hijack | 25 | rob | 35 | hitman |
| 6 | crime | 16 | hijacking | 26 | shoot | 36 | theft |
| 7 | criminal | 17 | kill | 27 | shooting | 37 | trafficking |
| 8 | criminals | 18 | killed | 28 | shot | 38 | victim |
| 9 | dead | 19 | knife | 29 | stab | 39 | victims |
| 10 | fraud | 20 | mug | 30 | stabbing | 40 | violence |

TABLE IV: Hashtags for crime tweets, $\overline{d_l}, l \in [20]$.

| $\overline{d_l}$ | Hashtag | $\overline{d_l}$ | Hashtag |
|---|---|---|---|
| 1 | #ActAgainstAbuse | 11 | #Missing |
| 2 | #ATMSafety | 12 | #MySAPSAPP |
| 3 | #CIT | 13 | #PoliceKillings |
| 4 | #CITRobbery | 14 | #PyramidSchemes |
| 5 | #CrimeStop | 15 | #ReportCrimes |
| 6 | #DrugsOffTheStreets | 16 | #ResponsiblefirearmUse |
| 7 | #ENDGBV | 17 | #RuralSafety |
| 8 | #Escaped | 18 | #SearchAndRescue |
| 9 | #GunsOffTheStreets | 19 | #StockThefts |
| 10 | #IllegalMining | 20 | #TipOFF |

TABLE III: $k$ objects $\overline{c_k}$, on/in which crime takes place, $k \in [10]$.

| $\overline{c_k}$ | Object | $\overline{c_k}$ | Object |
|---|---|---|---|
| 1 | car | 6 | phone |
| 2 | cars | 7 | truck |
| 3 | cash | 8 | van |
| 4 | home | 9 | vehicle |
| 5 | house | 10 | bod[y,ies], murder |

### B. *JFA for Crime Recognition*

Here we rely on the *shuffle property* (described in Remark 4) of JFAs to parse a tweet, such that, if the JFA consumes all keywords in the input string and ends up in an accepting state, we conclude that such a JFA accepts the string, and that this string, in fact, speaks about a crime. We begin by designing a JFA (using the previously set alphabet $\Sigma_M$) that recognises crime based on the union of $\Sigma_a$, $\Sigma_b$ and $\Sigma_c \subset \Sigma_M$.

Let $\mathcal{L}(M_1) = \{w \in \{a,b,c\}^+ : |w_a|, |w_b|, |w_c| \geq 1\}$ be the language of crime tweets that is accepted by the JFA $M_1$, then we present $M_1$ as follows:

$$M_1 = (\{S_0, S_1, S_2, S_3\}, \{\overline{a_i}, \overline{b_j}, \overline{c_k}\}, R, S_0, \{S_3\})$$

Graphically illustrated in Figure 4.

Figures 5 and 6 are real life tweets that are acceptable by the JFA, $M_1$. We choose to extract tweets from the @SA Police Service hashtag in Figures 5 and 6 because tweets that contain hashtags are more likely to be relevant than tweets that do not contain hashtags. The tweets are also semantically rich because of the number of entities that can be extracted from the tweet.

JFA $M_1$ accepts this tweet, for alphabet subsets, $\{KwaMsane\} \subseteq \Sigma_a$, $\{hitm[an, en], arrest, murder\} \subseteq$

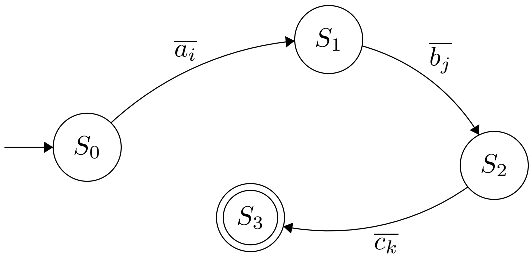


Fig. 4: JFA for crime tweets that accept any combination of of strings in the language: $\mathcal{L}(M) = \{w \in \{a_i, b_j, c_k\}^* : |w_{a_i}|, |w_{b_j}|, |w_{c_k}| \geq 1\}$


Fig. 5: Sample crime tweet reported in court proceedings.

$\Sigma_b$, and $\{murder \equiv bod[y, ies]\} \subseteq \Sigma_c$.

Both tweets in Figures 5 and 6 have a number of entities that makes it easy to be processed by JFA. For example, the entities in Figure 5 include $\{KwaMsane, arrest, murder, hitman, bodies$ etc while the tweet in Figure 6 contains the following entities: $\{Grabouw, Overberg, firearm, stole, house$ etc. By inspection, these entities form part of the alphabet look-up.


Fig. 6: A crime tweet.

JFA $M_1$ accepts this tweet, for alphabet subsets, $\{Grabouw, Overberg\} \subseteq \Sigma_a$, $\{firearm, stole[n], arrested, suspect\} \subseteq \Sigma_b$, and $\{house\} \subseteq \Sigma_c$.

JFA $M_1$ can be extended to be more accurate by including $\Sigma_d$, the sub-alphabet that takes care of crime hashtags. This assures a four-stage affirmation that a matching tweet speaks about a crime. Hence, we proceed to define another JFA, $M_2$, that makes use of the four subsets of $\Sigma$ as described in Section IV-A.

$M_2$ accepts a subset of the language accepted by $M_1$, i.e. $\mathcal{L}(M_1) \subseteq \mathcal{L}(M_2)$. Given the nature of hashtags (they are often topics themselves), we speculate that this definitely narrows the tweets down to specific type of crime tweets. A few times when this can be wrong (such as online marketers and trollers who use hashtags to get visibility, leading to false positives) are discussed later in this paper in the section on evaluation of the JFAs.

### C. *JFA for Crime, Place, and Category Recognition*

So far, we have presented JFAs that used four sub-alphabets to recognise tweets that contain crime reports. In this section, we describe how subsets of $\overline{a_i}$ can be used to categorise the recognised crimes into sub-domains of

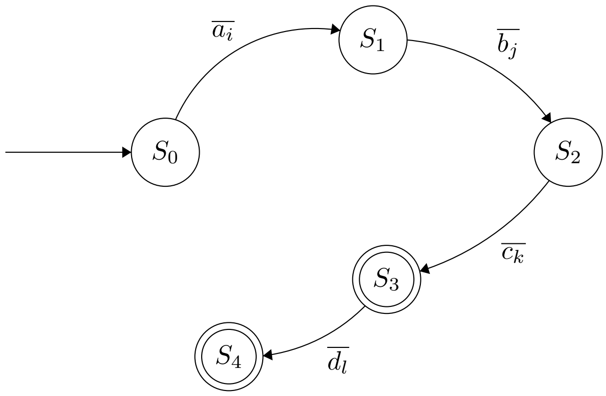


Fig. 7: JFA for the language of crime tweets that accept any combination for the language: $\mathcal{L}(M) = \{w \in \{a_i, b_j, c_k\}^* : |w_{a_i}|, |w_{b_j}|, |w_{c_k}| \geq 1\}$.

locations. Let $M$ be a JFA for the language of crime tweets, over the alphabet $\Sigma_M$, and let $\Sigma_a$ be a set of places and landmark (similar to the ones presented in Table I), we can further define sub-subsets of $\Sigma_M$, i.e. $\Sigma_{a_n}$. These sub-subsets can be used in defining new JFAs that are targeted towards recognising crime activities in specific areas, i.e. defining symbols (keywords for places) $a_i \in n$ for $\Sigma_{a_n}$.

### D. *Enhancing JFAs with Preprocessors and Normalisers*

Millions of users post thousands of tweets every minute. These tweets also have spelling errors, and since the JFA looks out for keywords, it is important to normalise these tweets before passing them to JFAs. Hence, we introduced a pre-processing module at the implementation stage based on the Levenshtein Distance algorithm to tackle spelling errors. We also introduced an algorithm for filtering unnecessary words (such as conjunctions). Finally, these algorithms are informed by a synonym database — making sure that a similar words in the alphabets can fire the JFA, even when the exact words in the alphabets are not found in the tweet.

In the next section, we present the implementation of the JFAs described in the section.

## V. Implementation, Results and Applications

In this section, we present the implementation of the JFAs presented in this paper. First, we describe the algorithm for simulating the transitions of our crime JFA for any given input string.

### A. *Algorithm for Crime JFAs*

Here we present the algorithm for an implementation of a *Parser* for the generalised form of the JFA, $M$, that uses $k$-sub-alphabets to recognise and classify tweets as either reporting crime or not. Algorithm 1 takes as parameters: a JFA M, an alphabet $\Sigma$ specified as a list of lists, each item in this list, representing a list of sub-alphabets for bag of word categories, such as places, crime action words, etc., and a piece of tweet text. The tweet is first preprocessed (on Line 2 with the `Preprocess` function) as discussed in Section IV-D and the outcome of the preprocessing stage is stored as a list of keywords in a list object, `T_strings`.

The current transition state of the JFA $M$ is initialised to zero on Line 2, before the simulation of the transitions begin. A loop that traverses the sub-alphabets of $\Sigma$ starts on Line 3. This loop contains three selective conditions:

1) if the JFA $M$ is at its final state and there are no more sub-alphabets to step through — in this case, $M$ has accepted the tweet string, and outcome is returned as success (see Lines 4 to 6),
2) if there exist a reason for the JFA to transition from one state to another — this is computed by checking if the current sub-alphabet has as similar symbols as the string of the tweet (see Line 7 to 10). The `List_Diff()` function returns the difference of two lists (i.e. A – B) the secondary parameter is subtracted from the first one. Line 9 makes sure that the symbols of the current sub-alphabet is "*popped*" from the tweet's string — since this JFA recognises one or more occurrences of the sub-alphabets, and
3) the last condition (the else part on Line 11) is the only possibility that makes this JFA reject this string of crime tweet. In this case, outcome is set to failure, and the loop is exited to return the final outcome.

Algorithm 1 saves some computational time by exiting loops on finding a any symbol that transitions the JFA from one state to another as seen on Line 6, and more time is saved as well on Line 10 to make the JFA continue with its parsing if at least a symbol has been seen in the sub-alphabet and the input tweet string, that allows for transition to the next state. Finally, a failed transition should not allow the loops to run, and this is managed on Line 13.

In the next section, we present how Algorithm 1 was implemented.

### B. Implementation and Results

The technique for parsing tweet strings using the JFA design proposed in this paper was presented in Algorithm 1. This algorithm was implemented with Python programming language (specifically Python 3). The implementation was in two phases:

1) **Data (Tweets) gathering**: We used Twitter API to crawl 10,000 tweets, and
2) **Parsing**: We implemented the JFA parsing algorithm as a piece of software written in Python and called this tool the `Crime-Ripper`. This application takes a JFA (specified as a list of strings in a text file, denoting alphabets) and a tweet string; and returns a success or failure if the JFA accepts the tweet or not respectively.

We ran our JFA Crime Tweets Parser through all the 10,000 tweets crawled, and the performance evaluation of the JFA Parser is discussed in Section VI.

## VI. Evaluation and Applications of Crime-Ripper

To evaluate the performance and/or accuracy of the `Crime-Ripper`, we had to get 45 participants that are very familiar with social media and understand what a crime reporting tweet should look like. These participants labelled the 10,000 tweets as either talking about crimes or not. The labelling exercise was *automated in parts* — where similar tweets to tweets labelled by humans as positive, were also labelled automatically as positive. This process was preceded by a survey of expert opinion or perception of this approach as either appropriate for crime estimation.

### A. Expert Perception of Crime-Ripper

As an initial part of the evaluation, we conducted an expert survey, consisting of 73 participants (all having a minimum of a graduate degree in Computer Science or Engineering, and have knowledge of algorithms in general).

TABLE V: Result of survey

| **Participants (73)** | **Yes** | **No** | **Maybe** | **%** |
|---|---|---|---|---|
| Knowledge of JFAs | 61 | 12 | 0 | 83.561 |
| Understanding of techniques | 68 | 5 | 0 | 93.150 |
| Perceived use of technique for crime estimation | 58 | 10 | 5 | 79.452 |

This survey was designed to measure the perception of these participants on the use of JFA algorithms (as presented in this paper) as technique for estimating crime using tweets. Table V shows the results of the survey.

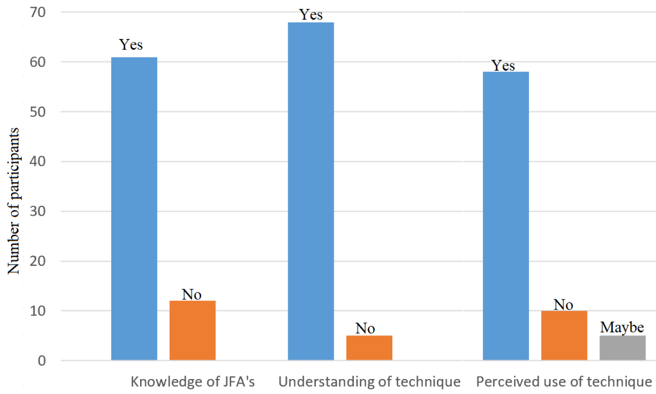


Fig. 8: Summary of survey.

Participants were asked if they had the knowledge of JFAs and if they understood the technique in general. 83.5% (61 of 73) declared that they understood JFAs in theory, while a larger percentage, 93.1%, said they understood the technique from an algorithmic standpoint as depicted in as shown in Fig. 8. The number of of respondents surveyed who do not have knowledge of JFAs and also do not understand the technique is an indication that the knowledge and understanding of JFA is good among participants. No participant responded that maybe they have knowledge of JFAs or its perceived use in estimating crime.

In total, 79% of the population perceived the use of this technique as a good solution for estimating crime based on tweets. Only 21% of the participants perceived JFAs as not usable for estimating crime is an indication that the knowledge and understanding of JFA is good among participants.

In general, a majority in our sample had a good knowledge of JFAs, understood the technique and perceived its use in crime estimation as shown in Fig. 9.

### B. Accuracy of Crime-Ripper

The distribution of the labelled tweets used for evaluation are as follows: 6,871 tweets are labelled by participants as

**Algorithm 1** JFA Parser for Crime Tweets.

**Data:** $M$: A JFA that accepts only crime tweets.
**Data:** $\Sigma = \{\Sigma_a, \Sigma_b, \ldots, \Sigma_n\}$: A list of lists (or matrix) containing all sub-aphabets of $\Sigma$.
**Data:** $T$: A tweet text to be parsed, or checked.
**Result:** outcome: *success* (if JFA accepts T, or *failure* if otherwise).
T_strings = `Preprocess` ($T$); `/* returns all strings in the tweet after preprocessing. */`
$M$.current_state = 0 **for** *(sub_alpha $\in$ Sigma)* **do**
  **if** *(`isFinalState` (M.current_state) $\wedge$ index (sub_alpha) == MAX (Sigma))* **then**
    outcome = success **exit for**
  **else if** *(sub_alpha $\cap$ T_string $\neq$ $\emptyset$)* **then**
    $M$.current_state += 1 T_strings = `List_Diff` (T_strings, sub_alpha) **continue** sub_alpha ; `/* moves to next sub-alphabets */`
  **else**
    outcome = failure **exit for**
  **end**
**end**
**return** outcome

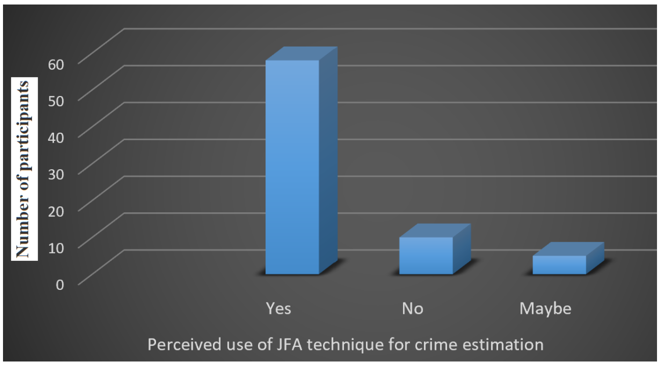


Fig. 9: Perceived use of technique for estimating crime.

talking about crime, while 3,129 are labelled as negative — not talking about crime. The 6,871 tweets labelled as crime tweets were presented to `Crime-Ripper` and 6,446 of these tweets was confirmed by `Crime-Ripper` as to be talking about crimes. This gives an accuracy of 93.81% $\approxeq$ 94% for the *true positives* — leaving 425 tweets as *false positives*, $\approxeq$ 6%. The tweets labelled by human participants as negatives (not about crimes) were given to `Crime-Ripper`, and 2876 tweets were rightly categorised as true negatives out of 3129; resulting in 91.91% $\approxeq$ 92%. This gives a true negative of $\approxeq$ 8%.

In this experimental set-up, `Crime-Ripper` was seen to perform very well. This can be attributed to the fact that tweets around crime within the context of this work is not totally unstructured. There is a recurring patten of place, type of crime, etc., and when these parameters are well modelled into sub-alphabets for a JFA, it becomes possible to pick up most of the true crime tweets, and also, rightly determine the tweets that are not crime tweets.

### C. Complexity of Algorithm

Here we comment on the complexity of the algorithm that implements the JFA Parser (i.e. Algorithm 1). The preprocessing task carried out with the function on Line 1 is the removal of unnecessary words from the tweet keywords, and the correction of spelling mistakes (based on Levenshtein Distance algorithm) — all these tasks are carried out in linear time, corresponding to the number of tweets; i.e. $O(n)$. Line 3 introduces another loop. This second loop contains a number of functions (such as checking if the final state is reached on Line 4, and the difference of two lists on Line 9) — all these functions are executable in constant or linear time, as none of them grows fast, with respect to the input size. Hence, we conclude that this algorithm has a complexity of $O(n)$.

During implementation, python offers many libraries for these functions, hence, the program execution is optimised.

### D. Limitations

We have not considered instances where the users have posted crime tweets in local languages, used a lot of vernacular that skews the message in these tweets, used abbreviated names for places or crime tokens, or other deeper embedded semantics such as slangs that are local to specific environments. By inspection, we realised that the false negatives were mostly instances that are of this nature — *these are limitations of this work.*

### E. Applications of Estimated Crimes

There are several applications of the JFA technique for recognising crime tweets presented in this paper. Here we list three major applications (that are within the context of crime reporting) as this is the theme of this paper, as follows:

1) Crime reports and alert in real time for law enforcement,
2) Crime reports for agencies to argument the current skewed data from unreported crimes, and
3) Hotspot identification and/or map annotation.

## VII. Conclusion and Future Work

In this section, we present the conclusion and future work.

### A. Conclusion

Predictive policing and predictive crime mapping have been proven to be very helpful [60, 61], with the police around the world beginning to engage in social media intelligence [62]. In this paper, we have proposed a new technique for the automatic comprehension of tweets (with a specific interest in crime tweets — tweets that report criminal activities), based on jumping finite automata, a type of finite automata that allows symbols of the input string to be read from random positions in the string. We have also designed three different JFAs for comprehending tweets, using bags of sub-alphabets of the alphabet for a JFA for the grouping of aspects that tweets that denote specific topics, such as place, crime action words, objects involved in the crime, and crime type hashtags.

We have designed and algorithm that implements a generalised form of the JFA for crime recognition, and presented a tool that uses this algorithm, named the `Crime-Ripper`. `Crime-Ripper` iterates through already labelled tweets and attempts to categorise these tweets as either crime tweets or otherwise — having successes in the range of 92% to 94% with minimal/acceptable errors. We pointed out that the advantages of the JFA method is that it allows different subtopics of the tweets to be recognisable using sub-alphabets of the main alphabet of the JFA. This also allows for anyone to design new JFAs for other types of tweets, with extensible alphabets as per the needs of their application (i.e. we can have more than four sub-alphabets).

We analysed the complexity of the algorithm for the JFA parser, and is computationally inexpensive, allowing for a linear time execution. Limitations that we have identified include the use of *slangs* and abbreviation by users on twitter that can skew the results. Applications of this technique is majorly in crime reporting and the annotation of maps to show crime areas. Knowing that the language of tweets, *in truth*, is semi-structured at some level of detail, we hypothesise that the JFA technique was able to perform exceptionally well (as seen in this paper) because of the task of crime identification that comes with a structured way of which users often report crime.

### *B. Future Work*

Following the results in this paper, future work in this space will include: the automatic building of dynamic JFAs from keyword repositories and exploring which JFAs do better based in an iterative space search algorithm. This will involve the steps of building JFAs from pre-defined alphabets, computing their accuracy, and ranking the automatically built JFAs by their accuracy — we might just find a combination of sub-alphabet that does well on tweets better than others. We will also explore other applications of JFAs for the automatic comprehension of (or information extraction from) texts.

## References


[1] Thitiphan Semangern, Wachit Chaisitsak, Twittie Senivongse, SI Ao, C Douglas, and HK Kim. Identification of risk of cyberbullying from social network messages. In *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science, San Francisco, USA*, pages 276–282, 2019.

[2] Hsi-Chuan Wang. Prioritizing compactness for a better quality of life: The case of us cities. *Cities*, 123:1–5, 2022.

[3] Anthony Abayomi Adebayo. Youths' unemployment and crime in nigeria: A nexus and implications for national development. *International journal of sociology and anthropology*, 5(8):350–357, 2013.

[4] Ted Enamorado, Luis F López-Calva, and Carlos Rodríguez-Castelán. Crime and growth convergence: Evidence from mexico. *Economics Letters*, 125(1):9–13, 2014.

[5] Anthony Kaziboni, Lizette Lancaster, Thato Machabaphala, and Godfrey Mulaudzi. Scapegoating in south africa: Busting the myths about immigrants. *ISS Southern Africa Report*, 2022(53):1–16, 2022.

[6] Gugulethu S Nkosi. The impact of crime on tourism in the city of umhlathuze, kwazulu-natal. *South Asian Journal of Tourism and Heritage*, 3(2):76–81, 2010.

[7] Irma Booyens. Rethinking township tourism: Towards responsible tourism development in south african townships. *Development Southern Africa*, 27(2):273–287, 2010.

[8] K. Lomahoza. The state of crime and safety in the city of joborg, 2019. City of Johanneburg. [Online]. Available: https://www.joburg.org.za/documents/Documents/Statistical.

[9] Likhona Mihlali Mlamla, Mary S Mangai, Tyanai Masiya, and Natasja Holzhauseni. Stakeholders' experience of the innovative ways of co-producing neighborhood security in johannesburg, south africa. *Technium Soc. Sci. J.*, 31:44, 2022.

[10] David Abrahams. A synopsis of urban violence in south africa. *International review of the Red Cross*, 92(878):495–520, 2010.

[11] Kathrin Leppert, Iris Saliterer, and Sanja Korać. The role of emotions for citizen engagement via social media–a study of police departments using twitter. *Government Information Quarterly*, 39(3):1–6, 2022.

[12] Rafael Prieto Curiel, Stefano Cresci, Cristina Ioana Muntean, and Steven Richard Bishop. Crime and its fear in social media. *Palgrave Communications*, 6(1):1–4, 2020.

[13] Laurie Greyvenstein. Mediation & the law victims of crime and adr part 2. *Servamus Community-based Safety and Security Magazine*, 115(5):70–71, 2022.

[14] Trevor Budhram and Nicolaas Geldenhuys. A losing battle? assessing the detection rate of commercial crime. *South African Crime Quarterly*, 61:7–18, 2017.

[15] Altje Agustin Musa, Jemmy Sondakh, Wempie Jh Kumendong, and Caecilia JJ Waha. Discretion of government officials in the perspective of corruption crime reviewed from the theory of criminal removal reason. *Journal of the Community Development in Asia (JCDA)*, 5(1):91–102, 2022.

[16] Kris Henning. Using data to improve public safety with kris henning. pages 1–4, 2022.

[17] Andrew C Gray and Karen F Parker. Race, structural predictors, and police shootings: Are there differences across official and "unofficial" accounts of lethal force? *Crime & Delinquency*, 65(1):26–45, 2019.

[18] Catrin Andersson and Lila Kazemian. Reliability and validity of cross-national homicide data: A comparison of un and who data. *International Journal of Comparative and Applied Criminal Justice*, 42(4):287–302, 2018.

[19] Seok Man Lee and Hyun-Ho Kim. A problem of crime control policy in south korea: A challenge of asymmetric information. *Journal of Scientific Research and Reports*, 26(3):80–85, 2020.
[20] Yingyu Chen and Sarah E Ullman. Women's reporting of sexual and physical assaults to police in the national violence against women survey. *Violence against women*, 16(3):262–279, 2010.
[21] David F Prindle. The university of texas "rape survey": A case study of politicized social science. *Academic Questions*, 31(1):77–83, 2018.
[22] S. Meixner. Police "frustrated' at people posting about suspicious activity on facebook without notifying law enforcement, 2019. ABC.NET News. [Online]. Available: https://www.abc.net.au/news/2019-06-22/reporting-crimes-on-social-media-not-helpful-say-police/11219404/.
[23] Gregory Gondwe and Evariste Some. Social media and catharsis in africa: Examining the role of whatsapp in venting stress in women. *Journal of Internet and Information Systems*, 9(1):1–7, 2020.
[24] Anne Vermeulen, Heidi Vandebosch, and Wannes Heirman. Smiling, venting, or both? adolescents' social sharing of emotions on social media. *Computers in Human Behavior*, 84:211–219, 2018.
[25] Matthew L Williams, Adam Edwards, William Housley, Peter Burnap, Omer Rana, Nick Avis, Jeffrey Morgan, and Luke Sloan. Policing cyber-neighbourhoods: tension monitoring and social media networks. *Policing and society*, 23(4):461–481, 2013.
[26] Rob Procter, Jeremy Crump, Susanne Karstedt, Alex Voss, and Marta Cantijoch. Reading the riots: What were the police doing on twitter? *Policing and society*, 23(4):413–436, 2013.
[27] Nick Malleson and Martin A Andresen. The impact of using social media data in crime rate calculations: Shifting hot spots and changing spatial patterns. *Cartography and Geographic Information Science*, 42(2):112–121, 2015.
[28] John R Hipp, Christopher Bates, Moshe Lichman, and Padhraic Smyth. Using social media to measure temporal ambient population: Does it help explain local crime rates? *Justice Quarterly*, 36(4):718–748, 2019.
[29] Xiaofeng Wang, Matthew S Gerber, and Donald E Brown. Automatic crime prediction using events extracted from twitter posts. *SBP*, 12:231–238, 2012.
[30] Somayyeh Aghababaei and Masoud Makrehchi. Mining social media content for crime prediction. In *2016 IEEE/WIC/ACM International Conference on Web Intelligence (WI)*, pages 526–531. IEEE, 2016.
[31] Mark Shaw and Simone Haysom. Organised crime in late apartheid and the transition to a new criminal order: the rise and fall of the johannesburg 'bouncer mafia'. *Journal of Southern African Studies*, 42(4):577–594, 2016.
[32] Steven Lawrence Gordon. Understanding xenophobic hate crime in south africa. *Journal of Public Affairs*, 20(3):1–4, 2020.
[33] Denise Buiten and Kammila Naidoo. Framing the problem of rape in south africa: Gender, race, class and state histories. *Current Sociology*, 64(4):535–550, 2016.
[34] Paula Meth. Informal housing, gender, crime and violence: The role of design in urban south africa. *The British Journal of Criminology*, 57(2):402–421, 2017.
[35] Gregory D Breetzke, Karina Landman, and Ellen G Cohn. Is it safer behind the gates? crime and gated communities in south africa. *Journal of housing and the built environment*, 29:123–139, 2014.
[36] A. Gloverr. Here are the top 10 most dangerous cities in the world, 2019. 7news.com. [Online]. Available: https://7news.com.au/news/travel/here-are-the-top-10-most-dangerous-cities-in-the-world/.
[37] P. Megaro. Top 10 cities with the highest crime rate in the world, 2019. appealslawgroup.com. [Online]. Available: https://www.appealslawgroup.com/top-10-cities-with-the-highest-crime-rate-in-the-world/.
[38] H. Mallinson. Most dangerous destinations in the world mapped – top 10 cities with the most crime, 2020. appealslawgroup.com. [Online]. Available: https://www.express.co.uk/travel/articles/1309270/holidays-most-dangerous-countries-in-the-world\\-numbeo-crime-rate-south-africa-brazil-map/.
[39] Anna Rypi, Veronika Burcar, and Malin Åkerström. Refraining from reporting crimes: accounts from young male crime victims with an immigrant background. *Nordic Social Work Research*, 9(2):131–146, 2019.
[40] Juha Kääriäinen and Reino Sirén. Trust in the police, generalized trust and reporting crime. *European journal of criminology*, 8(1):65–81, 2011.
[41] Alexander Meduna and Petr Zemek. Jumping finite automata. *International Journal of Foundations of Computer Science*, 23(07):1555–1578, 2012.
[42] VX Heavens. Metamorphism, formal grammar and undecidable code mutation. *International Journal of Computer Science*, 2(1):70–75, 2007.
[43] A John Kaspar, DK Christy, and DG Thomas. Lattice regular grammar-automata. *IAENG International Journal of Applied Mathematics*, 52(4):1123–1129, 2022.
[44] Vojtěch Vorel. On basic properties of jumping finite automata. *International Journal of Foundations of Computer Science*, 29(01):1–15, 2018.
[45] Simon Beier, Markus Holzer, and Martin Kutrib. Operational state complexity and decidability of jumping finite automata. *International Journal of Foundations of Computer Science*, 30(01):5–27, 2019.
[46] R. Sharma H. Singh and P. Tiwana. Conversion of jumping finite automata with null moves to deterministic jumping finite automata. *Journal of Critical Reviews*, 7(17):2526–2542, 2020.
[47] Matthew S Gerber. Predicting crime using twitter and kernel density estimation. *Decision Support Systems*, 61:115–125, 2014.
[48] Shiliang Wang, Michael J Paul, and Mark Dredze. Social media as a sensor of air quality and public response in china. *Journal of medical Internet research*, 17(3):1–5, 2015.
[49] Hongjian Wang, Daniel Kifer, Corina Graif, and Zhenhui Li. Crime rate inference with big data. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, pages 635–644, 2016.
[50] Varvara Ingilevich and Sergey Ivanov. Crime rate

prediction in the urban environment using social factors. *Procedia Computer Science*, 136:472–478, 2018.

[51] Hitesh Kumar Reddy ToppiReddy, Bhavna Saini, and Ginika Mahajan. Crime prediction & monitoring framework based on spatial analysis. *Procedia computer science*, 132:696–705, 2018.

[52] T. Head. The future is on its way: South african tech firm using ai to predict crime before it happens, May 2019. [Online]. Available: https://www.thesouthafrican.com/.

[53] Shiju Sathyadevan, MS Devan, and S Surya Gangadharan. Crime analysis and prediction using data mining. In *2014 First international conference on networks & soft computing (ICNSC2014)*, pages 406–412. IEEE, 2014.

[54] Tong Wang, Cynthia Rudin, Daniel Wagner, and Rich Sevieri. Learning to detect patterns of crime. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2013, Prague, Czech Republic, September 23-27, 2013, Proceedings, Part III 13*, pages 515–530. Springer, 2013.

[55] Gabriel Rosser, Toby Davies, Kate J Bowers, Shane D Johnson, and Tao Cheng. Predictive crime mapping: Arbitrary grids or street networks? *Journal of quantitative criminology*, 33:569–594, 2017.

[56] Ayisheshim Almaw and Kalyani Kadam. Crime data analysis and prediction using ensemble learning. In *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, pages 1918–1923. IEEE, 2018.

[57] Keith J Hayward and Matthijs M Maas. Artificial intelligence and crime: A primer for criminologists. *Crime, Media, Culture*, 17(2):209–233, 2021.

[58] Alif Ridzuan Khairuddin, Razana Alwee, and Habibollah Harun. Comparative study on artificial intelligence techniques in crime forecasting. In *Applied Mechanics and Materials*, volume 892, pages 94–100. Trans Tech Publ, 2019.

[59] Stephen Obare, Abejide Ade-Ibijola, and George Okeyo. Jumping finite automata for tweet comprehension. In *2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, pages 1–7. IEEE, 2019.

[60] Andrew Guthrie Ferguson. Policing predictive policing. *Wash. UL Rev.*, 94(5):1–5, 2016.

[61] Jessica Fitterer, Trisalyn A Nelson, and Farouk Nathoo. Predictive crime mapping. *Police Practice and Research*, 16(2):121–135, 2015.

[62] David Omand, Jamie Bartlett, and Carl Miller. Introducing social media intelligence (socmint). *Intelligence and National Security*, 27(6):801–823, 2012.