UNIVERSITY OF CALIFORNIA,
IRVINE

Crime Intelligence 2.0: Reinforcing Crowdsourcing using Artificial Intelligence and Mobile Computing

THESIS

submitted in partial satisfaction of the requirements
for the degree of

MASTER OF SCIENCE

in Information and Computer Science

by

Sanket Subhash Khanwalkar

Thesis Committee:
Professor Gloria Mark, Chair
Professor Ramesh Jain
Professor David Redmiles

2016

TABLE OF CONTENTS

LIST OF FIGURES

Page

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to thank everyone who has supported me throughout this project. This journey would not have been the same without you all.

More than a year back, Professor Ramesh and I shared our concern towards crime and terrorism, and got together to discuss over how technology could be used to counter these pressing issues. After countless number of meetings and emails since then, I express my deepest gratitude towards him for believing in my ideas and shaping them into reality. His vast experience, insightful vision and determination for social good will always remain matchless.

I am sincerely grateful to Professor Gloria, the chairperson of the thesis committee, who gave me the confidence to explore this research area without fearing the final outcome of the project. With the hurdles that we faced during the course of this project, her advice was instrumental in taking important decisions.

Despite joining the committee towards the later stages of the project, Professor Redmiles' contribution had an outstanding impact. I am really thankful to him for expressing his concerns then and pushing my limits to evaluate our research, and bringing all the pieces together.

Apart from my thesis committee, I would like to thank UC Irvine Police Department. Their understanding of crime intelligence and willingness to improve current infrastructure was significant.

I would also express love for my family who raised me with great values, and my determination to solve pressing problems in the society is certainly their giving.

On a concluding note, I certainly believe that this is just the beginning . . .

CURRICULUM VITAE

Sanket Subhash Khanwalkar

EDUCATION

Master of Science in Information and Computer Science **2016**
University of California *Irvine, CA*

Bachelor of Engineering in Information Technology **2007**
University of Mumbai *Mumbai, India*

TEACHING EXPERIENCE

Teaching Assistant **2014–2016**
University of California *Irvine, CA*

REFEREED CONFERENCE PUBLICATIONS

uBeacon: Configuration based Beacon tracking **Mar 2016**
2016 IEEE International Conference on Pervasive Computing and Communication

ABSTRACT OF THE THESIS

Crime Intelligence 2.0: Reinforcing Crowdsourcing using Artificial Intelligence and Mobile Computing

By

Sanket Subhash Khanwalkar

Master of Science in Information and Computer Science

University of California, Irvine, 2016

Professor Gloria Mark, Chair

Despite heavy investments by local and national governments, crime continues to remain a serious problem in the society. Current state-of-the-art is below average and relies heavily on crime intelligence and efforts of human personnel across several law enforcement agencies.

Over the last few years, crowdsourcing has been promoted worldwide through several Suspicious Activity Reporting (SAR) campaigns. However, such systems have proven to be largely unsuccessful owing to complicated reporting processes, excessive noise, limited authorized personnel, lack of motivation and one-way communication.

To address these issues and improve counterintelligence, this thesis proposes two approaches: (i) Combination of Machine Learning and Natural Language Processing to promote generic reporting while automating the process of crime detection, summarisation and delegation (ii) Improving crowd intelligence by enabling better citizen involvement through newer information sharing techniques, sophisticated sensors and responsive feedback. In this thesis, we also investigate lone wolf terrorism as a special case of crime, and evaluate application of our proposed methods for its mitigation.

Chapter 1

Introduction

1.1 Background on Crime

In 2014, according to U.S Bureau of Justice Statistics (BJS) National Crime Victimization Survey (NCVS), US residents (age>12) experienced approximately 5.4M violent victimizations and 15.3M property victimizations [4].

In 2015, UK police recorded around 4.3M offences (violent crimes, sexual offense, knife/firearm crime, etc), amounting to a 6% increase in crime reported as compared to the previous year. The increase in crime earlier this year, for the month of February and March, is clearly evident from Figure 1.1 [36].

Crime	Feb 2016	Mar 2016	Change
Anti-Social Behaviour	129,283	144,780	+15,497
Robbery	4,130	4,327	+197
Robbery	33,502	34,369	+867
Vehicle Crime	29,900	31,280	+1,380
Violent Crime	89,703	98,534	+8,831
Violent Crime	89,703	98,534	+8,831
Other Crime	157,381	168,460	+11,079
Theft - Shoplifting	29,170	30,109	+939
Drugs	11,316	11,884	+568
Criminal Damage and Arson	44,383	47,932	+3,549
Theft - Other	37,252	39,463	+2,211
Bike Theft	4,812	5,547	+735
Theft From the Person	6,504	6,773	+269
Possession of Weapons	2,204	2,381	+177
Public Order	15,934	18,290	+2,356
Total Crimes Plus ASB	**443,899**	**481,750**	**+37,851**

Figure 1.1: UK Crime Statistics 2016

Even in a developing country like India, where 30% of crimes are not reported and 50% of crime reports are turned away by authorities, the official crime rate is high. According to the report by National Crime Records Bureau in 2014, the overall crime rate has increased by 8.9% from the year before [28].

Even if we disregard the increase in the overall crime rate, the staggering numbers themselves depict the sorry state of law enforcement authorities in curbing crime. In most countries, there is a wide gap between crime victimization surveys and official crime statistics. In order to reduce this gap, in addition to improving law and order reforms, crime reporting mechanisms need special attention.

1.2 Motivation

Why Crowdsourcing for crime reporting?

Crime reports provided by crowdsourced initiatives like SAR clearly help the law enforcement agencies in identifying existing crimes that need action. However, most importantly, it also provides data that needs to be aggregated for long-term strategy in mitigating crime. Analysis of such reported data can give us higher level information about crime. This information, if communicated effectively, leads to knowledge creation and further insights. Such insights enable authorities to build strategies and utilize the knowledge in their decision making and future policies [44].

Given the disparity between actual and official crime statistics, crowdsourcing intelligence can play an influential role in shaping public perception about crime and safety in a given community. This may further lead to effective and realistic approaches towards policy making and policing strategies for safety perception management in the future [52].

Existing Issues

Failure of crowdsourcing in crime reporting can be attributed to several factors - inefficient technology, lack of awareness, organizational incompetence, outdated policies, etc. However, the focus of this thesis is on addressing issues whose impact can be mitigated using technology. In this context, crowdsourcing initiatives in crime have not been successful owing to the following factors:
(i) Limited personnel capable of analyzing and acting upon the high number of crime reports generated through popular campaigns (such as 'See Something, Say Something') [58].
(ii) Most of the traditional crime reporting in mobile and web applications continue to support clunky forms requesting users to furnish too many details in a fixed format, making the

entire reporting process cumbersome.

(iii) Opinions, apprehensions and racial complaints dominate suspicious reports and crime reports in general with actual numbers varying based on different geographical and economical factors [33].
(iv) Crowdsourcing in general faces a tricky problem of incentivising or motivating the users to contribute to expected information or tasks. To make matters worse, in crowdsourcing crime reporting, there's an additional risk of putting lives in danger.
(v) Most of the existing systems allow only one way communication - from citizens to law enforcement agencies. This leaves the citizens in a state of wonder over the 'real' status of their contributed report - reviewed, acted upon, closed or discarded.

1.3 Special Case: Lone Wolf Terrorism

In addition to addressing existing issues and building a powerful system to mitigate general crime, we also explore opportunities where we can build systems to counter lone wolf terrorism.

Most of the SAR campaigns are tailored around finding suspicious activity, behavior, patterns, objects, etc. which allow law enforcement agencies to use citizen's eyes to keep an additional watch on potential threats to national security, lone wolf terrorism being the most prominent one.

Lone Wolf Terrorism: By definition, A lone wolf or lone-wolf terrorist is someone who commits violent acts in support of some group, movement, or ideology, but who does so alone, outside of any command structure and without material assistance from any group [24].
According to Global Terrorism Index, since 2000, the number of deaths related to terrorism

have increased five-fold [16]. Intelligence services are continuously working towards identifying and eliminating terrorist groups that are organised in structure and substantial in size and activity. Until 2014, terrorist organisations like Al-Qaeda, ISIL, Boko Haram, etc. spread terrorism through planned heavy attacks or on-ground- war-like attacks. Owing to tremendous effort by the Intelligence, a majority of terror attacks have been foiled, forcing the terrorist groups to alter their strategies. These terror groups are fully aware that fighting any of these powerful governments on ground will only weaken them, hence, some groups like ISIL have now devised a more decentralised and independent operative strategy - Lone Wolf Terrorism. Dealing with unconventional and low-key terrorism tactics like lone-wolf, needs designing and implementing approaches that are unorthodox in nature. And thus, we investigate the fitness of crowdsourcing intelligence as an approach to countering lone-wolf terrorism.

1.4 Thesis Statement

This thesis emphasises on a literature study related to the factors that have influenced the history and evolution of crime intelligence, crime reporting and understanding lone wolf terrorism in depth. To support our literature, we will also be studying a variety of applications/initiatives relevant to crime intelligence and reporting, and also some of the major issues in such systems. Based on this, we discuss the design and implementation of our 2 approaches: (i) Artificial Intelligence and (ii) Mobile Computing and how they address these issues by improving quality of reports, automatic crime classification, effective utilisation of personnel, two-way communication, ease of reporting, etc. We will also ascertain the success of our approaches using statistical measures and pilot studies.

With the premise set for the research, this thesis begins with chapter 2 describing some of the related work, highlighting literature study and study of existing systems and initiatives

in the field of crime intelligence and reporting, and lone wolf terrorism. With the issues in current state-of-the-art clearly laid out, Chapter 3 introduces our two proposed approaches for mitigating previous issues and improving existing systems, followed by evaluation of both the approaches. Finally, Chapter 4 will conclude our research and discuss related future work.

Chapter 2

Related Work

This section investigates literary work in crime intelligence and lone wolf terrorism ideology in depth. After this, we discuss existing crowdsourcing sytems/intitiatives implemented to counter crime and lone wolf terrorism and also existing issues with such systems/initiatives.

2.1 Understanding Crime

Almost a century back, crime intelligence started with maintaining dossiers that contained critical information about suspects. Over the years, state of national and global affairs evolved drastically, but crime intelligence continued to be restricted to dossiers. In late 1960s, drastic changes in the civil rights against law enforcement agencies started to change the dossier intelligence landscape. Law enforcement agencies were compelled by the courts to purge these dossiers containing personal information of innocent citizens tagged as suspects. The cost of damage and embarrassment to law enforcement agencies suggested an important lesson: information must be collected, maintained, and disseminated in a manner that is consistent with legal and ethical standards [47]. With the reshaping of role of the FBI and

the CIA, each local law enforcement agency was required to setup an intelligence unit that executed regulations and coordinated with regional, state and all the way up to the federal units. However, most of these local units were counter productive owing to the following reasons: (i) Lacking analysis skills for the data collected, (ii) Not a proactive but instead a reactive approach towards taking action using crime intelligence, (iii) Slow evolution of intelligence data and products causing lack of effectiveness in law enforcement operations. In the last decade, the 9/11 Commission report addressed several suggestions for improving relations between intelligence units at different levels and re-engineering of the intelligence community.

Until few years back, the law enforcement structure needed a criminal or terrorist activity to occur, before it can react. Primary reason for this reactive nature of action is due to the legal system that is focused on punishing criminal after a crime has been committed, and not before. Two, a major part of law enforcement finds its difficult to monitor criminals, as they are occupied with extremely high incarceration practices in countries like the US. Three, in case of terrorism related activities, the occurrence is relatively low which causes such activities to be obscured by everyday events [53].

For research, development and execution of effective crime reporting strategies, understanding the determinants of citizen reporting is critical. Crime reporting activity can be classified into three main categories, viz. incident- or crime-specific, individual-specific and environment-specific. Contrary to the general assumption that crime reporting might be heavily weighted towards individual-specific correlates, study conducted in a developed nation like USA proves that incident-specific correlates are the biggest predictors of crime reporting [42].

However, another study performed over a wider spectrum of countries with varying income-per-capita, shows different results. Results of this study show that crime reporting has a strong correlation with police presence, institutional development, degree of urbanisation,

level of education and corruption [61].

With several reforms in the working of law enforcement agencies, perception towards crime reporting for helpfulness and a considerate action from the law enforcement has been improving. According to a study conducted in New York, from the late 80s to the last decade, the number of victims citing "police wouldn't help" has been decreasing tremendously. Some of the additional police- and nonpolice- related factors influencing the increase in reporting are: rise in community policing, social movements strengthening citizen rights, improvements in 911 and mobile systems [65].

In the last few years, reporting crime via text messages to 911 and local law enforcement agencies have also increased. These text tip lines transfers the received crime reports to central information center and then delegates it to an appropriate local department for effective action. Such an initiative clearly showcases efforts of authorities to stay updated with current trend of frequent emailing and messaging as against the older trend of calling.

In some cases of minor crimes, crime reporting may be a mere formality for claiming damages from insurance companies. In such cases, online reporting generate a crime report number thereby saving time and resources for both residents, and officers who can focus on other more important and in-progress crimes.

According to Police Executive Research Forum (PERF) Survey of around 200 law enforcement agencies, 82% believe that online crime reporting will become grow in the next 2-5 years [15].

> **"Do we want to police in the future the way we are policing today? I dont think so. How did we get to where we are today? We looked; we listened; we learned; we changed; we grew with the times. Thats our job as leaders today."**

- Jim Fox, Chief, Newport News, Virginia, Police Department

2.2 Understanding Lone Wolf Terrorism

According to Ramon Spaaij, lone-wolf is the terrorism that is carried out by persons who,

- operate individually,
- do not belong to an organized terrorist group or network, and
- whose modi operandi are conceived and directed by the individual without any direct outside command or hierarchy [62].

In his study, *The Enigma of Lone Wolf Terrorism*, Ramon points out that lone-wolf terrorism should not be confused with violent attacks by individuals that are motivated by financial gain or personal vengeance. Lone-wolf terrorism boasts of larger pursuits that are strongly influenced by politics, ideology or religion.

The structure of lone-wolf terrorism has an interesting design. In his famous essay *Leaderless Resistance*, Louis Beam envisioned an approach that can be highly correlated with the current lone-wolf terrorism strategy [41]. This approach consists of two tiers - one tier for organs of information and other tier for phantom cells. The first tier is responsible for spreading the propaganda and recruiting potential lone-wolves in the process. The second tier allows these influenced individuals to carry out real attacks. Even though this approach seems interwoven, theres often a disconnect between the two tiers which allows the terrorist organisations to maintain their flexibility and scale both on political and ideological levels.

In the last decade, this 2-tier approach has become more lethal owing to the emergence of the internet, smartphones and social media. In *Leaderless Jihad*, Marc Sagemen mentions that

future generation of terrorists will consist of home-grown wannabes, who will be disconnected to the leadership, but will be globally connected via the internet [60]. He also emphasises that it is this structure (or the lack of structure) that will make them more destructive and volatile than their previous counterparts.

This disconnect gives lone-wolf terrorists advantages at various levels. First, since these are individuals that act independently, their identification and detection at any given stage of the attack, is extremely difficult. Second, it is tricky to differentiate between real threats and hoaxes from extremists. Finally, young lone-wolves try to imitate their idols which causes confusion for the Intelligence in identifying patterns [citation needed].

In his 1518 page-long manifesto, *A European Declaration Of Independence*, Anders Breivik, described topics like Islam colonization of Western Europe, current state of Jihad and how Jihad movements should move forward with solutions and strategies on multiple fronts of politics [43]. This book also acts as a guide presenting several pointers and techniques for individuals who practice lone-wolf terrorism. The following checklist is just a summary from his manifesto:

- Mobile phones must be turned off to avoid location triangulation.
- Treat email/surfing assuming that you are being monitored. Use secure email client and internet in public places like McDonalds.
- Increase credibility of your travels by carrying brochures or signing up for congresses or exhibitions and also liking similar pages on Facebook. Avoid using credit cards and always use cash during your travel.
- As a general rule. You will increase your chance of being apprehended by 100% for every person you involve. Dont trust anyone unless you absolutely need to (which should never be the case). Do absolutely everything yourself.

On one hand where this disconnect strengthens the execution and survival of lone-wolf terrorism, on the other hand, it also exposes weaknesses that are yet to be fully exploited by the counterterrorism agencies. In *Capitalising on the disconnect*, Fred and Scott emphasize on certain mistakes that to-be-lone-wolves are likely to commit that can assist in Counter Terrorist (CT) activities [45]. First, they tend to commit errors by working out illogical or suspicious deals for resources like firearms, bombs, chemicals, etc. required to carry out terror attacks. Second, they also tend to look out for like minded extremists with whom they can share their early apprehensions and views to backup their clarity and fuel their motivation further. Third, since most of them get trained online or in an environment lacking direct trainers, these individuals tend to simulate their attacks in the real world before actually executing their final attacks. All these areas clearly show weaknesses in the disconnect and can be heavily exploited by CT to foil such attacks in future.

2.3 Existing crowdsourcing systems/initiatives

In addition to studying the literary work, it is also important to understand different scenarios in which crime reporting/intelligence and crowdsourcing have worked in a synergistic manner in the past.

***London Riots*, 2011**

In order to crack down on the crime offenders of the infamous London riots in 2011, London Metropolitan Police (LMP) turned towards crowdsourcing for a solution. More than 2800 CCTV photos were uploaded by the police on Flickr and Facewatch ID app. Users were then requested to sort these photos by postal codes and report identified users to the authorities. Later, Greater Manchester Police (GMP) launched another crowdsourcing campaign to find

information about the looters. Eventually, more than 770 people were arrested and around 167 of those were charged with several offences [11].

Texas Virtual Border Watch

Texas Virtual Border Watch, an initiative implemented by the US defence around its Mexican border, crowdsources labor for surveillance. In this program, citizens can keep a watch over the US-Mexico border through the live footage available to them via surveillance cameras mounted around the border. In the journal *Surveillance and Society*, Doug Tewksbury, argues that these techniques are not effective and cost-efficient currently. However, with the rise of web 2.0 as a participatory platform, they do offer a new approach to governance and surveillance under Neoliberalism, shifting the power of surveillance from the state to the citizens [63].

Internet Reporting, France

The French government has setup a web portal that allows citizens to report any crime witnessed over the internet. It includes crimes like spam, fraud, incitement to violence, illicit trafficking, etc. witnessed over websites, blogs, forums, social networks, etc. The users can choose to remain anonymous and/or sign-up for follow up via arbitrary accounts created for the users [14].

SpotCrime

Crowdsources crime data from police departments, news reports and users. It then plots the crime on Google maps and sends out alerts on social media and other messaging platforms [32]. The crime data is used for categorizing state wise, the daily crime reports and most

wanted criminals in different states.

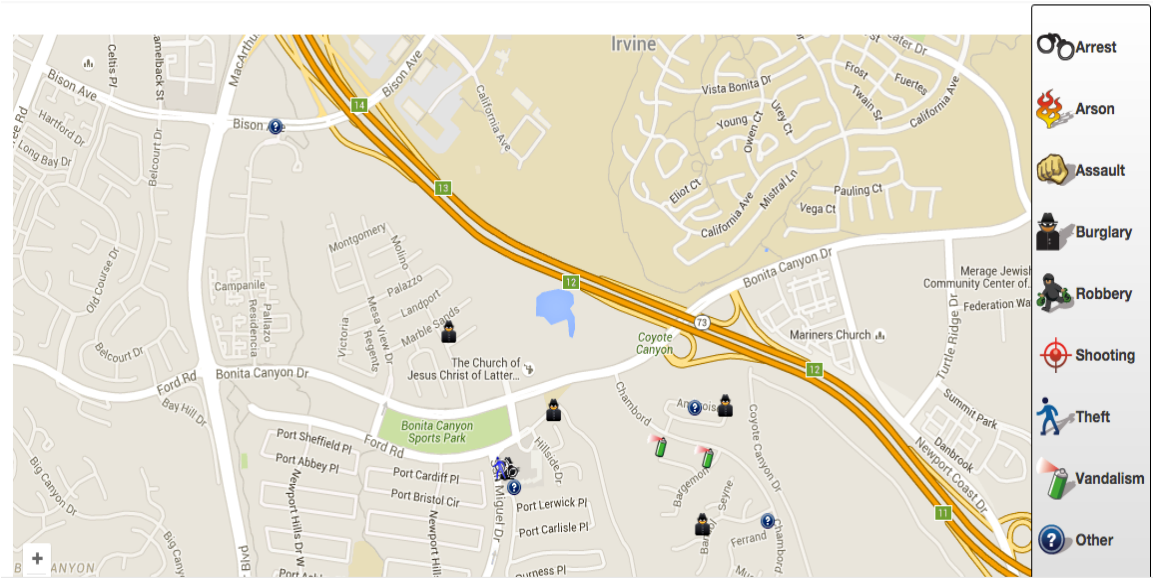


Figure 2.1: Crime reporting on SpotCrime

Ushahidi

During the presidential elections of 2007 in Kenya, Ushahidi created a platform to enable citizens of Kenya report incidents of violence through text messages and emails. According to a Harvard study, at that time, data collected via Ushahidi was far more helpful than that reported by mainstream media since it crowdsourced data from rural areas too. Post this successful campaign, Ushahidi has used crowdsourcing in several projects related to crime, corruption, natural disasters, etc [39].

Microblogs/Social Media Networks

Recently, microblogs and social media networks have been extensively used by law enforcement agencies to share crime related warnings/updates, look out for clues in profiles of suspects, tracking more updated information on public gatherings, events, etc [49]. But most importantly, these platforms not only empower citizens to report crime incidents but

they also enable them to proactively monitor and highlight problems thriving in the society, at times even allowing them to 'publicly shame' other citizens to ensure compliance [56, 55, 54].

See Something, Say Something

National campaign launched by US Department of Homeland Security (DHS) to increase public awareness on the indicators of terrorism and terrorism-related crime through reporting of suspicious activity to state and local law enforcement. Suspicious activity includes spotting unattended bags or vehicles in public places, chemical substances, behavior of trespassing restricted areas, etc. National Suspicious Activity Reporting Initiative (NSI) standardizes the protocols related to documenting, analyzing and sharing information with the FBI led Joint Terrorism Task Forces for investigation and with state Fusion Centers for analysis [30]

A large number of mobile applications are implemented by local and state authorities where users can send anonymous tips about suspicious activity related to terrorism. Some popular mobile applications around this campaign are LAPD's *iWatchLA*, *See Send*, MARTA's *See Say*, etc [23, 29, 30].

Wikistrat

Wikistrat, apparently, the worlds first crowdsourced consultancy, tackles complex strategic challenges by offering solutions collaborated over an online network of more than 2000 subject matter experts. In the past, Wikistrat was hired to predict how ISIS would infiltrate Jordan. After several rounds of brainstorming and critiquing, the Wikistrat team presented a set of mutually agreed upon possibilities. The results ranged from capture of Amman, creating chaos over Jordans peace treaty and recruiting vulnerable-to-radicalization citizens [7].

Similarly, Wikistart has also crowdsourced an in-depth analysis and strategy for illicit trafficking activities in Trans-Sahel region, US messaging tactics to counter ISIS popularity, impact of Ansar Bayt Al-Maqdiss oath of allegiance to the ISIS, etc. According to Tim Haffner, a US Africa Command Analyst, crowdsourcing offers multiple perspectives for processing and synthesizing complex data much faster [40].

FBI Tips and Public Leads

FBI also uses crowdsourcing at their web portal for collecting any information related to suspected terrorism or criminal activity by allowing users to enter information using a web form and optionally sharing their personal information. This information is then reviewed by an FBI special agent or professional staff member and acted upon accordingly [13]. Similar web portal *Anti Terrorist Hotline* has been deployed by UK Metropolitan Police to allow its citizens to report terrorist-related information [37].

2.4 Additional Issues

Even with all the benefits around the establishment of a crowdsourced platform for crime reporting, it certainly opens up several fronts vulnerable to attacks.

The False Positives

While crowdsourcing intelligence during investigations have helped in cracking down certain cases, some other incidents like the Boston Marathon Bombings (April 2013) reveal a completely different picture. The investigation to find the bombers took a different course, when countless number of photos surfaced on Reddit uploaded via Flickr and Google drive. An

army of Reddit users started hunting for clues and identifying suspicious objects and people in the photographs. However, this eventually led to witch hunting and negative consequences for innocent individuals, further consolidating the general observation that opinion of crowd is resistant to change. Despite several warnings and rules setup on Reddit subthreads, some suspects were falsely accused and harassed on social media [7, 11]. This clearly signifies the lack of operational readiness of platforms like Reddit to investigate such cases independently.

In the past, crowdsourced campaigns like Hoaxmap [17] in Germany have been initiated to debunk false information. In Hoaxmap, allegations and rumors over crimes committed by the immigrants were corrected by citizens by pointing to more accurate and official sources of underlying information.

False positives based out of opinions, assumptions, sentiments and apprehensions eventually become causes for conflicts or damages on several social, political and ideological fronts [48].

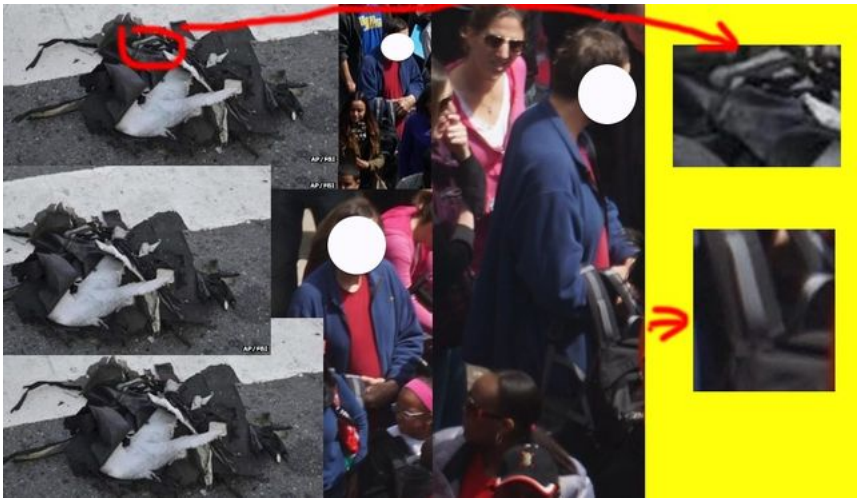


Figure 2.2: Boston Bombings investigation on Reddit

Identity Management

When crime reporting is crowdsourced, participants share personal information with tremendous faith that privacy, trust, legality and ethics will not be breached under any circum-

stances. In the developed countries, with a solid technical and infrastructure in place, there's a higher likelihood of participants feeling more secured while using such systems. However, in other more under resourced and unstable economies, the lifeblood for solving crime, may itself spill the lifeblood of those who share such critical information [46]

In his 2012 article in *Culturally Digital*, Andres delineates that, when it comes to crime reporting, information itself reveals identity [8]. Information leaks during Drug War in Mexico precisely pointed to sources being local to the community. Even on a smaller scale, say domestic violence, a victim reporting such a crime, unknown to anyone outside, directly brings the victim in the spotlight. In such cases, irrespective of the type of identity management deployed, algorithms implemented or information channel used, the information points to the source unambiguously.

Motivation

In general, non-paid and non-expert crowdsourced data lacks credibility and is prone to errors [11]. One of the trickiest enigma in crowdsourcing crime reporting is motivating citizens to go that extra mile for their information contribution. Since, recognition in the society is considered to be the most favorable, social media networks like Facebook, Twitter, etc. may act as potential platforms for offering social incentives. However, in future, in order to motivate people globally, the nature of rewards must be adjusted to suit the nature of crime and demographics.

Selection Bias

Social media being majorly accessible to middle and upper class creates a selection bias. This bias indirectly eliminates a substantial segment in the society and data crowdsourced with this bias may not be potentially correct [11].

Crimesourcing

In a recent article published to *The Huffington Post*, Dr. Maha Hosain Aziz, argues that if terrorist organizations like ISIS can deploy crowdsourcing, why is US and the West, not dedicating enough resources to this unconventional approach [10]. ISIS has previously used crowdsourcing to strategize new policies, to brainstorm ideas to kill Moaz al-Kasasba - a Jordanian coalition fighter pilot, to collect Saudi Arabian Intelligence officials names, to name a few. Significant research and resources must be dedicated to consider crowdsourcing as an alternative approach to counter global terrorism. In the past, crimesourcing has been implemented extensively in setting up fake accounts, recruiting middlemen, purchasing resources, transferring money across nations, enticing vulnerable participants for committing certain activities, etc. [51].

Non-standardization of Systems

Larger federal agencies like DHS, FBI, LAPD, NYPD, etc. tasked with terrorism related responsibilities and hence have deployed their own SAR programs. However, despite strong recommendations in the post 9/11 homeland security strategy documents, there appears to be a gap in the implementation of SAR initiative/system in other local law enforcement agencies, specially the ones serving lower number of population [59]. In addition to this, despite recommendations in the Congressional Research Service report of 2009, current processes of information gathering, analysis and dissemination of law enforcement and homeland security related intelligence, lack co-ordination [57]. In addition to NSI, even the proposed use of social media by law enforcement has been managed in an 'ad hoc and relatively unorganized' manner, major reasons attributed to lack of awareness and varying prohibitory regulations on the use of such systems [64].

Chapter 3

Proposed Work

3.1 Overview

3.2 Introduction to Emerging Technologies

Crowdsourcing

Crowdsourcing is the practice of obtaining needed services, ideas, or content by soliciting contributions from a large group of people and especially from the online community rather than from traditional employees or suppliers [9]. With sporadic increase in smartphones and internet, crowdsourcing has gained tremendous traction for a myriad of applications. Most popular applications deploying crowdsourcing are as follows:

- *Wikipedia* - a free encyclopedia, by far the most comprehensive knowledge repository available today, enables general users to generate and contribute information on any topic across several languages.

- *Stackoverflow* - a Q&A platform for collaboratively answering questions across a wide range of topics related to computer programming.

- *Waze* - Like Google maps, Waze offers GPS-based navigation services. But additionally, it crowdsources real-time traffic information and also obstacles like road blocks, police patrols, etc thereby creating a more socially appealing navigation experience for the drivers.

- *Uber* - a worldwide app based cab booking service, an alternative to the traditional taxi service, uses crowdsourced drivers to offer rides to requested users.

- *Mechanical Turk* - an emerging online marketplace for work, where users can work on Human Intelligence Tasks (HITs) created by other users and get paid for their service accordingly.

Crowdsourcing systems currently face challenges in the four key areas: (i) Recruiting contributors, (ii) their role and functions, (iii) aggregating multiple contributions, (iv) managing user abuse [50].

Crowdsourcing intelligence allows citizens to identify and share information that contributes to national security and surveillance. This information then needs to be collected and analysed by Intelligence services for determining consequent actions. However, crowdsourcing intelligence requires the development of a system that demands an in-depth understanding and analysis of several social and political factors.

Machine Learning

Machine learning is a type of artificial intelligence (AI) that provides computers with the ability to learn without being explicitly programmed. Machine learning focuses on the

development of computer programs that can teach themselves to grow and change when exposed to new data [25].

- *Recommendation Systems* - Companies like Google, Netflix, Amazon heavily rely on machine learning to offer content recommendations to their users.
- *Object Recognition/Tracking* - Companies like Facebook have built sophisticated algorithms to detect faces and identify people from images.
- *Medical Diagnosis* - Health care research centers and organisations use machine learning to detect patterns in data and images for identifying problems and discovering solutions.

Natural Language Processing

Natural language processing (NLP) is a branch of artificial intelligence that deals with analyzing, understanding and generating the languages that humans use naturally in order to interface with computers in both written and spoken contexts using natural human languages instead of computer languages [27].

- *Intelligent Assistant* - Bots like *Siri* and *Cortana* rely on highly developed NLP algorithms for understanding user queries and fetching appropriate answers accordingly.
- *Fighting Spam* - Most of the top email service providers use NLP to detect certain keywords and their relationship with other words in the content for finding anomalies and classifying them as spam.
- *Information Extraction/Summarization* - With exponential rise in content creation, extracting critical information from the web in real time for algorithmic trading or summarizing massive content for social media requires advanced NLP algorithms.

3.3 Approach 1: Artificial Intelligence

3.3.1 Data

Traditionally, most of the crime reporting was either done in person or over telephonic hotlines like 911. In the last decade, proliferation of personal computers and smartphones along with reliable internet, has led to the development of various crime reporting web portals and mobile apps.

Irrespective of the platform or methodology, crime reporting data gets collected at local, state and federal levels. However, making this data available to the public can have the following issues:

Consistency: There is not consistency in the data variables for most of the organisations at different levels. This creates an issue in combining and consolidating data from different sources.

Maintainence: Maintaining this data digitally needs dedicated resources with state-of-the-art tools. Some organisations lack the funding and resources for making this data up-to-date or available as per requirements.

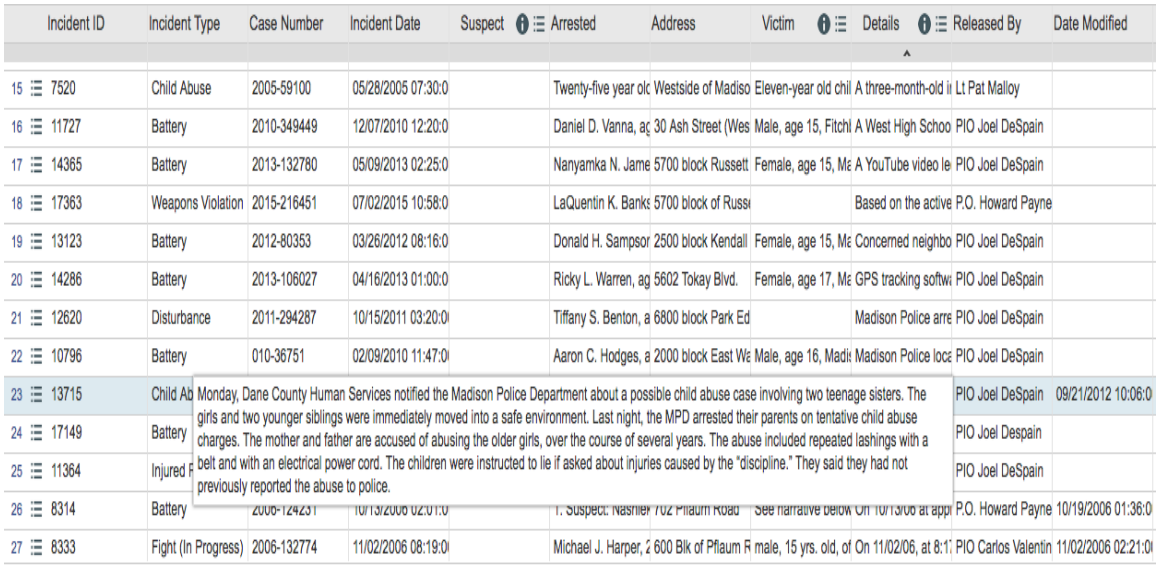
Cleaning: If this data is not cleaned for sensitive information, sometimes, information can easily point to the source itself [8].

Due to these reasons, requesting a crime reporting dataset, particularly SAR data, from any of law enforcement agency is difficult.

Simulation of real crime reporting process and building automated tools capable of taking responsive action, heavily relies on crime reporting data. Even in our system, we use machine learning for building our automatic crime classifier and this is fully dependent on the data used for its training. Even for identifying key entities and evaluating this system, data

plays a key role to gauge the effectiveness of the proposed system in real world. Even more specifically, we need crime reports that have sufficient description of the crime for them to be of any use.

After dozens of dataset requests getting rejected eventually and extensive research for open-sourced crime report datasets that have descriptive details, we finalised on City of Madison crime reporting dataset [5] shown in the figure 3.1. In addition to crime report attributes like incident type, incident date, address time, etc., this dataset described the details of the crime that happened. These details allowed us to analyse how a victim or a tipster reports or narrates a crime incident to the police.



	Incident ID	Incident Type	Case Number	Incident Date	Suspect	Arrested	Address	Victim	Details	Released By	Date Modified
15	7520	Child Abuse	2005-59100	05/28/2005 07:30:0		Twenty-five year olc	Westside of Madiso	Eleven-year old chil	A three-month-old i	Lt Pat Malloy	
16	11727	Battery	2010-349449	12/07/2010 12:20:0		Daniel D. Vanna, ag	30 Ash Street (Wes	Male, age 15, Fitchl	A West High Schoo	PIO Joel DeSpain	
17	14365	Battery	2013-132780	05/09/2013 02:25:0		Nanyamka N. Jame	5700 block Russett	Female, age 15, Ma	A YouTube video le	PIO Joel DeSpain	
18	17363	Weapons Violation	2015-216451	07/02/2015 10:58:0		LaQuentin K. Banks	5700 block of Russ		Based on the active	P.O. Howard Payne	
19	13123	Battery	2012-80353	03/26/2012 08:16:0		Donald H. Sampsor	2500 block Kendall	Female, age 15, Ma	Concerned neighbo	PIO Joel DeSpain	
20	14286	Battery	2013-106027	04/16/2013 01:00:0		Ricky L. Warren, ag	5602 Tokay Blvd.	Female, age 17, Ma	GPS tracking softw	PIO Joel DeSpain	
21	12620	Disturbance	2011-294287	10/15/2011 03:20:0		Tiffany S. Benton, a	6800 block Park Ed		Madison Police arre	PIO Joel DeSpain	
22	10796	Battery	010-36751	02/09/2010 11:47:0		Aaron C. Hodges, a	2000 block East We	Male, age 16, Madi	Madison Police loca	PIO Joel DeSpain	
23	13715	Child Ab								PIO Joel DeSpain	09/21/2012 10:06:0
24	17149	Battery								PIO Joel Despain	
25	11364	Injured F								PIO Joel DeSpain	
26	8314	Battery	2006-124231	10/13/2006 02:01:0		1. Suspect: Nashier	702 Pflaum Road	See narrative below	On 10/13/06 at appr	P.O. Howard Payne	10/19/2006 01:36:0
27	8333	Fight (In Progress)	2006-132774	11/02/2006 08:19:0		Michael J. Harper, 2	600 Blk of Pflaum R	male, 15 yrs. old, of	On 11/02/06, at 8:1	PIO Carlos Valentin	11/02/2006 02:21:0

Monday, Dane County Human Services notified the Madison Police Department about a possible child abuse case involving two teenage sisters. The girls and two younger siblings were immediately moved into a safe environment. Last night, the MPD arrested their parents on tentative child abuse charges. The mother and father are accused of abusing the older girls, over the course of several years. The abuse included repeated lashings with a belt and with an electrical power cord. The children were instructed to lie if asked about injuries caused by the "discipline." They said they had not previously reported the abuse to police.

Figure 3.1: Crime Report Dataset Snapshot

3.3.2 System Architecture

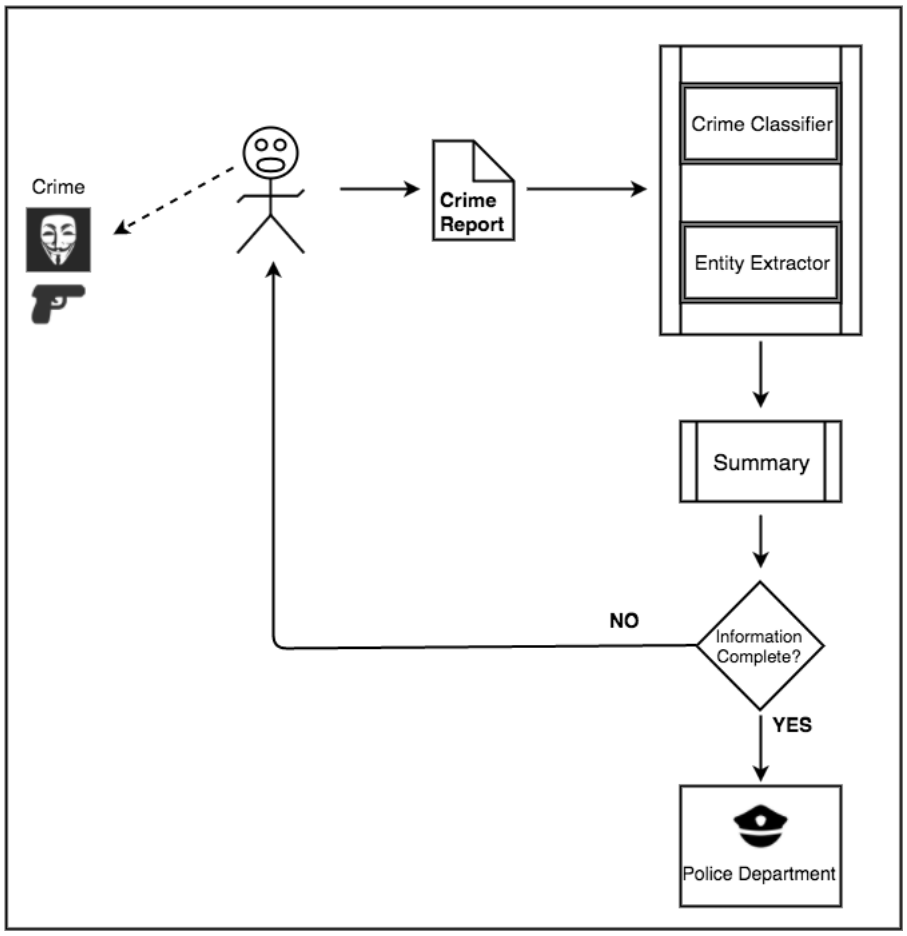


Figure 3.2: Crime Reporting using Artificial Intelligence

Crime Classifier

During crime reporting, classification of the report into a specific crime is critical for the following reasons:

- Allowing the system to prompt for missing entities in the report based on the crime type. For example, in case of a crime report that involves a suspected vehicle, the system can immediately prompt the victim or tipster to provide description of the

vehicle (type of vehicle, color, license plate, etc.).

- Passing the crime report to departments capable of taking faster and effective action. For example, a vandalism and a suspected weapon violation report need to be delegated to their corresponding departments with different priorities.

- Organisation and aggregation of crime data for visualisation and analytics. For example, before dispatching a team, law enforcement can also analyse their data and consider possibility of two crime events being connected based on their location or suspect involved, thereby improving their response.

The crime classifier in our system uses machine learning for identifying the crime via the text-based crime details available in the above mentioned dataset. This classifier has been trained on around 1435 samples using two algorithms, viz. Multinomial Naive Bayes and Support Vector Machines. The distribution of these samples into different categories and subcategories is given below:

Abuse(Total)	:	294
Domestic/Physical Abuse	:	30
Drug Abuse	:	119
Sexual Abuse	:	145
Suspicion(Total)	:	711
Suspicious Activity	:	269
Suspicious Person	:	147
Weapon Possession/Violation	:	295
Theft(Total)	:	430

For the ease of development, *MonkeyLearn* was used for creating web services and exposing end points for our web application to communicate with our crime classifier [26]. Web application, developed using Node.js, received crime report from the user, passed it as a

request to our crime classifier, which in turn returned the crime category results back to the web application.

Entity Extractor

This component is responsible for extracting important concepts present in the given piece of textual data. In our case, this component uses natural language processing for identifying key entities present in the crime report. Building an efficient entity extractor heavily relies on the data that it has been trained on. Thus, choosing the right entity extractor for our use case was a challenge.

A comprehensive research and testing across multiple NLP based concept extractors like *Thomas Reuters' Open Calais* [34], *Bitext* [3], *AlchemyAPI* [1], etc. was conducted. Finally, *IBM Watson Relationship Extraction* API [21] was chosen for the following reasons:

- This beta API, a series of IBM Watson, identifies entities from text and then extracts relationship between these entities
- Allows faster deployment on IBM Bluemix [19], thus allowing access to IBM Watson using RESTful web services.
- Developed using Node.js and Express.js, hence offers compatibility and easier integration with crime classifier discussed before.
- Additionally, being deployed on IBM Bluemix enabled addition of *Dialog* [20] component to existing web application

Note: IBM Bluemix is licensed, and required a prior procurement of the proprietary license for getting access to the above mentioned services.

3.3.3 Missing Information Prompting

Intially, we tried connecting our Crime Classifier and Entity Extractor with a component that could analyse the entities extracted, and based on the crime, it would prompt the user if there is any important entity missing in the report. But, due to technical issues, we could not integrate it within the given time frame. However, in order to create a prototype and make it available for evaluation, we created a simple bot and hosted it on *Facebook* messenger [18] using *api.ai* [38] and used the following strategy to create:
(i) A bot agent specific to our crime application. (ii) 'Entities' that represent concepts like crime-type or suspect-description. This allowed us to map natural language phrases to canonical phrases to understand the meaning. (iii) 'Intents' that allows mapping of user needs to what action our agent should take. For example, when the user says 'I want to report crime', the system should respond asking for certain details.

Note: Refer Appendix D for some examples of entities and intents for our crime application.

3.3.4 Demo Walkthrough

Demo 1

As shown in Figure 3.3, the proof-of-concept of such an intelligent system is realised using a basic web application deployed on local servers.
Note: Since we have used City of Madison Police Crime Data, this web application will be not be deployed on commercial web servers for public use, until permission has been granted.

Working of this prototype is explained step-wise as follows:

1. A sample crime report is entered in the text box provided in the web application.

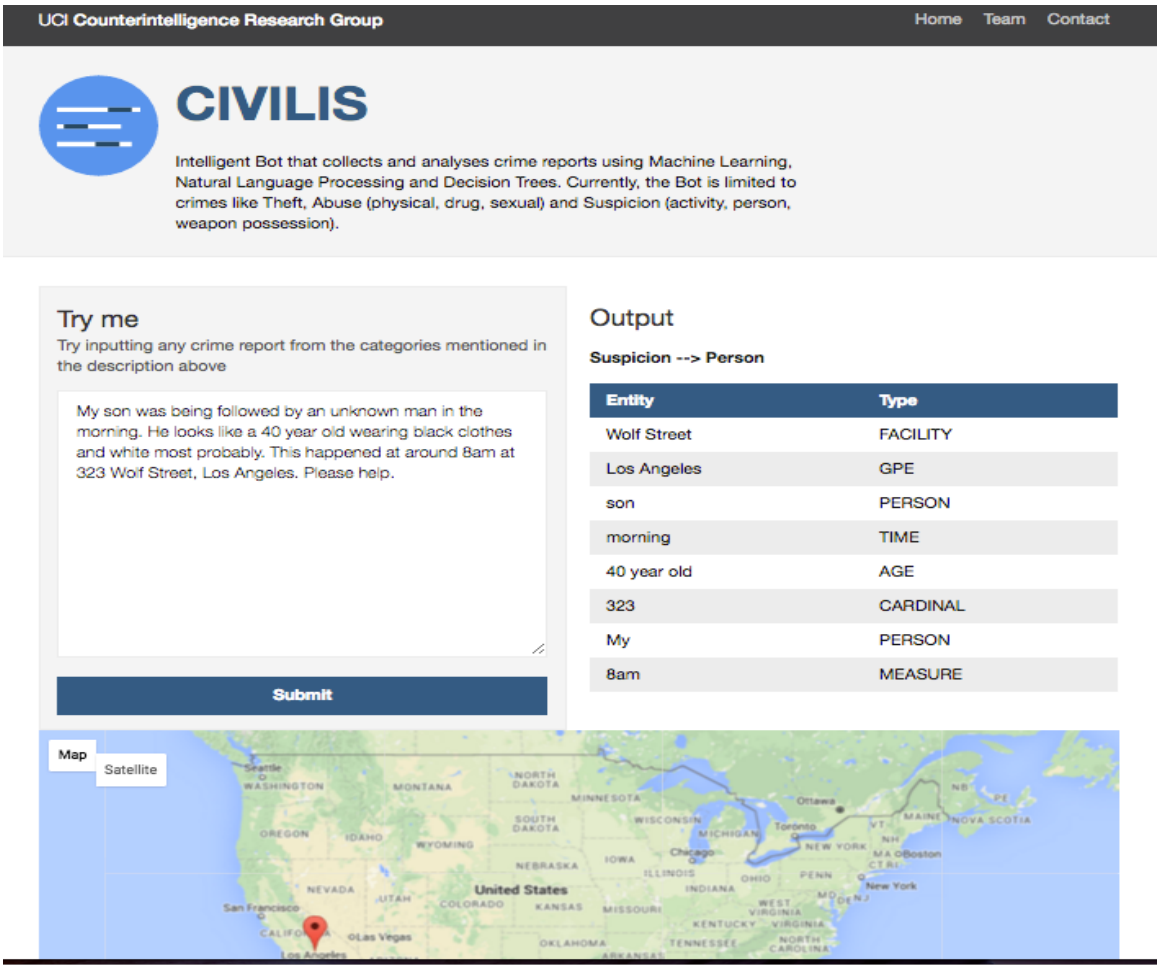


Figure 3.3: Web Application Prototype

2. Once the user clicks the submit button, the crime report is first sent to our crime classifier to fetch the crime category. In this example, 'Suspicious Person' is returned as the crime category.

3. In parallel, the crime report is also fed to our entity extractor and key entities are displayed in a tabular fashion to the user. In this example, *Los Angeles* is identified as the GPE (Geo Political Entity), 'morning' is the time, '40 year old' is age, so on and so forth.

4. If there is a location provided, it will be shown on map, for user to further fine tune it if needed. Knowledge of precise crime location is critical to offer appropriate action from the police department.

5. Most importantly, if the entity extractor recognises a missing entity, the user is immediately prompted to provide the missing details.

Demo 2 (Only for information prompt showcase)

1. A user/tipster wants to report a crime. He/She simply types our agent name 'Civilis' in Facebook Messenger. And mentions that he/she wants to report a crime.

2. The bot responds to this request and asks user to freely narrate the entire crime report.

3. User narrates the crime, but misses out on critical information like date/time and location. So, the bot immediately prompts the user to provide these missing details.

4. Using built-in machine learning and natural language processing, the bot classifies the crime report, into 'physical abuse' in our case.

5. Now with the basic crime report complete, User is then prompted for additional information like suspect description which may be critical for the law enforcement for taking appropriate action in finding the suspect.

6. Once the entire crime report is complete with all the details, the user is asked for a confirmation, and the report is submitted. In the ideal scenario, this report must reach the dashboard of a law enforcement agency which will then issue a dispatch accordingly.

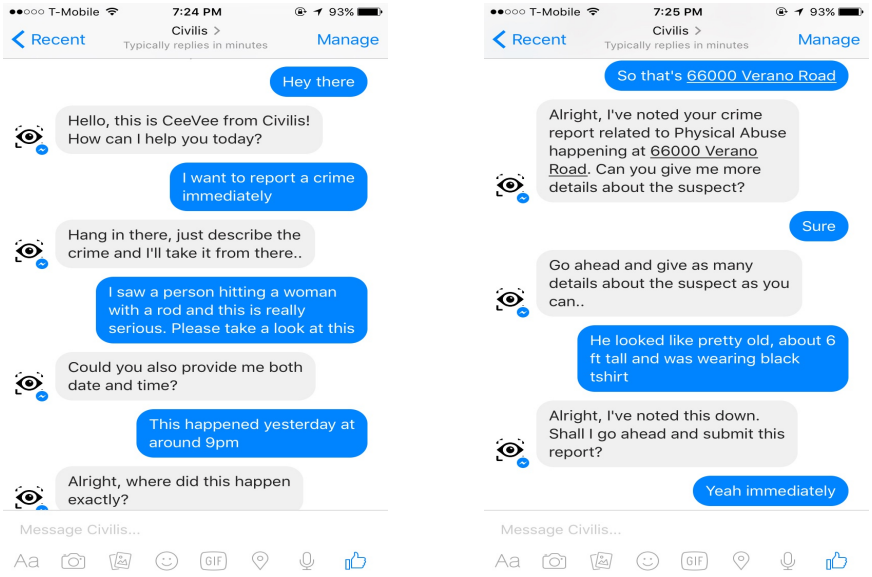


Figure 3.4: Crime Reporting using Bots

3.3.5 Evaluation

This system being fairly complex, can be evaluated in multiple ways. One, each of the components of the system can be evaluated separately. Two, analysing overall effectiveness of this system using real-world evaluation model. Both these methods have been detailed in the following subsections.

Crime Classifier Evaluation

This component, based out of machine learning, can be evaluated by simply using traditional success measures of accuracy, precision and recall. Also, *MonkeyLearn* directly offers calculation of these measures based on the following rubric:

(i) Split the training dataset randomly in four equal-length sets of samples.

(ii) For each set, train a model with the remaining samples (a 75% of the total samples) and use the 25% of the samples in the set for testing and getting the effectiveness of the model.

(iii) Repeat this process with the other 3 sets of samples.

(iv) The performance measures displayed on a classifier (accuracy, precision and recall) is the average obtained in the evaluation process of these 4 tests.

Note: Each sample was categorized by a classifier trained with other samples to avoid overfitting.

The overall accuracy of the classifier to correctly classify the given text-based crime report into a specific crime type was **90%**, a fairly reasonable number for us to move to the next stages of our prototype development.

For each of the crime categories, the table 3.1 lists the corresponding values of precision and recall.

Crime Category	**Precision**	**Recall**
Domestic/Physical Abuse	91%	63%
Drug Abuse	97%	97%
Sexual Abuse	93%	99%
Suspicious Activity	77%	92%
Suspicious Person	97%	78%
Weapon Possession/Violation	88%	81%
Theft	88%	95%

Table 3.1: Crime Classifier Evaluation Results

Entity Extractor and Information Prompt Evaluation

Unlike the evaluation of the crime classifier detailed above, the other components could not be evaluated statistically. However, evaluation of the importance and success of these components were included as a part of the survey that will be discussed in the *Mobile Computing* approach evaluation.

3.4 Approach 2: Mobile Computing

As previously discussed in the literature study, in the last few years, advancements in mobile technology have revolutionized the way information is being shared today. Interestingly, mode of information exchange has shifted from text-only to images. Modern smartphones now come with sleek built-in camera and sophisticated motion and environmental sensors that can calculate current orientation, location, temperature, humidity, ambient light, fingerprint, etc [31].

The current transition towards image-based information-sharing can be estimated based on the fact that in one minute on an average - *Instagram* users liked more than 1.7 million photos, *Pinterest* pinners pinned nearly 10,000 images and *Facebook* users uploaded nearly 136,000 photos [2, 12].

Thus, in this approach, we capitalize on crime reporting with image of the crime as the pivot, and other smartphone-enabled data like date, time, location, audio and text notes.

3.4.1 System Architecture

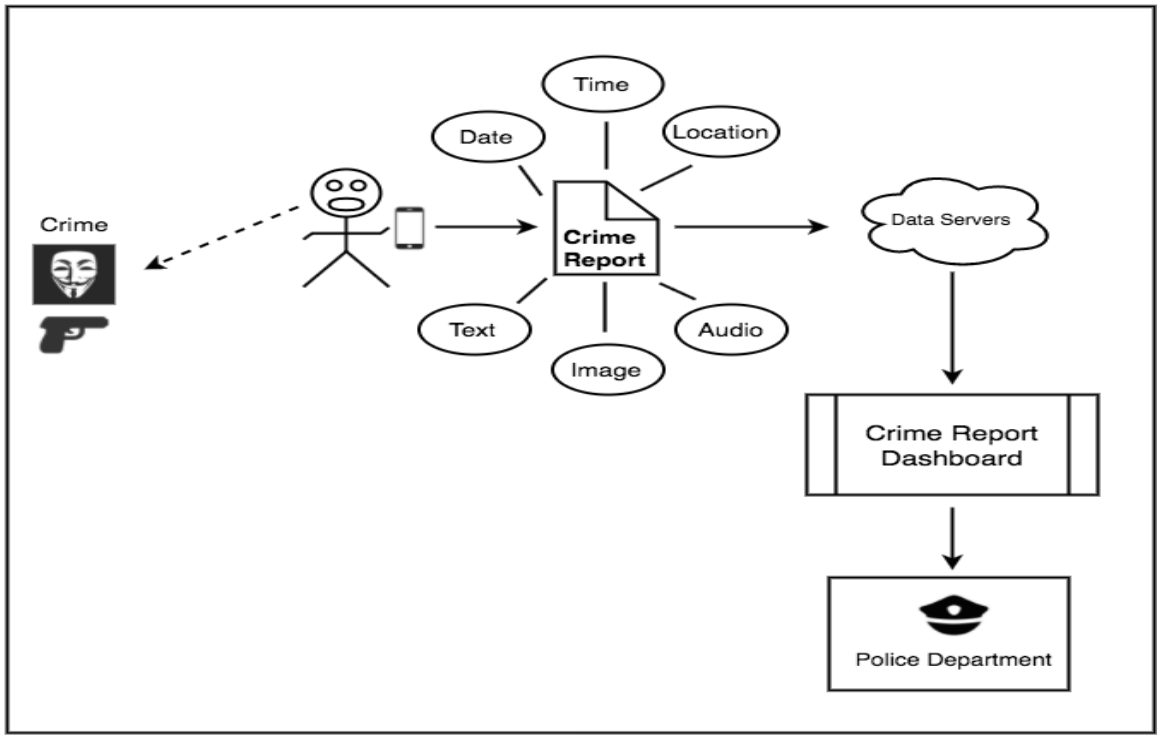


Figure 3.5: Crime Reporting using Mobile Computing

The basic components assembling the architecture for mobile computing approach can be detailed as follows:

1. **Crime Report** via Smartphones

   This basically represents the interface on the client-side, or 'tipster-side', that basically allows a tipster to record a crime on smartphone by using built-in sensors and supplying manual information minimally. In our prototype, we use Android platform to build basic interface that allows an Android smartphone user to report crime. This crime is submitted in the form of a JSON that contains all the data like date, time, location, image URL, etc.

2. **Data Servers**

Once the crime is captured by the tipster, the report is submitted to data servers in the backend, that are responsible for collecting, storing and distributing the information to appropriate dashboards. In our prototype, we use *Krumbs* SDK [22] for creating RESTful web services to collect information, parse the JSON and make the data available on corresponding dashboards. This component may also interact with other external web services like *ClarifAI* [6] for fetching concepts out of a given image. **Note:** For more details about the JSON structure, please refer the Appendix C.

3. **Information Dashboard**

   Based on the requirements, visualisation and analytics can be presented on the dashboard, that can be accessible to tipsters and/or citizens and/or local police. This information dashboard allows people to understand crime situation in real time and thereby take appropriate action accordingly. In our prototype, Krumbs was used to allocate a designated dashboard for visualising all the crime reports and their details like date, time, brief textual description, image and location of the crime.

3.4.2 Demo Walkthrough

Working of this prototype is explained step-wise as follows:

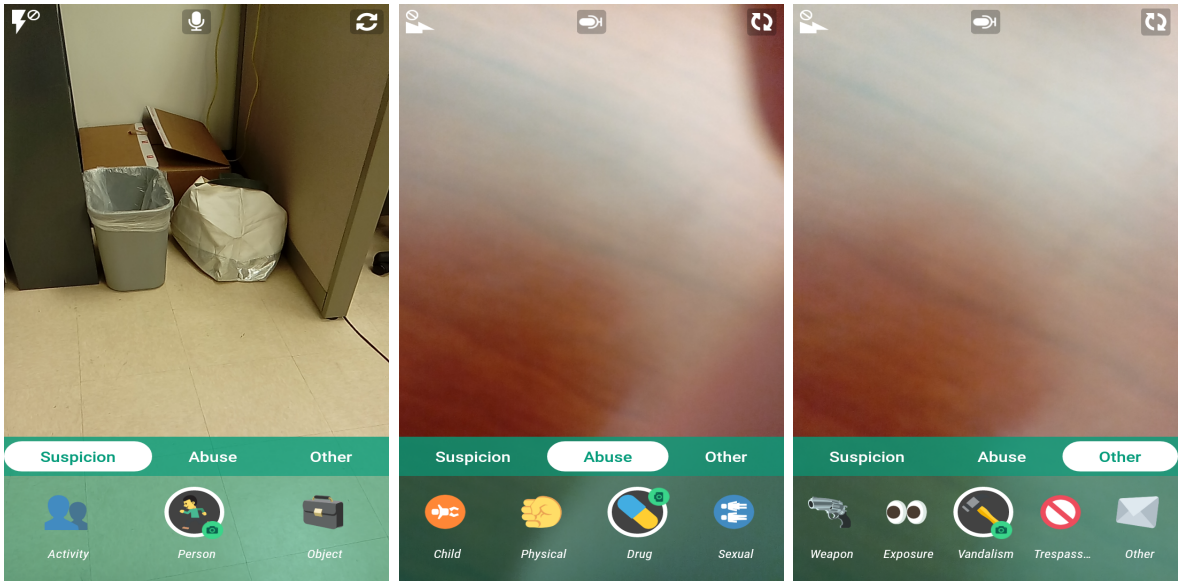


Figure 3.6: Available Crime Categories

1. After starting the Android app prototype, a camera is invoked and the user is presented with crime categories and sub categories as emojis at the bottom 'intent panel'. As shown in the Figure 3.6, currently the prototype supports the following crimes:

   Suspicion with sub categories as Activity, Person and Object.

   Abuse with sub categories as Child, Physical, Drug and Sexual.

   Other with sub categories as Weapon, Exposure, Vandalism, Trespassing and Other.

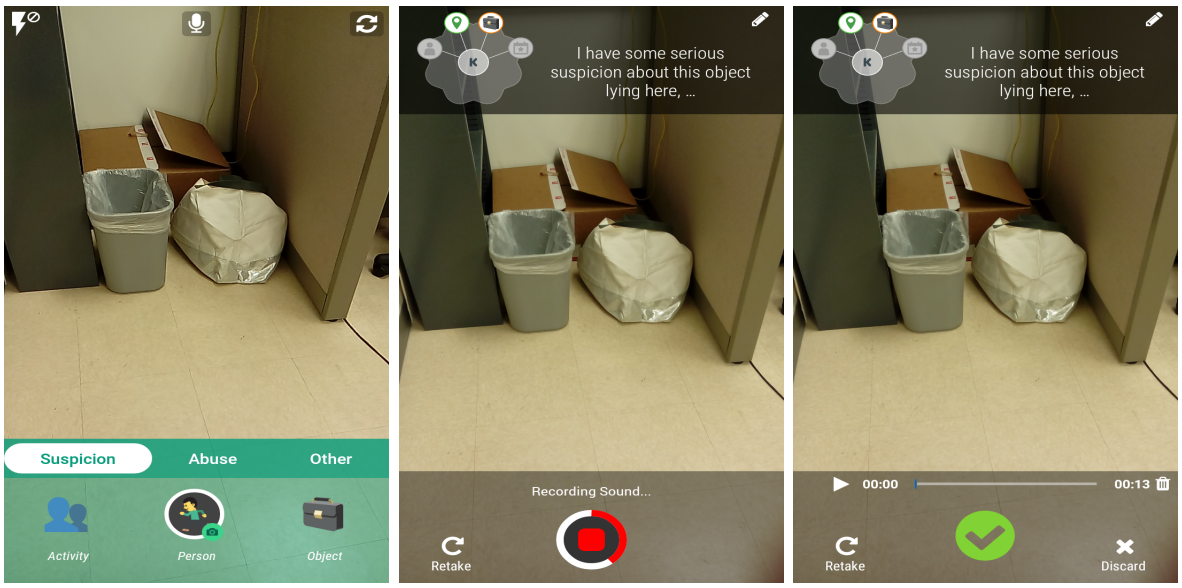


Figure 3.7: Reporting Suspicious Object

2. Once the user has decided what crime category and sub category he wants to report, the next step is to simply point your smartphone to an entity that describes the corresponding crime, and significantly helping the law enforcement in understanding the crime report/tip. In the example show in Fig 3.7, the user identified an abandoned bag lying around in the community hall. User can now simply select on 'Suspicion' as the category and then tap on 'Object' emoji. This action will enable the camera to click a photo immediately. After this, the user can add audio notes that may help in describing additional relevant details. Finally, the user can simply click on the green 'check' icon to submit the crime report. User is also offered a choice to retake the photo taken, discard the audio note or edit the placeholder textual caption at the top.

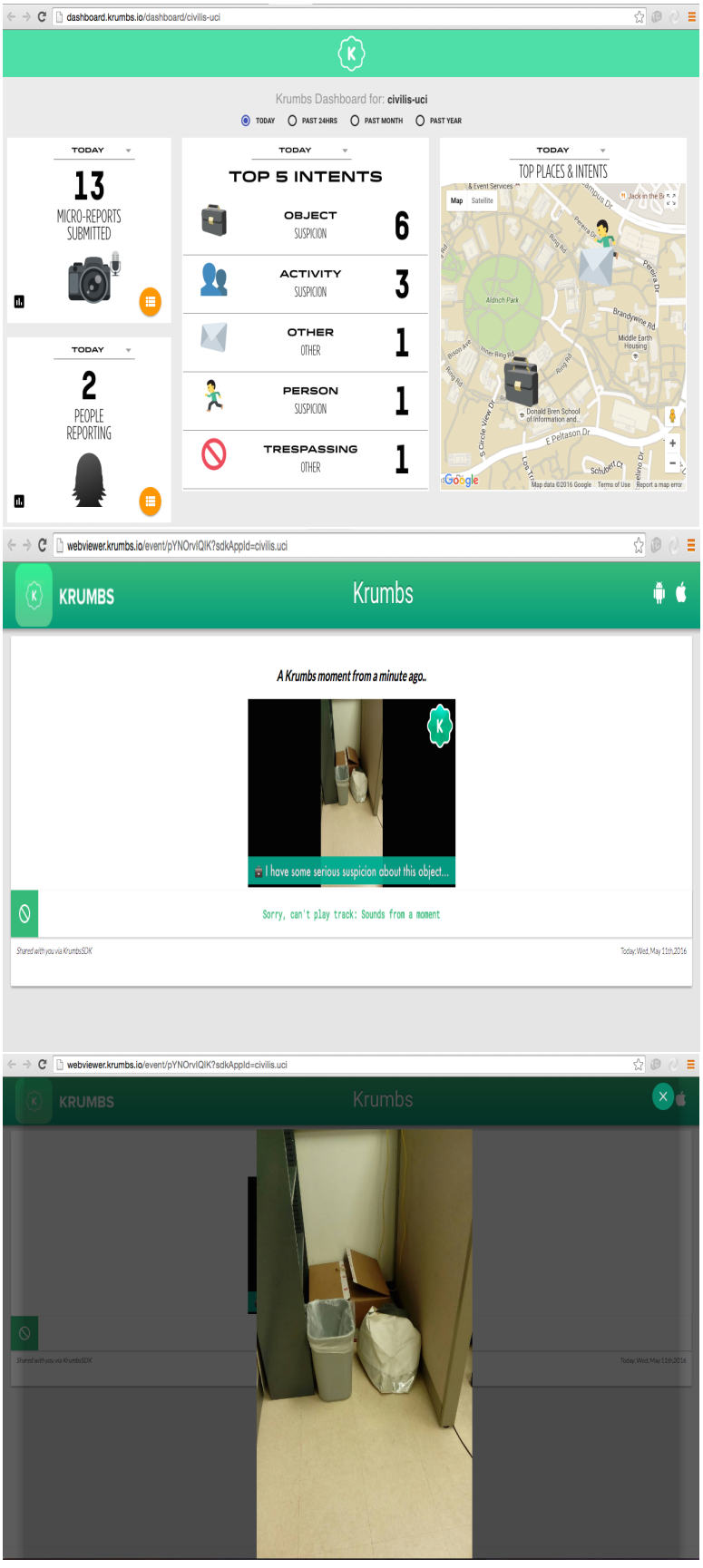


Figure 3.8: Viewing submitted report on Dashboard

3. After a user/tipster has submitted the crime report, the crime details are made available on an information dashboard as seen in Fig 3.8. User or a law enforcement personnel can now browse through the crime reports based on time, location, crime type or users. Additional details like text caption, location, audio notes and enlarged images can be obtained by clicking on any report.

3.4.3 Evaluation

Evaluation of our Mobile Computing approach was conducted in the form of a pilot study with two main components - tipsters and law enforcement agency.

Citizen/Tipster Experiment

1. Participants:

   15 Android smartphone users with age between 21-65 years, different professional backgrounds and geographically distributed across the US, were selected for this pilot study. An instruction list containing steps to install the app, basic app reference guide and Android build apk was distributed to the participants.

2. Process:

   After understanding the instructions on using the app, the participants were requested to record 'mock' crime, emphasizing on something that is very close to 'apparent' reality. Since, recording 'real' crime was not possible, few examples were also suggested to participants to convey the idea of mock crime. Reason for this process was two-fold: (i) Collect crime reports that can be displayed on the dashboard and later evaluated by a law enforcement agency (Details will be discussed in the next sub section) (ii) Offer participants the ability to report crime assuming its real, only to enable them to take the survey more honestly and accurately.

3. Post-reporting survey:

   Once participants finished reporting mock crime via the app, they were requested to take a 10-question survey. The objective of this survey was to collect data related to user experience while reporting crime and offer us insights into the strengths and weaknesses of our approaches (via this app).

4. Results: Responses of this survey were collected on SurveyMonkey over a period of one week. A summarized version of the survey results is described as follows:

Parameter	**Result**
Ease of use	4.3
Speed of crime reporting	4.1
Image as pivot to crime reporting	4.2
Two-way communication	80% (Agreed)
Crime report feedback	3.5
Motivation	53% (Altruism)
Future use	4.5

Table 3.2: Citizen Pilot Study Evaluation Results

Most of the results in Table 3.2 were computed on a Likert scale of 1-5, with 1 being 'Not satisfied at all' and 5 being 'Extremely satisfied'. Factors like ease of use, speed of reporting and image as pivot were rated quite high in our prototype. Crime report feedback, particularly, did not receive great response, which will allow us to improve feedback mechanism in our future prototypes. Surprisingly, around 80% of the total participants were ready to 'Trust' such a system and continue to provide additional crime details through a communication channel. For motivation to report crime, contrary to the 'logical' belief that monetary benefits or social recognition are more likely, more than 50% participants chose altruism over all other options. Finally, most of the

participants agreed to use our application for crime reporting in future.

Note: Detailed analysis of each question and corresponding responses is available in the Appendix A towards the end of this document.

Law Enforcement Agency Experiment

1. Participant:

   In order to simulate a real world scenario where crime reports are being analysed by a law enforcement agency, evaluate the usefulness of our prototype and both our approaches in general, we collaborated with University of California Irvine Police Department (UCI PD) [35]. One Lieutenant and one Police Officer participated with us in evaluating our prototypes.

2. Process:

   After collecting all the mock crime reports submitted by the participants and displaying the crime reports on the dashboard, we requested UCI PD personnel to investigate the crime report and evaluate the usefulness of the information from crime intelligence perspective. Prior to starting the investigation, a general walkthrough of our dashboard was presented. After this, we followed a process that involved selecting few crime reports based on images, crime types or location. For each of these crime reports, the personnel tried to look for all available clues via text, date/time, image concepts, location, etc. After going through about 20 crime reports, all the personnel were requested to take the survey.

3. Post-investigation survey:

   After investigating several reports, two personnel from UCI PD were requested to take a 10-question survey about their investigation experience and usefulness of available information. They were also requested for subjective feedback and suggestions for highlighting strengths, weaknesses and future work recommendations.

4. Results:

   Results of the experiment with the law enforcement personnel (UCI PD) have been summarized on a Likert scale of 1-5, with 1 being 'Not satisfied/important at all' and 5 being 'Extremely satisfied/important'. As shown in Table 3.3, UCI PD rated the quality of our data quite high, in terms of accuracy, time to understand and elimination of noise. Using machine learning and natural language processing (from *AI Approach*) for automatically classifying crime reports and extracting important entities from them, was considered valuable addition to their existing systems.

   To improve our systems from crime intelligence perspective, We also wanted to analyse what kind of information law enforcement emphasises on, for effective action, when they receive a crime report. As shown in Table 3.4, based on traditional approaches to crime intelligence: date, time and location are seen to be of paramount importance even now. Having image and text was considered important, while audio and concepts were mere value-additions for the law enforcement.

   Overall, they gave a positive response to using such mobile-computing-centric systems in future for more efficient crime report investigation.

Parameter	**Result**
Information Accuracy	4.5
Time to understand crime	4.5
Automatic crime report classification and entity identification	4.5
Elimination of Information Noise	4.5
Future use	4.5

Table 3.3: Pilot Study Evaluation Results - UCI PD

Sensor/Multimedia Data	**Result**
Date/Time	5
Location	5
Text	4.5
Audio	4
Image	4.5
Concepts	4

Table 3.4: Sensor and Multimedia data Feedback - UCI PD

Note: Detailed analysis of each question and corresponding responses is available in the Appendix B towards the end of this document.

Chapter 4

Conclusion and Future Work

Realising the gravity of the crime situation across the globe, there is a pressing need for understanding the loopholes in existing approaches to crime reporting. After researching current state-of-the-art, our proposed work offered solutions to some of the existing issues like complicated reporting processes, excessive noise, limited authorized personnel, lack of motivation and one-way communication with fairly positive evaluation results. Without mutual exclusiveness, both Artificial Intelligence and Mobile Computing approaches offer different solutions to similar or different issues at hand. With both these areas of technology continuing to progress at a rapid pace, these approaches will only become more sophisticated and reliable in future.

Our comprehensive study of literary work and existing initiatives also allowed us to explore some of the other critical issues with such systems like elimination of false positives, managing identities, motivating citizens to report crime, standardizing existing systems, etc.

Role of the discussed emerging technologies and approaches will play a pivotal role in shaping the future of crime reporting. However, though the research of our thesis was limited to application of technology to mitigate existing issues, we believe that other areas like public

policy and law will be instrumental for progress in this direction.

In addition to working on the issues mentioned above, we also propose the following areas for future work:

Managing Redundancy

In cases like Boston Marathon Bombings, law enforcement usually gets flooded with too many calls, emails and messages, more or less conveying same piece of information. This causes more 'unique' information to be lost in the huge traffic of redundant information. Thus, our system should be able to address this issue of pulling only unique crime details from a stream of crime reports.

Live action

In order to improve productive and faster response from the law enforcement, a mechanism to prioritize live vs past crimes is important. Although, law enforcement authorities continue to emphasize that 911 must be used for crime-in-action, provision must be made in such crowdsourcing apps to handle emergencies and non-emergencies differently, thus providing the law enforcement with optimum mechanism for effective action.

Intermediate Curators

For crowdsourcing approaches like SAR campaigns to be effective, a middle layer between the citizens and the law enforcement agency is critical. This middle layer, a combination of humans and computers can filter the information/report noise, thereby further improving the nature and scalability of response.

Reasons for crime

Reporting a crime and punishing an identified suspect is like recording the symptoms and then curing the corresponding disease. However, looking at a bigger picture, for preventing such crimes to happen in future, our system needs to accommodate provisions to analyse the root causes behind people committing a crime in the first place.

Bibliography

[1] AlchemyAPI Cognitive APIs. http://www.alchemyapi.com/.

[2] Big Data Usage Per Minute. http://www.inc.com/larry-kim/15-mind-blowing-statistics-reveal-what-happens-on-the-internet-in-a-minute.html.

[3] Bitext Entity Extractor. https://www.bitext.com/text-analysis-api/.

[4] BJS Criminal Victimization 2014. http://www.bjs.gov/content/pub/pdf/cv14.pdf.

[5] City of Madison, Crime Report Dataset. https://data.cityofmadison.com/Police/Police-Incident-Reports/d686-rvcw.

[6] ClarifAI Visual Recognition API and services. https://www.clarifai.com/.

[7] Crowdsourcing Aftermath. http://www.thewire.com/national/2013/04/reddit-find-boston-bombers-founder-interview/64455/.

[8] Crowdsourcing Crime Reporting Issues. http://culturedigitally.org/2012/07/the-problem-with-crowdsourcing-crime-reporting/.

[9] Crowdsourcing Definition. http://www.merriam-webster.com/dictionary/crowdsourcing.

[10] Crowdsourcing for fighting ISIS. http://www.huffingtonpost.com/maha-hosain-aziz/how-crowdsourcing-can-hel_b_7433356.html.

[11] Crowdsourcing for National Security. https://www.rsis.edu.sg/wp-content/uploads/2015/03/PR150317Crowdsourcing-for-National-Security.pdf.

[12] Facebook Statistics. https://zephoria.com/top-15-valuable-facebook-statistics/.

[13] FBI Tips. https://tips.fbi.gov/.

[14] French Government Crime Reporting Portal. https://www.internet-signalement.gouv.fr/.

[15] Future Trends in Policing. http://www.policeforum.org/assets/docs/Free_Online_Documents/Leadership/future%20trends%20in%20policing%202014.pdf.

[16] Global Terrorism Index 2015. http://economicsandpeace.org/wp-content/uploads/2015/11/Global-Terrorism-Index-2015.pdf.

[17] Hoaxmap. http://hoaxmap.org/ueber.html.

[18] Hosting a Facebook Messenger Bot. https://developers.facebook.com/docs/messenger-platform/quickstart.

[19] IBM Bluemix. http://www.ibm.com/cloud-computing/bluemix/.

[20] IBM Watson Dialog. http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/dialog.html.

[21] IBM Watson Relationship Extraction. http://www.ibm.com/smarterplanet/us/en/ibmwatson/developercloud/relationship-extraction.html.

[22] Krumbs Microreporting and Analytics. https://krumbs.net/.

[23] LAPD's iWatch LA. http://www.lapdonline.org/iwatchla.

[24] Lone Wolf Terrorism Definition. https://en.wikipedia.org/wiki/Lone_wolf_(terrorism).

[25] Machine Learning Definition. http://whatis.techtarget.com/definition/machine-learning.

[26] MonkeyLearn for machine learning. http://monkeylearn.com/.

[27] Natural Language Processing Definition. http://www.webopedia.com/TERM/N/NLP.html.

[28] NCRB Crime Statistics 2014. http://ncrb.nic.in/StatPublications/CII/CII2014/snapshots.pdf.

[29] New York State's See Send. https://www.ny.gov/programs/see-something-send-something#.

[30] See Something, Say Something. https://www.dhs.gov/see-something-say-something/about-campaign.

[31] Smartphone Sensors. https://developer.android.com/guide/topics/sensors/sensors_overview.html.

[32] Spot Crime. http://spotcrime.com/.

[33] Suspicious Activity Reports from Mall of America. http://www.npr.org/2011/08/18/139756444/database-mall-of-america-suspicious-activity-reports.

[34] Thomas Reuters' Open Calais. http://www.opencalais.com/.

[35] UC Irvine Police Department. http://police.uci.edu/.

[36] UK Crime Statistics. http://www.ukcrimestats.com/.

[37] UK Metropolitan Police Anti-Terrorist Hotline. https://secure.met.police.uk/athotline/.

[38] Understanding bot structure, api.ai. https://docs.api.ai/docs/key-concepts.

[39] UShahidi. https://en.wikipedia.org/wiki/Ushahidi.

[40] Wikistrat. http://www.wikistrat.com/.

[41] L. Beam. Leaderless resistance. *The Seditionist*, 12:12–13, 1992.

[42] R. R. Bennett and R. B. Wiegand. Observations on crime reporting in a developing nation*. *Criminology*, 32(1):135–148, 1994.

[43] A. B. Breivik. 2083: a european declaration of independence. *Retrieved December*, 6(2013):2010–2019, 2011.

[44] C. W. Bruce. Fundamentals of crime analysis. *Exploring rime Analysis*, pages 11–36, 2004.

[45] F. Burton and S. Stewart. The lone wolfdisconnect. *Terrorism Intell. Rep., Stratfor*, 2008.

[46] M. Byrne Evans, K. O'Hara, T. Tiropanis, and C. Webber. Crime applications and social machines: crowdsourcing sensitive data. In *Proceedings of the 22nd international conference on World Wide Web companion*, pages 891–896. International World Wide Web Conferences Steering Committee, 2013.

[47] D. L. Carter. Brief history of law enforcement intelligence: Past practice and recommendations for change. *Trends in Organized Crime*, 8(3):51–62, 2005.

[48] T. Cincotta. Platform for prejudice. *ThePublicEye*, page 17, 2010.

[49] L. Cohen. Ways law enforcement uses social media to fight crime. *Retrieved on*, 12, 2011.

[50] A. Doan, R. Ramakrishnan, and A. Y. Halevy. Crowdsourcing systems on the world-wide web. *Communications of the ACM*, 54(4):86–96, 2011.

[51] M. Goodman. From crowdsourcing to crime-sourcing: The rise of distributed criminality, 2011.

[52] M. Hamilton, F. Salim, E. Cheng, and S. Choy. Transafe: a crowdsourced mobile platform for crime and safety perception management. *ACM SIGCAS Computers and Society*, 41(2):32–37, 2011.

[53] G. Markowsky. Crowdsourcing, big data and homeland security. In *Technologies for Homeland Security (HST), 2013 IEEE International Conference on*, pages 772–778. IEEE, 2013.

[54] I. A. of Chiefs of Police/National Law Enforcement Policy Ctr and U. S. of America. Social media: Concepts and issues paper. 2010.

[55] D. Osimo. Web 2.0 in government: Why and how. *Institute for Prospectice Technological Studies (IPTS), JRC, European Commission, EUR*, 23358, 2008.

[56] L. A. Peters. *Utilizing Social Media to Further the Nationwide Suspicious Activity Reporting Initiative*. PhD thesis, Monterey, California. Naval Postgraduate School, 2012.

[57] M. A. Randol. Terrorism information sharing and the nationwide suspicious activity report initiative: Background and issues for congress. 2009.

[58] P. M. Regan, T. Monahan, and K. Craven. Constructing the suspicious data production, circulation, and interpretation by dhs fusion centers. *Administration & Society*, 47(6):740–762, 2015.

[59] B. A. Roberts. *Managing Suspicious Activity Reporting Systems at Small Agency Police Departments*. PhD thesis, Monterey, California. Naval Postgraduate School, 2012.

[60] M. Sageman. *Leaderless jihad: Terror networks in the twenty-first century*. University of Pennsylvania Press, 2011.

[61] R. R. Soares. Crime reporting as a measure of institutional development*. *Economic Development and cultural change*, 52(4):851–871, 2004.

[62] R. Spaaij and M. S. Hamm. Key issues and research agendas in lone wolf terrorism. *Studies in Conflict & Terrorism*, 38(3):167–178, 2015.

[63] D. Tewksbury. Crowdsourcing homeland security: The texas virtual borderwatch and participatory citizenship. *Surveillance & Society*, 10(3/4):249, 2012.

[64] J. Woodcock. Leveraging social media to engage the public in homeland security. Technical report, DTIC Document, 2009.

[65] M. Xie. Area differences and time trends in crime reporting: Comparing new york with other metropolitan areas. *Justice Quarterly*, 31(1):43–73, 2014.

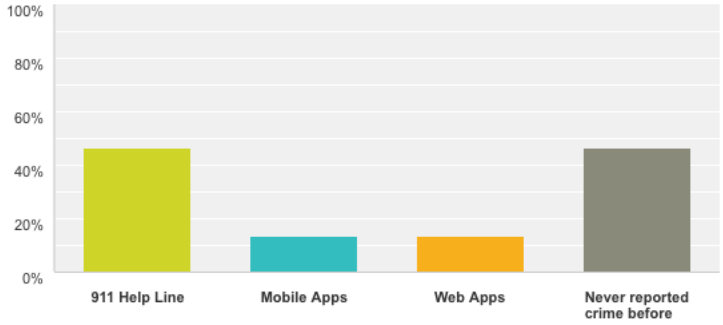
Appendix A

Citizen Survey Results

The charts below are actual questions and responses from the survey taken by 15 participants (citizens/tipsters). Some of the responses have been visualised as Pie charts for better understanding of the data distribution.

In general, how do you report crime?

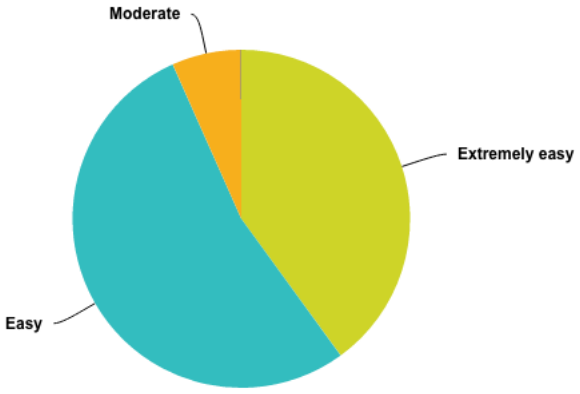
Answered: 15 Skipped: 0



Answer Choices	Responses	
911 Help Line	46.67%	7
Mobile Apps	13.33%	2
Web Apps	13.33%	2
Never reported crime before	46.67%	7
Total Respondents: 15		

How easy was the process of reporting crime using our app?

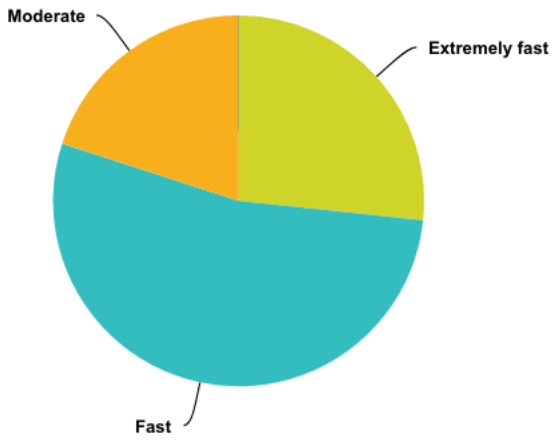
Answered: 15 Skipped: 0



	Extremely easy	Easy	Moderate	Difficult	Extremely difficult	Total	Weighted Average
(no label)	40.00% 6	53.33% 8	6.67% 1	0.00% 0	0.00% 0	15	1.67

Based on your previous experience with apps (crime or non-crime), how would you consider crime reporting process using our app, timewise?

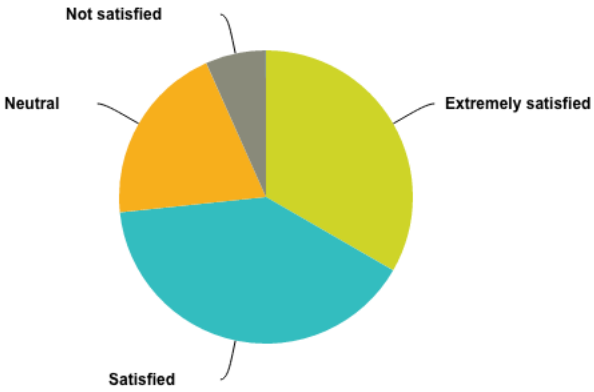
Answered: 15 Skipped: 0



	Extremely fast	Fast	Moderate	Slow	Extremely slow	Total	Weighted Average
(no label)	26.67% 4	53.33% 8	20.00% 3	0.00% 0	0.00% 0	15	1.93

Unlike some apps that request you to manually enter information like time, date, location, user ID (if chosen), etc, how satisfied are you with built in sensors automatically taking input in our app?

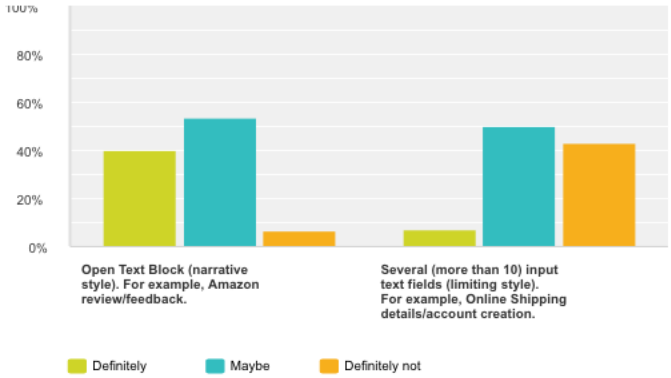
Answered: 15 Skipped: 0



	Extremely satisfied	Satisfied	Neutral	Not satisfied	Not satisfied at all	Total	Weighted Average
(no label)	33.33% 5	40.00% 6	20.00% 3	6.67% 1	0.00% 0	15	2.00

Based on your previous experience with apps (crime or non-crime), would you prefer a narrative approach for crime reporting with just one input text block or 10-15 different form fields?

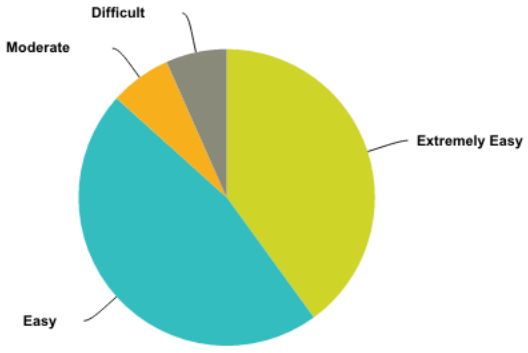
Answered: 15 Skipped: 0



	Definitely	Maybe	Definitely not	Total
Open Text Block (narrative style). For example, Amazon review/feedback.	40.00% 6	53.33% 8	6.67% 1	15
Several (more than 10) input text fields (limiting style). For example, Online Shipping details/account creation.	7.14% 1	50.00% 7	42.86% 6	14

In our app, how easy was it for you to convey your crime related information using image as the main component?

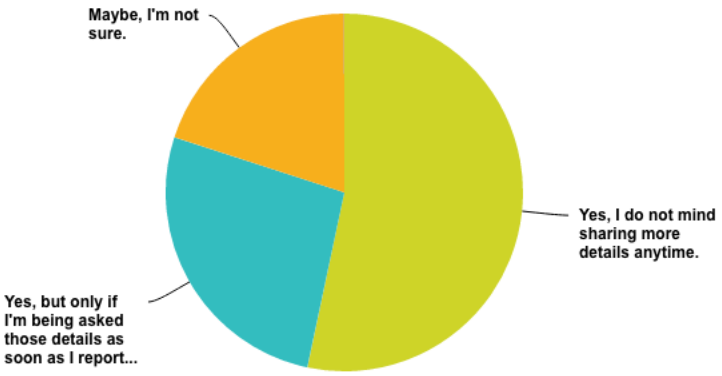
Answered: 15 Skipped: 0



	Extremely Easy	Easy	Moderate	Difficult	Extremely Difficult	Total	Weighted Average
(no label)	40.00% 6	46.67% 7	6.67% 1	6.67% 1	0.00% 0	15	1.80

In general, after you report crime, would you like to be asked additional details (anonymously or non-anonymously) related to the crime?

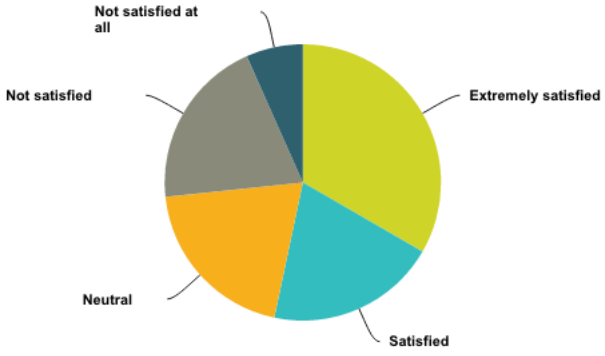
Answered: 15 Skipped: 0



Answer Choices	Responses	
Yes, I do not mind sharing more details anytime.	53.33%	8
Yes, but only if I'm being asked those details as soon as I report crime.	26.67%	4
Maybe, I'm not sure.	20.00%	3
No, I do not want any further communication after I have submitted the crime report.	0.00%	0
Total		15

In our app, how satisfied are you with the instant feedback showing that your report has reached the authorities?

Answered: 15 Skipped: 0

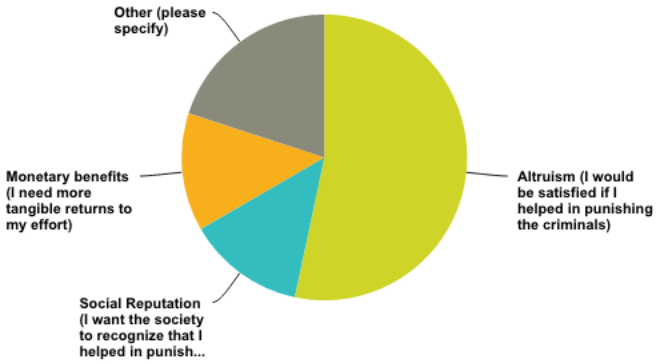


	Extremely satisfied	Satisfied	Neutral	Not satisfied	Not satisfied at all	Total	Weighted Average
(no label)	33.33% 5	20.00% 3	20.00% 3	20.00% 3	6.67% 1	15	2.47

bb

In general, what would motivate you to report crime in future?

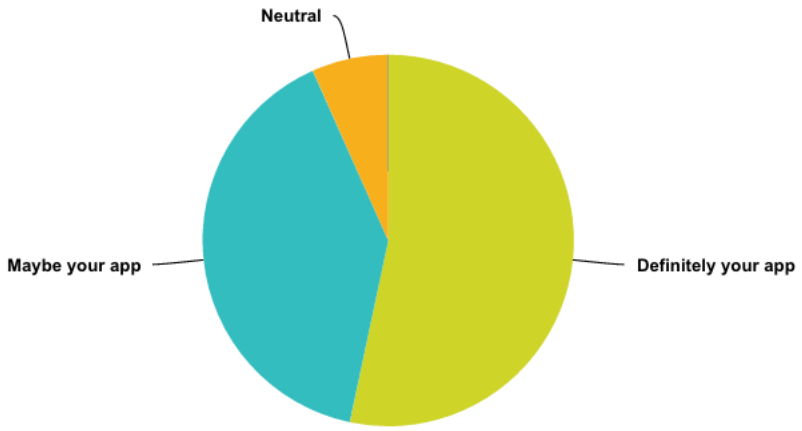
Answered: 15 Skipped: 0



Answer Choices	Responses
Altruism (I would be satisfied if I helped in punishing the criminals)	53.33% 8
Social Reputation (I want the society to recognize that I helped in punishing the criminals)	13.33% 2
Monetary benefits (I need more tangible returns to my effort)	13.33% 2
Other (please specify) Responses	20.00% 3
Total	15

For reporting crime, in comparison between other existing apps/methods and our app, which one would you prefer?

Answered: 15 Skipped: 0



	Definitely your app	Maybe your app	Neutral	Maybe other existing apps/methods	Definitely other existing apps/methods	Total	Weighted Average
(no label)	53.33% 8	40.00% 6	6.67% 1	0.00% 0	0.00% 0	15	1.53

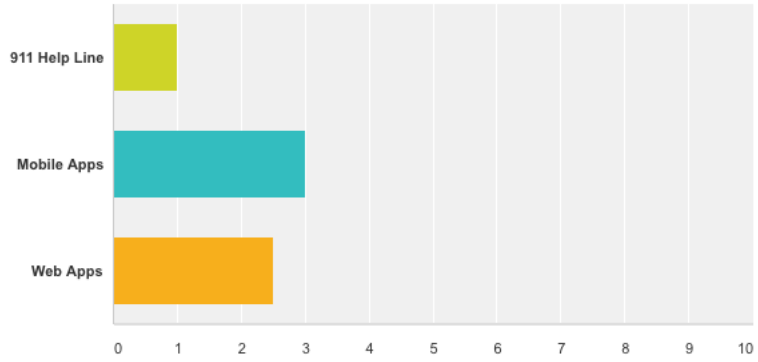
Appendix B

UCI PD Survey Results

The charts below are actual questions and responses from the survey taken by two law enforcement personnel (UCI PD).

In general, what is the crime reports distribution across different sources?

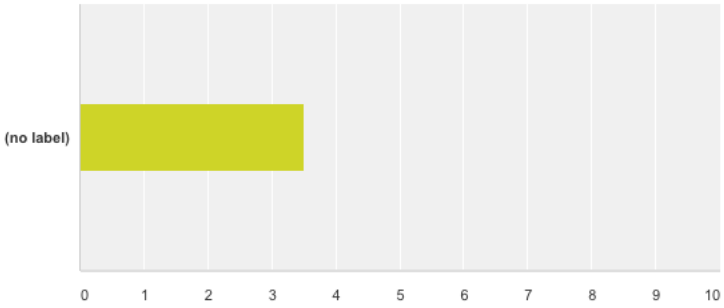
Answered: 2 Skipped: 0



	High	Moderate	Low	Total	Weighted Average
911 Help Line	100.00% 2	0.00% 0	0.00% 0	2	1.00
Mobile Apps	0.00% 0	0.00% 0	100.00% 2	2	3.00
Web Apps	0.00% 0	50.00% 1	50.00% 1	2	2.50

In general, in terms of information accuracy, how satisfied was your experience with crime reported through other existing apps?

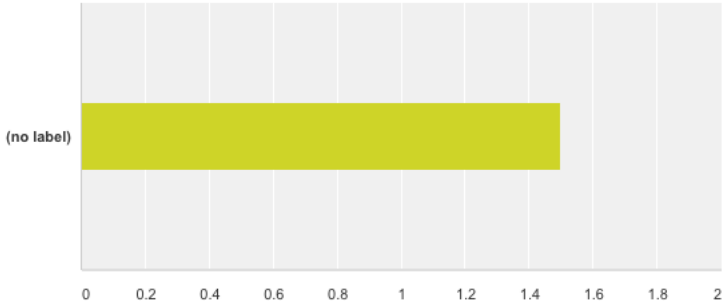
Answered: 2 Skipped: 0



	Extremely satisfied	Satisfied	Neutral	Not satisfied	Not satisfied at all	Total	Weighted Average
(no label)	0.00% 0	0.00% 0	50.00% 1	50.00% 1	0.00% 0	2	3.50

In terms of information accuracy, how satisfied was your experience with crime reported using our app?

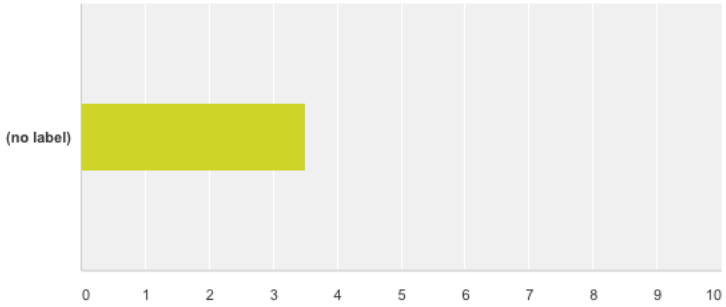
Answered: 2 Skipped: 0



	Extremely satisfied	Satisfied	Neutral	Not satisfied	Not satisfied at all	Total	Weighted Average
(no label)	50.00% 1	50.00% 1	0.00% 0	0.00% 0	0.00% 0	2	1.50

In general, how long does it take to understand a crime reported through other existing apps, for taking appropriate action?

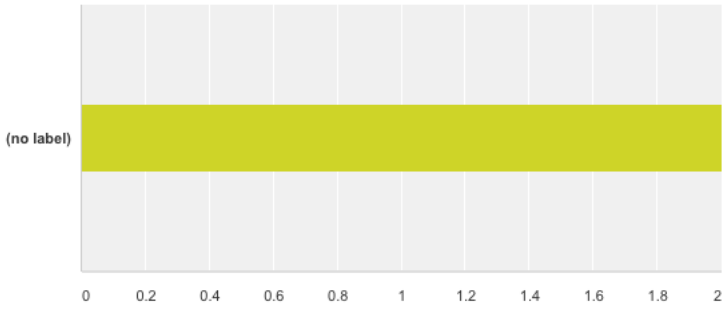
Answered: 2 Skipped: 0



	Extremely fast	Fast	Moderate	Slow	Extremely slow	Total	Weighted Average
(no label)	0.00% 0	0.00% 0	50.00% 1	50.00% 1	0.00% 0	2	3.50

How long did it take to understand a crime reported through our app, for taking appropriate action?

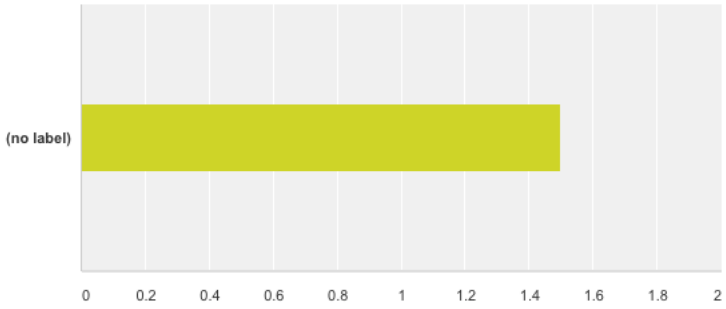
Answered: 2 Skipped: 0



	Extremely fast	Fast	Moderate	Slow	Extremely slow	Total	Weighted Average
(no label)	0.00% 0	100.00% 2	0.00% 0	0.00% 0	0.00% 0	2	2.00

How useful was the automatic crime report categorisation and entity identification for taking appropriate action?

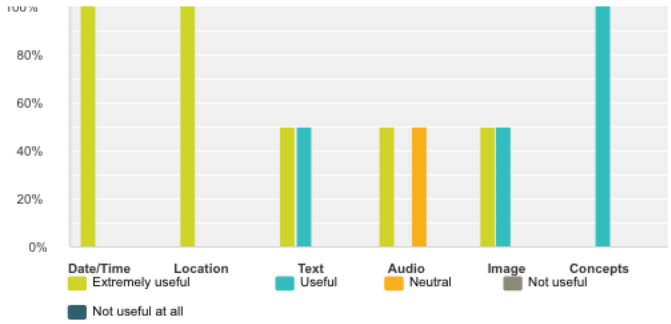
Answered: 2 Skipped: 0



	Extremely useful	Useful	Neutral	Not useful	Not useful at all	Total	Weighted Average
(no label)	50.00% 1	50.00% 1	0.00% 0	0.00% 0	0.00% 0	2	1.50

How useful was the sensor and multimedia data like date/time, location, image, audio, concepts, etc. in understanding the crime report?

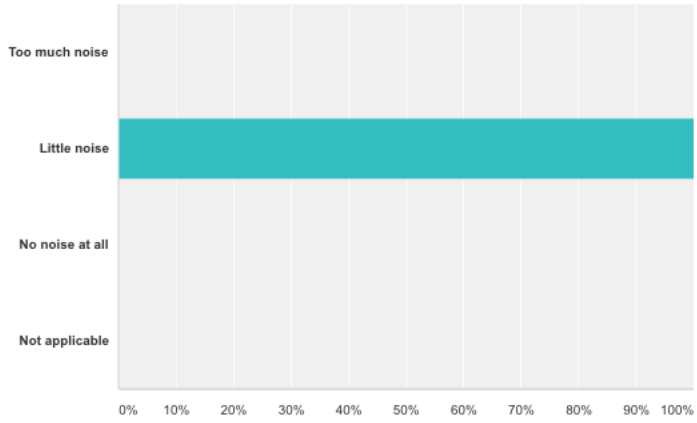
Answered: 2 Skipped: 0



	Extremely useful	Useful	Neutral	Not useful	Not useful at all	Total
Date/Time	100.00% 2	0.00% 0	0.00% 0	0.00% 0	0.00% 0	2
Location	100.00% 2	0.00% 0	0.00% 0	0.00% 0	0.00% 0	2
Text	50.00% 1	50.00% 1	0.00% 0	0.00% 0	0.00% 0	2
Audio	50.00% 1	0.00% 0	50.00% 1	0.00% 0	0.00% 0	2
Image	50.00% 1	50.00% 1	0.00% 0	0.00% 0	0.00% 0	2
Concepts	0.00% 0	100.00% 2	0.00% 0	0.00% 0	0.00% 0	2

In general, through other existing apps, how much information noise do you receive in your crime reports? Example: Inaccurate details, racist complaints, pranks, etc.

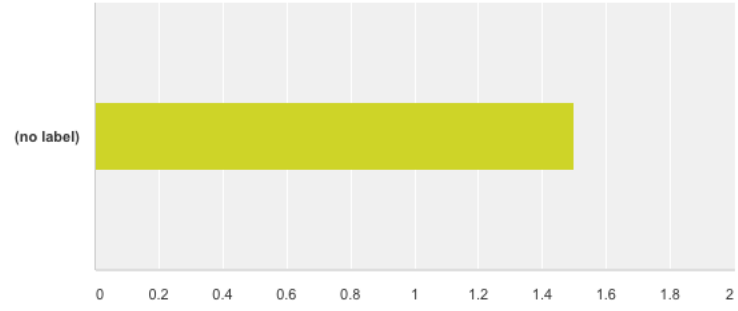
Answered: 2 Skipped: 0



Answer Choices	Responses	
Too much noise	0.00%	0
Little noise	100.00%	2
No noise at all	0.00%	0
Not applicable	0.00%	0
Total		2

How useful was our app (using AI and sensors multimedia data) in reducing crime report information noise?

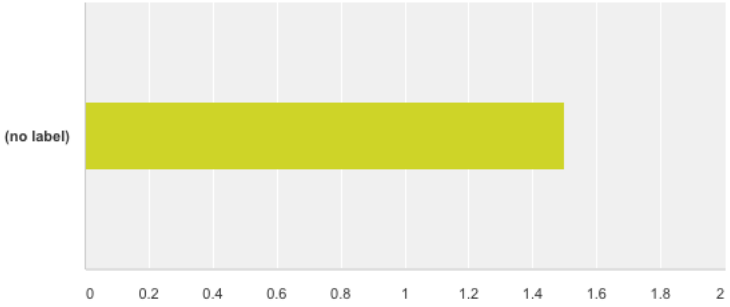
Answered: 2 Skipped: 0



	Very useful	Useful	Neutral	Not useful	Not useful at all	Total	Weighted Average
(no label)	50.00% 1	50.00% 1	0.00% 0	0.00% 0	0.00% 0	2	1.50

For receiving crime reports, in comparison between other existing apps/methods and our app, which one would you prefer?

Answered: 2 Skipped: 0



	Definitely your app	Maybe your app	Neutral	Maybe other existing apps/methods	Definitely other existing apps/methods	Total	Weighted Average
(no label)	50.00% 1	50.00% 1	0.00% 0	0.00% 0	0.00% 0	2	1.50

Appendix C

Example Media JSON

The following is a sample Media JSON containing all the data like ID, concepts present in the image, date/time, location, crime category, image URL, etc. This JSON is generated on the client side (Android app), and then posted to our web servers for visualisation on our dashboards.


```
{
id: "rjO298GObn",
media: [
{
what: [
{
concept_name: "wheel",
confidence: 0.99959755
},
{
concept_name: "bike",
```

```
confidence: 0.99781644
},
{
concept_name: "no person",
confidence: 0.99115336
},
{
concept_name: "transportation system",
confidence: 0.9730878
},
{
concept_name: "bicycle",
confidence: 0.9640044
},
{
concept_name: "spoke",
confidence: 0.96384513
}
],
when: {
start_time: "2016-05-27T00:54:55.2Z",
end_time: "2016-05-27T00:54:55.2Z"
},
where: {
geo_location: {
latitude: 33.643856,
longitude: -117.84157
```

```
},
revgeo_places: [
{
name: "Ring Rd",
country: "United States",
latitude: 33.643898,
longitude: -117.84146,
city: "Irvine",
street: "Ring Rd",
unformatted_address: "Ring Rd,Irvine, CA 92617"
}
]
},
why: [
{
intent_name: "Vandalism",
intent_index_in_category: 3,
intent_used_synonym: "Vandalism",
intent_category_name: "Other",
intent_emoji_id: "jLszIRTfVz",
intent_emoji_unicode: "1f528"
}
],
who: {
creator: [
{
username: "312f6945-873d-42e6-81ee-a579e3f3328d"
```

```
}
]
},
media_source: {
default_src: "https://d3j4aoik7k8ki7.cloudfront.net/
315f80aa-2b1d-46c9-98d9-2fb7d3b75efd.jpeg",
mime_type: "image/jpeg",
raw: "https://d3j4aoik7k8ki7.cloudfront.net/
315f80aa-2b1d-46c9-98d9-2fb7d3b75efd.jpeg"
},
caption: "Why would someone cause damage to this, At Ring Rd",
media_type: "photo"
}
],
start_time: "2016-05-27T00:54:55.2Z",
end_time: "2016-05-27T00:54:55.2Z"
}
```

Appendix D

Messenger Bot Sample Logic

The following two figures represent sample intent (bot action based on user input) and entity (concepts like suspect description) logic for our messenger bot. These are actual screenshots from *api.ai* developer console.

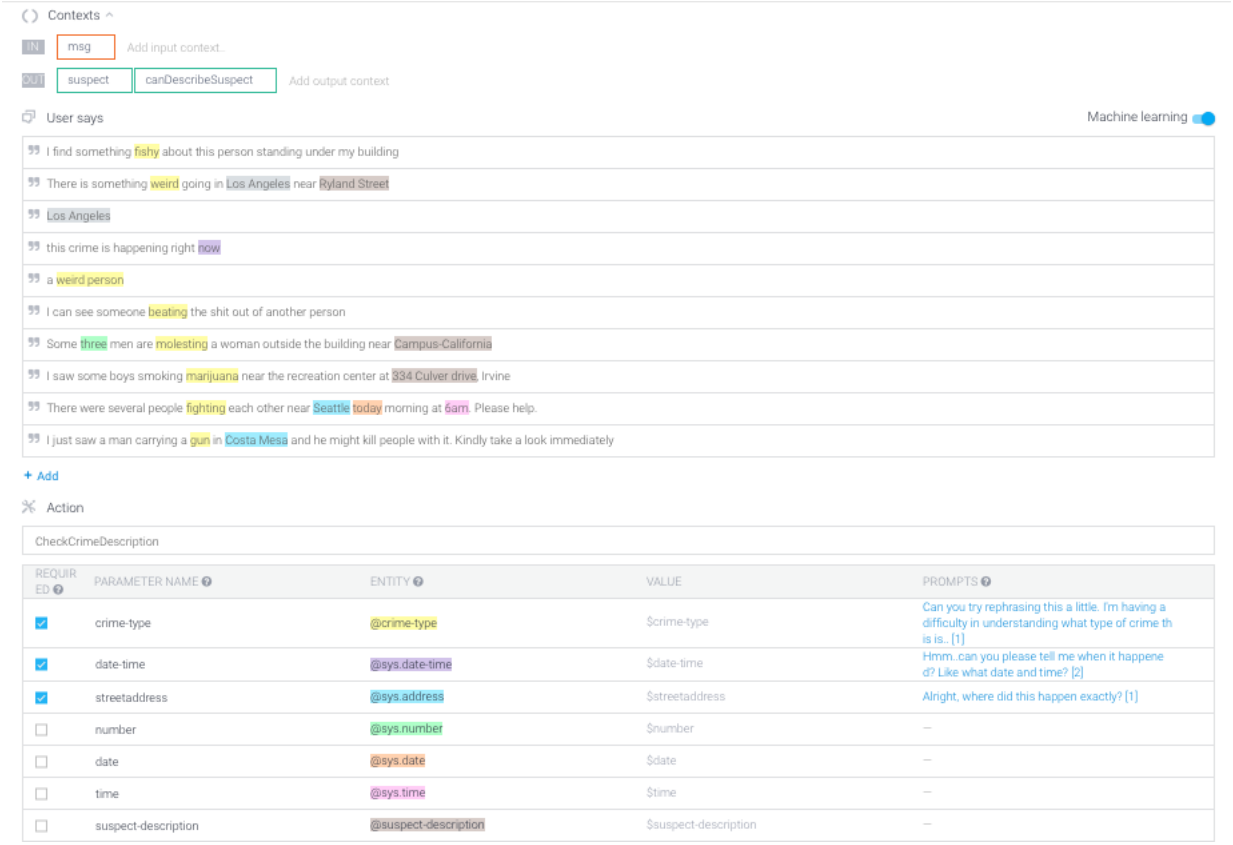


Figure D.1: Intent setup to create decision trees

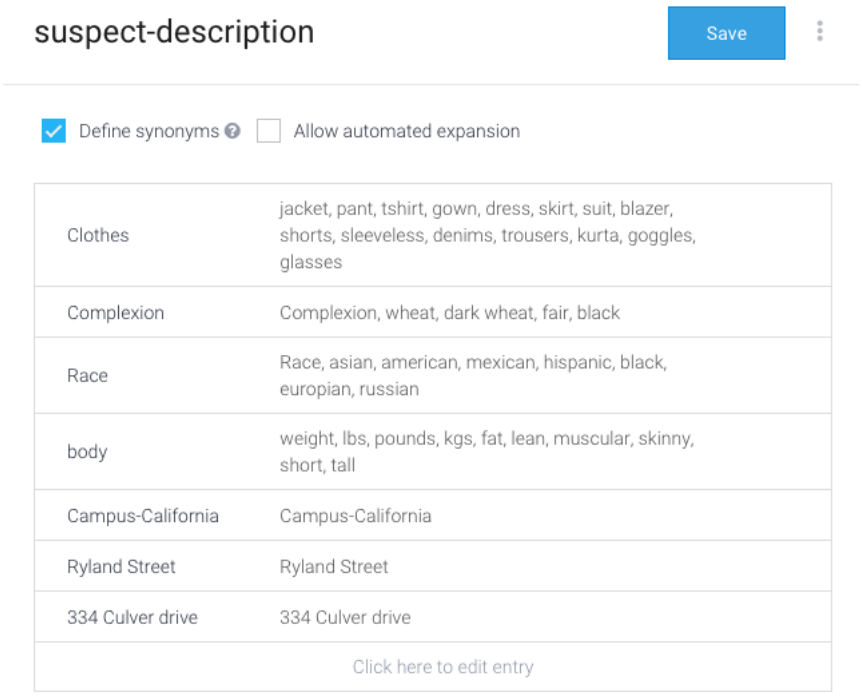


Figure D.2: Entity description

Appendix E

Messenger Bot Sample JSON

The following JSON represents a snapshot of a crime reporter's conversation with our messenger bot. It highlights some of the entities that have been extracted from user input and corresponding bot response.


```
{
  "id": "2a89a13a-e62d-4309-a197-dd558315df8c",
  "timestamp": "2016-06-02T03:55:32.022Z",
  "result": {
    "source": "agent",
    "resolvedQuery": "66000 Verano Road",
    "action": "CheckCrimeDescription",
    "actionIncomplete": false,
    "parameters": {
      "city": "",
      "crime-type": "Physical Abuse",
      "date": "",
```

```
  "date-time": "2016-06-01T21:00:00-07:00",
  "number": "",
  "streetaddress": "66000 Verano Road",
  "suspect-description": "",
  "time": ""
},
"contexts": [
  {
    "name": "msg",
    "parameters": {
      "date": "",
      "number": "",
      "city": "",
      "date-time": "2016-06-01T21:00:00-07:00",
      "crime-type": "Physical Abuse",
      "streetaddress": "66000 Verano Road",
      "time": "",
      "suspect-description": ""
    },
    "lifespan": 2
  },
  {
    "name": "suspect",
    "parameters": {
      "date": "",
      "number": "",
      "city": "",
```

```
      "date-time": "2016-06-01T21:00:00-07:00",
      "crime-type": "Physical Abuse",
      "streetaddress": "66000 Verano Road",
      "time": "",
      "suspect-description": ""
    },
    "lifespan": 5
  },
  {
    "name": "candescribesuspect",
    "parameters": {
      "date": "",
      "number": "",
      "city": "",
      "date-time": "2016-06-01T21:00:00-07:00",
      "crime-type": "Physical Abuse",
      "streetaddress": "66000 Verano Road",
      "time": "",
      "suspect-description": ""
    },
    "lifespan": 5
  }
],
"metadata": {
  "intentId": "7913f405-34e8-45e7-b14c-5d5f30f07a87",
  "webhookUsed": "false",
  "intentName": "cr-content"
```

```
    },
    "fulfillment": {
      "speech": "Alright, I've noted your crime report
       related to Physical Abuse happening at
                 66000 Verano Road. Can you give me more
                 details about the suspect?"
    },
    "score": 1
  },
  "status": {
    "code": 200,
    "errorType": "success"
  }
}
```