

**PREDICTION OF GEOGRAPHICAL LOCATION  
BASED ON SOCIAL MEDIA BIG DATA ANALYTICS**

التنبؤ بالمواقع الجغرافية اعتماداً على تحليل بيانات مواقع التواصل  
الاجتماعي

**Prepared by:**

**Raneem Naseem Abu Hussein**

**Supervised by:**

**Prof. Alaa Hussein Al-Hamami**

**This thesis is submitted in partial fulfillment of the requirement for the degree of Master in**

**Computer Science**

**Department of Computer Science**

**College of Computer Science and Informatics**

**Amman Arab University**

**July 2015**



AMMAN ARAB UNIVERSITY

Form (9)

College of Scientific Research and Graduate Studies

Authorization

We, the undersigned, pledge to grant Amman Arab University for discretion in the publication of the academic content of the thesis, so that the intellectual property rights of a Master thesis be back to the university in accordance with the laws, regulations and instructions relating to intellectual property and patent.

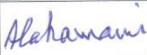
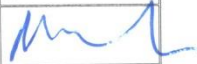

Advisor Name	Co-advisor Name	Student Name
Prof. Alaa Hussein Al Hamami		Raneem Naseem Abu Hussein
Signature: <i>Alaahamami</i> Date: 7.11.2015	Signature: Date:	Signature: <i>[Signature]</i> Date: 7.11.2015

Amman Arab University - Jordan  
 Amman Arab University - Jordan  
 Amman Arab University - Jordan  
 Amman Arab University - Jordan

## Committee Members' Decision

The thesis entitled: "PREDICTION OF LOCATION BASED ON SOCIAL MEDIA BIGDATA ANALYTICS" was submitted by the student **Raneem Naseem Abu Hussein**, was examined and approved on 29/8/2015.

## Committee Members

Name		Signature
Prof. Alaa Hussein Al Hamami	Advisor	
Prof. Omar Aljarráh	Member	
DR. Mahmoud Al Shbool	External/ Member	

## Acknowledgments

There are a number of people without whom this thesis might not have been written and to whom I am greatly indebted.

To my mother, Maisa, who continues to learn, grow and develop and who has been a source of encouragement and inspiration to me throughout my life, a very special thank you for providing a 'writing space' and for nurturing me through the months of writing. In addition, I give thanks for the myriad of ways in which throughout my life you have actively supported me in my determination to find and realize my potential, as well as to make this contribution to the world.

I cannot begin to express my unfailing gratitude and love to my fiancé, Qusai for his continued and unfailing love, support and understanding during my pursuit of my Master degree that has made the completion of this thesis possible. You were always around at times I thought that it is impossible to continue, but you helped me to keep things in perspective. I greatly value your contribution and deeply appreciate your belief in me.

I am grateful to some people who worked hard with me from the beginning until the completion of the present research particularly my supervisor Prof. Alaa Al-Hamami, who has always been generous during all phases of the research, and I highly appreciate his efforts extended upon me.

I also highly appreciate efforts expended from my university instructors especially Dr. Mohammed Otair who help me to develop my skills and to become stronger, Lastly, I would to thank my friends who supported me throughout his process.

## DEDICATION

I dedicate my thesis work to my family and many friends.

A special feeling of gratitude goes to my loving parents, Naseem and Maisa.

My Mom, Maisa, who is her words of encouragement and push for tenacity ring in my ears.

My lovely fiancé Qusai, who has supported me throughout this process and his constant encouragement when the tasks seemed arduous and insurmountable.

My amazing brother Moauth and my sisters Raghad and Mera who never left my side.

Also, I dedicate this work and give special thanks to my supervisor Prof. Alaa Al-Hamami for being with me throughout the entire masters program.

# Table of Contents

<b>AUTHORIZATION.....</b>	<b>II</b>
<b>COMMITTEE DECISION .....</b>	<b>III</b>
<b>ACKNOWLEDGMENTS .....</b>	<b>IV</b>
<b>DEDICATION.....</b>	<b>VI</b>
<b>TABLE OF CONTENTS.....</b>	<b>VII</b>
<b>LIST OF FIGURES.....</b>	<b>VIII</b>
<b>LIST OF ABBREVIATIONS .....</b>	<b>IX</b>
<b>ABSTRACT ARABIC .....</b>	<b>XI</b>
<b>ABSTRACT .....</b>	<b>XIII</b>
<b>CHAPTER ONE: INTRODUCTION.....</b>	<b>1</b>
<b>CHAPTER TWO: LITERATURE REVIEW .....</b>	<b>28</b>
<b>CHAPTER THREE: THE PROPOSED SYSTEM DESIGN .....</b>	<b>48</b>
<b>CHAPTER FOUR: RESULT AND DISCUSSIONS.....</b>	<b>75</b>
<b>CHAPTER FIVE: CONCLUSION AND FUTURE WORKS .....</b>	<b>88</b>
<b>REFERENCES.....</b>	<b>90</b>

## List of Figures

<b>Figure</b>	<b>Content</b>	<b>Page</b>
1	Computer Network	2
2	TCP/IP Architecture	5
3	Internet Users in the World	8
4	Wireless Local Networks	11
5	Wireless Communications	11
6	Use Social Network Sites	16
7	Arab Publics Share	16
8	How Use Social Network Site	17
9	Social Media Site	18
10	Industry lookup	51
11	Locations lookup	52
12	Entity lookup	52
13	Facebook, .Net app	59
14	Industry data dictionary	61
15	Data normalization process	63
16	Database schema	64
17	Sample of collected posts	72
18	Sample of normalized posts	74
19	Classification process results	76
20	Calculated post rate	77



## List of Abbreviations

API	Application Programming Interface
ASP.Net	Active Server Page
ATM	automated teller machine
CID	Crime Investigation Department
CRM	Customer relationship management
CSS	Cascading Style Sheets
E	Electronic
FB	Facebook
Geo	geographic
HTML	Hyper Text Markup Language
HTTP	Hypertext Transfer Protocol
IETE	Institution of Electronics and Telecommunication Engineers
IP	Internet Protocol
IPV6	Internet Protocol version 6

JSON	JavaScript Object Notation
RST	Representational State Transfer
SNA	Social Network Analysis
SOAP	Simple Object Access Protocol
SQL	Structured Query Language
TCP	Transmission Control Protocol
XML	Extensible Markup Language

## التنبؤ بالمواقع الجغرافية اعتماداً على تحليل بيانات مواقع التواصل الاجتماعي

إعداد

رنيم نسيم أبوحسين

إشراف

أ.د. علاء حسين الحمامي

### الملخص

الجريمة هي مشكلة اجتماعية وقضية خطيرة تؤثر على المجتمع بأكمله. تؤثر على المجرم وعلى الضحية وعلى أسرهم. بالإضافة إلى ذلك، يؤدي ضعف القدرة على حل المشكلة إلى ظهور بعض السلوكيات الإجرامية. وقد تزايد معدل الجريمة بشكل كبير خلال العقد الماضي. وقد بذلت الحكومة محاولة للحد من الجريمة من خلال تحليل وتحديد السلوك الإجرامي المحتمل وهو أمر حيوي للوقاية والتدخل، وتحديد سمات الشخصية التي تميل إلى أن تؤدي إلى الانحراف.

التوجه للحد من الجريمة هو تشارك كل من المكان والزمان في التنبؤ بحدوث الجريمة، هذا من شأنه أن يساعد الحكومة على تأمين مكان الجريمة وحماية الأفراد من ان يكونوا ضحايا لجرائم مختلفة، ومن جانب آخر العمل

على فض أي نزاع قائم بدون وجود ضحايا بسبب تواجد العناصر الأمنية في الموقع المتوقع حدوث الجريمة فيه. نجاح هذا الحل هو مقياس للحد من عدد الجرائم الحاصلة الى أن تختفي.

لتحقيق هذا الهدف سوف نستخدم قاعدة بيانات وسائل الاعلام الاجتماعية المخزنة في السحابة الإلكترونية، ثم تطبيق خوارزمية خاصة لتحليل هذه البيانات الكبيرة والتتبؤ بموقع الجريمة. في النهاية، سيتم بناء هذا الحل لأن يكون حزمة تقدم خدمة متوفرة ضمن السحابة الإلكترونية.

# **PREDICTION OF GEOGRAPHICAL LOCATION BASED ON SOCIAL MEDIA BIG DATA ANALYTICS**

**Prepared By**

**Raneem Naseem Abu Hussein**

**Supervised By**

**Prof. Alaa Hussein Al-Hamami**

## **Abstract**

Crime is a social problem, and it is a serious issue that affects everyone in society. It affects the victims, perpetrators and their families. In addition, lack of problem-solving capability does account for some criminal behaviors. Crime has increased drastically within the last decade. The government has made an attempt to reduce crime by analyzing and identifying potential criminal behavior which is vital for prevention and intervention, as well as to identify personality traits that tend to lead to delinquency.

The new orientation for crime reduction is to predict the crime based on the location and time, which will help the government in securing the crime locations and protect people from potential crimes before they occur, while reducing the number of crimes, this being the measure of success for this solution

To achieve this goal, a social media database stored at cloud infrastructure is used, and then a special algorithm is applied to analyze this big data and to predict crime locations. in the end, this solution is constructed to be a package as a service plug in a cloud infrastructure.

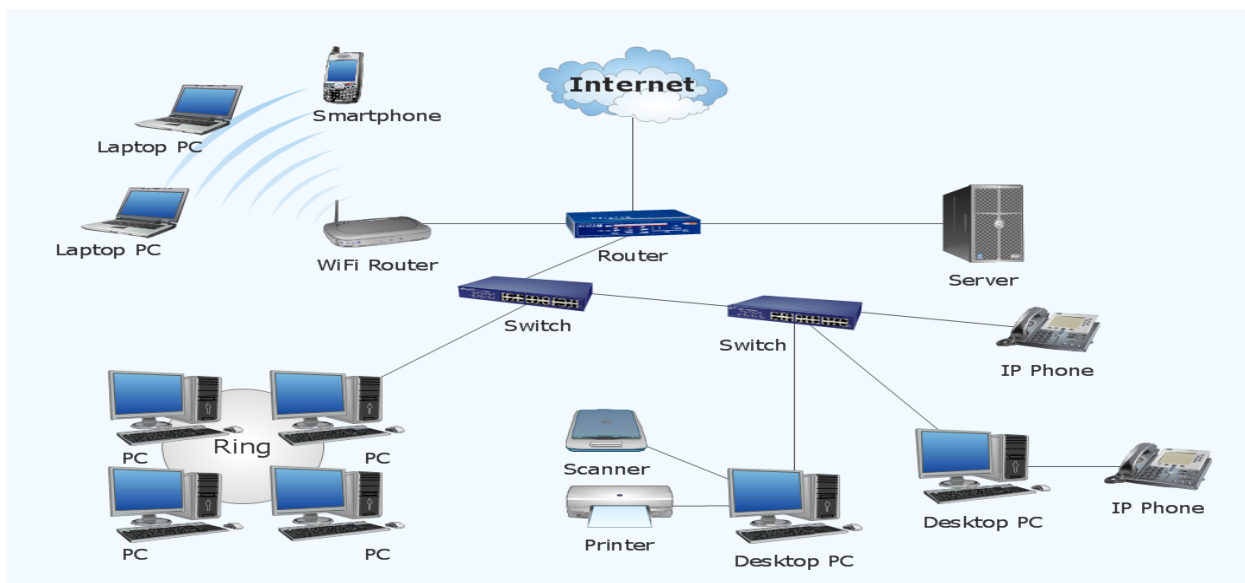
# Chapter One

## Introduction

### 1.1 Introduction

Communication is the activity of conveying information through the exchange of messages, information, and thoughts, by signals, visuals, writing or behavior. It is an essential part of the human life when it comes to interacting and developing our society. Today however we have taken communication to a new level, another form of written language: social media and social networking. This sensation affects us in numerous ways and nowadays we can access it at any times, because we are dealing with it all the time through our cell phones. Social media has created a new way to communicate with people from all around the world, so communications has been simplified.

People are linked together so that they can communicate with each other through a network that enables multiple users to share information, as in Figure 1 [Janssen, 2015; Odessa, 2015. Users interact with others by using dynamic web pages, IP phones, parallel Computing, and instant messaging (Tutorialspoint, 2015).



**Figure 1: Computer Network [Odessa, 2015]**

All previous features and multiple types of connection channels to link people evolve social networks.

Social networks is the basis of virtual communities. creating social conventions through software and technology, people can connect and interact with each other like in the real world, and this network allow each person to list his/her interest like sports, food, opinions as well as share in all things at virtual communities. all of this amount of data added a new concept to the technology world which is called big data.



Big data gives a lot of opportunities to companies, organizations, and the government to help the decision maker from different sides. This is because social networks have a profound impact on the economy because individuals take cues from one another rather than from a company, the press and the media.

Government can use this data to maintain security and raise awareness of different aspects.

## 1.2 Computer Communications

There are many devices and software that help in communication between computers. Some of these terms are the following:

**A. TCP/IP Protocol:** Defense Advanced Research Projects Agency (DARPA) has developed this protocol. This protocol was used in building ARPANET to connect the existing computers in this agency, universities, and other research centers. This network has been expanded and connected with many other networks that exist in several countries and form what is now called as Internet. TCP/IP protocol becomes general in providing services and for that, it is used in constructing the open systems.

It is a family that consists of several protocols and each one is allocated to deal with one limited application within the miscellaneous Internet applications. TCP/IP is not considered as a collection network because it works with any computer type and with all kinds of operating systems. In addition it has enough flexibility to work with any type of technology (e.g. it is possible to work with x.25, frame Relay and ISDN Techniques). This is one reason for making TCP/IP expand rapidly to become one of the more public for the network technology.

Servers and routers need TCP/IP protocol. Anyway, routers are not using protocols in all layers, Router does not use first layer protocol such as File Transfer protocol, because router does not execute such as these applications.

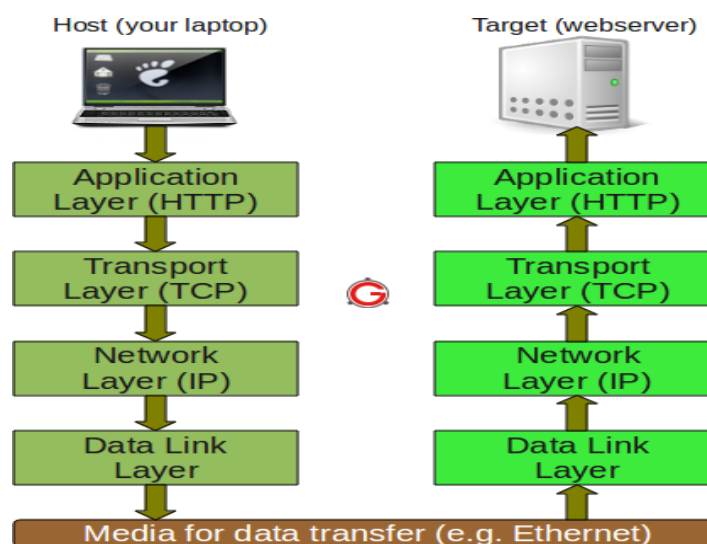
TCP/IP protocol consists of four layers and these are: Application Layer, Transport Layer, Internet Layer, and Network Layer. List the programs and devices that are deal with each layer are the following:

Application Layer: deals with the following software: SMTP, HTTP, FTP, TelNet, DNS, BootIP DHCP, SNMP, etc.

Transport Layer: deals with the Transport Control Protocol TCP and UDP.

Internet Layer: deals with Internet Protocol, RARP, ARP, ICMP.

Network Layer: Internet. Figure 2 shows the TCP/IP Architecture.



**Figure 2: TCP/IP Architecture (Arora, 2011)**

**B. Internet:** A network of networks, based on Contact and communication between individuals and devices (Investintech, 2015). Internet is an infrastructure of communication that include WWW, Video conferencing, ftp, Email, Instant messaging (Ricci, 2011). Internet does not represent communication role only, but it evolves to become the base of many different areas and it achieves positive results that is the reason behind finding the internet in all aspects life (Social, Commercial).

Today we find a number of organizations provide number of services to their clients through the internet because it has become one of the essential communication channel for most individuals

Social Networks Sites is the first step to move the internet from virtual world to real world in order to reflect the real world activities within a virtual world. In addition, Social Network Sites make the organizations have the ability to expand their scope, depending on the number of followers under the social networks sites especially Facebook and to motivate it to make process reengineering for all services to achieve more profit (Vermesan & Friess, 2013). Today, the internet has become the internet of things. The social networks sites have succeeded in representing real-world social aspect and are still developing, and providing many services. Additionally it holds a positive impact on the concept of the Internet of Things through all domains from cars, cameras, machines and any other product or machine the individual using it on social media that has its portion even in disparagement or praise (Wieber, 2015). Social networks sites users present to the company decision makers golden chances to test their products , compare them with other competitors products, measure and monitor customers satisfaction, detect the weaknesses and strengths, and helps identify exactly individual's needs. This information is the input to identify the organization position in the market and to help it to develop and execute its own strategy to compete and survive in a very competitive market.

However, at this time, individuals work on the Internet is targeting a particular category and a particular sector, but the results of the work and its benefits accrue to the entire world, that is, he/she is not touching the lives of individuals directly.

The advent of social networking sites was the beginning that made the Internet enters into a new revolution, and these sites were able to gain the confidence of individuals and succeeded in transferring the social life of individuals from the real world into the virtual world with full details.

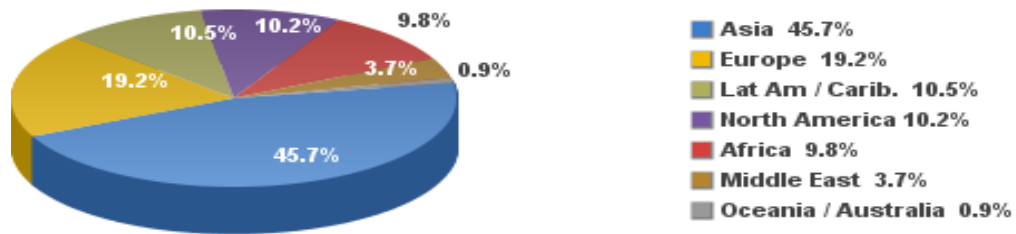
Social networking sites have created an atmosphere of interaction and communication between individuals. Internet of things creates an atmosphere of interaction between the individual and the things surrounding their homes, cars, hometown, and so forth.

To get to the so-called E-Life, we do the following.

E-Life = Social Network Sites+ Internet of things.

People around the world tend to use the internet in all life aspects ; using internet help to solve a lot of problems and the number of different activities done over the internet makes the life more easier . many statistics as in Figure 3 shows the increasing of using the internet

### Internet Users in the World Distribution by World Regions - 2014 Q2



Source: Internet World Stats - [www.internetworldstats.com/stats.htm](http://www.internetworldstats.com/stats.htm)  
Basis: 3,035,749,340 Internet users on June 30, 2014  
Copyright © 2014, Miniwatts Marketing Group

**Figure 3: Internet Users in the World (Internet World State, 2014)**

**C. IPv6:** IPv4 will not be able to cover the increased size of the Internet because there are more online users and more devices for each user and to connect these devices to the internet which is finally becomes internet of things. Because of the growing number of devices and users connecting to the internet IETF (IETF, 2015) will need to accommodate this growth and to preserve the connection on the internet and different services.

IPv6 can accommodate a number of users and devices with providing communication to the Internet by creating IP addresses with 128 bit; IPv4 creates IP addresses with 32 bit.

Using IPv6 will provide feature of enabled networks and services without any intervention from the user. IPv6 allow the sharing single IP address with multiple client (Apple, 2015).

Social Network Sites are the main reasons for the growth the number of users and devices, users not aware about IPv6 but it does one of the reasons they have to ability communicate within social networks sites.

"On social networks, users want to have multiple friends, partners, or relations with other users. Therefore, it can be expected that there is a heavily meshed network among these users. This will provide for good IPv6 connectivity because each user (IPoSN router) will be IPv6 connected to all his/her friends (IPoSN neighbor routers).

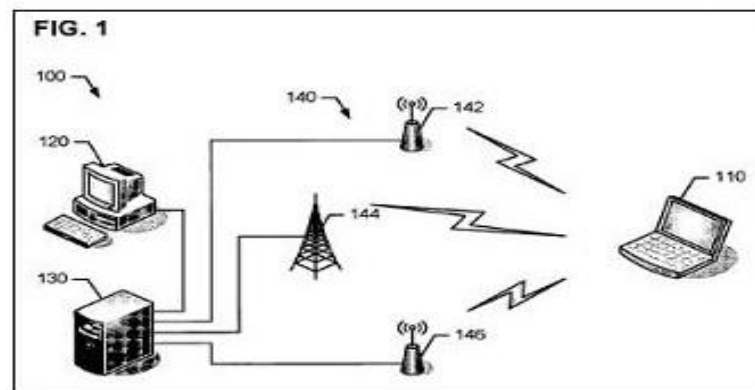
"Several Social Network Applications (SNAs) allow for plug-ins or for other applications to be meshed with the social network. Those applications can then generate IPv6 packets on the behalf of the users. Those packets can then be transferred hop by hop, or rather user by user, over the mashed SNA/IPv6, until they reach their destination (Vyncke, 2009)."

**D. Wireless Network:** Is one of the most used technique in data transfer technology trends with and growing expectation of ubiquitous connectivity. For example it connect with social networks sites, check email, voice conversation, web browsing, or myriad other activities. We now expect to be able to access these online services regardless of location, time, or circumstance (Grigorik, 2013).

Wireless networks are at the forefront of this trend, and it is a flexible data communications system, which uses radio frequency technology to transmit and receive data over the air, minimizing the need for wired connections (GC, 2013).

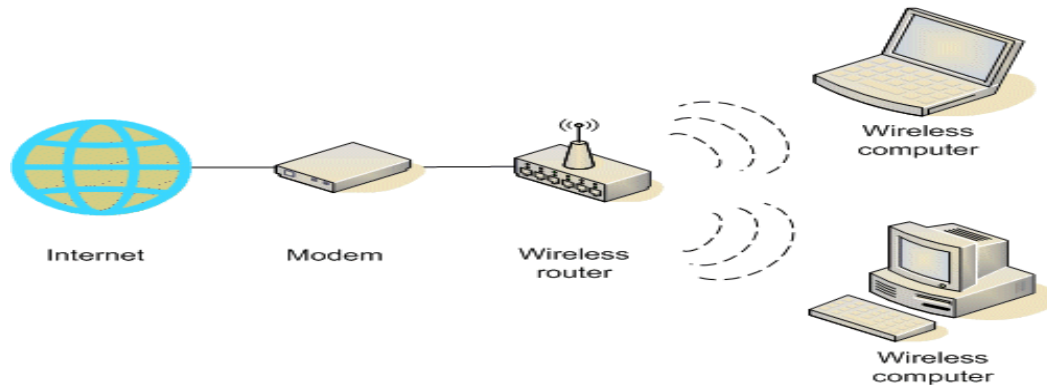
The spread of electronic services and individuals need it at various fields of life, so we should expect to find number of different wireless technologies to meet the needs, each with its specific characteristics based on its task. Today, we already have number of widespread wireless technologies: Wi-Fi, LTE, Bluetooth, WiMAX, ZigBee, HSPA, NFC, EV-DO, earlier 3G standards, satellite services, and more (Grigorik, 2013).

Wireless networking proved to be very useful in public places – libraries, guest houses, hotels, cafeterias, and schools are all places where one might find wireless access to the Internet, as in Figure 4Error! Reference source not found., and Figure 5Error! Reference source not found. (Tabona, 2004).





**Figure 4: Wireless Local Networks (Neo, 2015)**



**Figure 5: Wireless Communications (Enquiries, 2015)**

To achieve the greatest benefits from the wireless technology all manufactured devices are becoming keen to be able to deal with wireless technology, and become wireless networks. The first solution in any networks should be the flexibility and freedom they provide, and then the evolution trend of mobile devices to support the wireless technology, known as mobile computing.

**E.** Mobile computing: this refers to devices that allow people to access data and information to the network anytime, anywhere (Bucky, 2015).

Mobile has many users in different fields like commercial, social and political. Mobile computing has become the essential part of all individuals life, because its ensures permanent internet connection, that allows individuals of practice all social life within the social network sites. Also, mobile is able to have GPS technology to identify your location and to search for different location.

### 1.3 Social Networking

In this section, a brief description and its advantages will be given for Social networking.

**A. Social networking:** A social networking service is an internet-based service or site that allows building social relations among people and can expand number of connections with family, and friends who have same interests, activities and knowledge backgrounds, as well as serving business purposes, educational, economic and political. Examples of social networking include Facebook, LinkedIn, Twitter and Myspace.

Social Networking Site can be a good way to make connections and Interaction between people with the advent of social network and the rapid deployment and adoption of by it people; it leads to the question about the advantages of these networks.

**B. Advantages of social networking:** There are many advantages in the social Networking and can be listed as follows:

(1). For entrepreneurs Freelancers can find contacts via professional groups on LinkedIn and Twitter, while business owners can make use of the large user bases of Facebook and Twitter to market their products and services. Facebook

has a range of services designed to help businesses market themselves more effectively, because it reaches millions of people in a very short time and this deployment is free, and has the ability to target with advertisements at the precise demographic groups that are likely to respond favorably (Mikoluk, 2013).

(2). Contact and continuous communication between individuals is a great benefit of social networking sites because through these sites help individuals re-connect with old relationships and maintain them, as well as make new friends with individuals who have the same fields and Interests. Also, they help you to identify and gain new skills through members who share their interests and different opinions. In addition, to create an opportunity for you were not available to you before. They communicate and collaborate with members of different cultures and environments, such as students, staff, and university professors, while helping spread ones particular thought or idea reach to thousands of individuals [Weiyi and Zhen, 2013].

(3). Seeking a new job, Getting and giving product and service referrals, Receiving support from like-minded individuals , Making or receiving advice on career or personal issues (Cosmato, 2013).

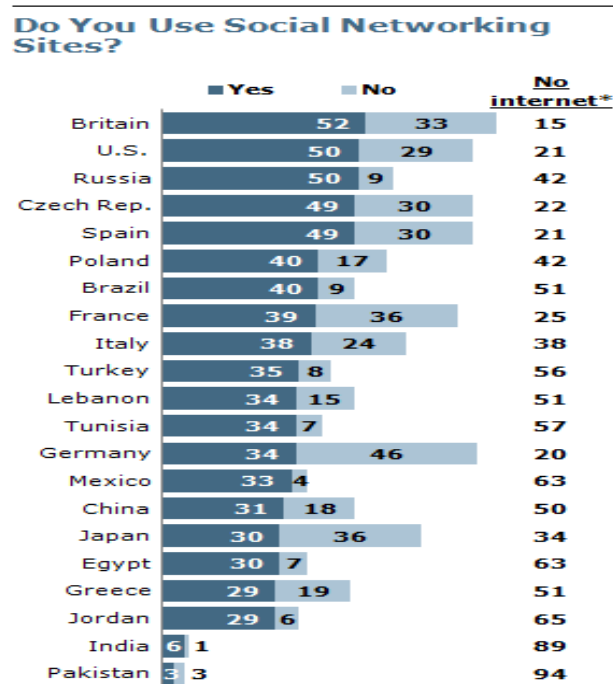
(4). Networking social sites have become a substitute for the weekly and daily in the family-scale phone calls because these sites allow individuals to exchange things and events of their daily lives with each other all the time and allowed them to participate in these events through a comment on the slate publication of articles or pictures. They cross the stage to be inform individuals of what happened, but they became remote in participation, also social networking sites include instant messaging feature that is, through chat exchange of information at the same time and they is a big advantage in multiple areas and including the field of education where can use this chat discussions between students or between students and the lecturer to facilitate the exchange of ideas and discussion through these sites people can participate in conferences long distance or even just send out invitations and organize things for the conference (Cosmato, 2013).

(5). According to professionals and academics, social networking sites allow users to create networks of like-minded people. Academics find other scholars to share research or ideas, or simply to talk to and "get their name out there". Professionals find mutual friends and possible clients, employers, or business partners. Social networking expands the horizons of what sort of contacts people can make in their professional lives (G.S, 2015).

(6). An additional area for politics such as politics social networking sites have become very important. This area is a resource for political news, information and finding likeminded issue-oriented people. In addition, it is a tool for voter outreach in the run-up to elections.

## 1.4 Social Networking Survey report

Here is a survey report about social networking spread around the world and how it used, this analysis from Pew Research Center, 2012:



**Figure 6: Use Social Network Sites (Wike, Horowitz, 2012)**

The percentage of social networking sites used around the world.

**Arab Publics Share Views about Politics, Community, Religion Online**

*Do you ever use social networking sites to share your views about...\*\**

	% Saying they use social networking sites*	% Saying they use social networking sites to share their views about...		
		Politics	Community issues	Religion
Lebanon	34	68	81	8
Tunisia	34	67	82	63
Egypt	30	63	74	63
Jordan	29	60	80	62
<i>21-nation median</i>	<i>34</i>	<i>34</i>	<i>46</i>	<i>14</i>

**Figure 7: Arab Publics Share (Wike & Horowitz, 2012)**

The percentage of Arabs who share views about politics, community, and religion

**Social Networking Usage**

*Do you ever use social networking sites to share your views about...\*\**

	% Saying they use social networking sites*	% Saying they use social networking sites to share their views about...				
		Music and movies	Community issues	Sports	Politics	Religion
U.S.	50	63	47	49	37	32
Britain	52	49	36	35	30	8
France	39	59	14	40	18	8
Germany	34	50	42	29	27	7
Spain	49	62	48	48	34	13
Italy	38	75	64	44	36	16
Greece	29	83	51	41	34	13
Poland	40	55	40	35	19	9
Czech Rep.	49	61	45	37	27	9
Russia	50	67	34	38	31	15
Turkey	35	78	63	61	57	53
Egypt	30	67	74	53	63	63
Jordan	29	68	80	68	60	62
Lebanon	34	48	81	22	68	8
Tunisia	34	77	82	64	67	63
Pakistan	3	--	--	--	--	--
China	31	86	50	50	35	10
India	6	85	41	82	45	40
Japan	30	59	32	33	22	1
Brazil	40	74	38	51	31	43
Mexico	33	84	40	42	21	15
<b>MEDIAN</b>	<b>34</b>	<b>67</b>	<b>46</b>	<b>43</b>	<b>34</b>	<b>14</b>

**Figure 8: How Use Social Network Site [Wike & Horowitz, 2012]**

The percentage of Social Networks Usage around the world and the percentage of sharing content around the world is very high, according to Figure 8Error! Reference source not found..

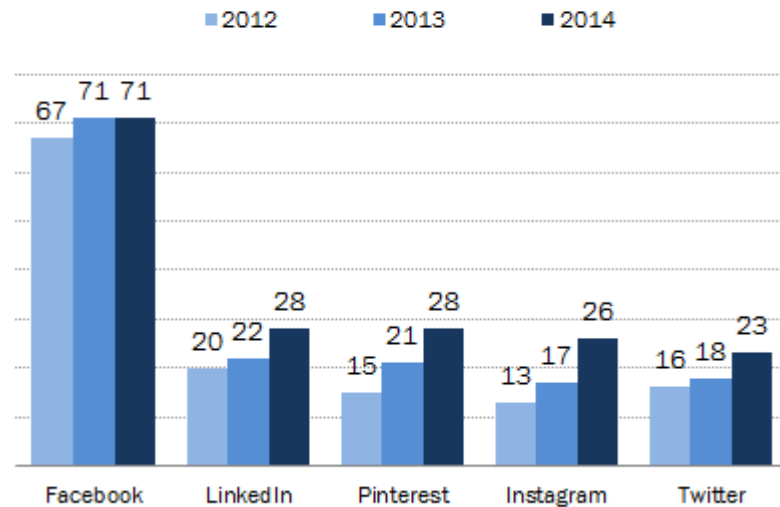
### **1.5 Information provided by social networking sites:**

Social network sites allow users to build a personal page that need some of information to become useful and enable all activities and features that allow people to “friend” others and share content with other users.

In a new survey conducted in September 2014 of social networks site, we find that Facebook remains by far the most popular social media site, as in Figure 9Error! Reference source not found..

### Social media sites, 2012-2014

% of online adults who use the following social media websites, by year



**Figure 9: Social Media Site [Duggan & et al, 2015]**

Compare between different percentages for different social media sites used by adults over a period of three years.

On Facebook, the largest social networks site, median user has 300 friends with a wider variety of friends, wider variety of information from share more personal information, and share their friends personal information, interest, and activities.



### **Type of data share with friends:**

Real name, a photo of themselves, Birthdates, School name, University name, company or organization of work name, city or town where they live, relationship status, email address, videos of themselves, personal cell phone number, and interests like (movies, music, books, quotes, political opinion) can be shared between friends.

Additional use of Facebook method is to invite individuals to conference, wedding, seminars, and meeting. All this data makes Facebook a treasure, because if we like to analyze this data we can find rich information in all life aspects. Large volume of data will open the door to political analysts and economists to raise different issues affecting the area of the work of each of them, and get satisfactory results that are impacted positively in various aspects.

### **1.6 Cloud Computing**

Social networks have become a reflection of real-world relationships; Social and practical through the virtual world. Facebook is one of the most important of these sites that contains the largest number of users. Facebook facilitates the communication between individuals and information exchange among themselves at different times with different images. Facebook provide many ways for interaction between users contribute to build new relationships and to expand the knowledge circle in this virtual world.

Facebook viewed the users as a data, since the registration at Facebook requires the definition of some information and documentation for each user to be able to get permission to enter and activate their personal account. Facebook always need data, because it is the basis of development, and this is the reason to stay Facebook the top of social network sites. Facebook innovated of how collect diversity data from users.

When we participate on the Facebook site, each steps through our account is data; activate our account, our chat, our interest, our opinion, our friends, and many others. That is not enough for Facebook, as more data is needed, and that is the reasons behind app, tests, and advertisements.

This data is wealth, but how can Facebook control this data?

Cloud computing is the savior for the Facebook success story. Cloud new business model delivered the reason for the solution of many problems and birth of new ideas. Cloud Computing is a blessing to share resources in return for payment of a bill of requirements used such as storage space, networks, and computer processor (Dialogic, 2010).

Cloud in the sky means to cover something, and that really is the basic aim behind the cloud. the Cloud collects number of people, organization, hardware, internet and collaboration between it.

First to achieve this collaboration and adopt it, and cloud should maintain availability to any object at each time and for each place.

Second, facilitate each of single user and organization with multiple views:

A: Infrastructure of hardware and software, Servers and storage.

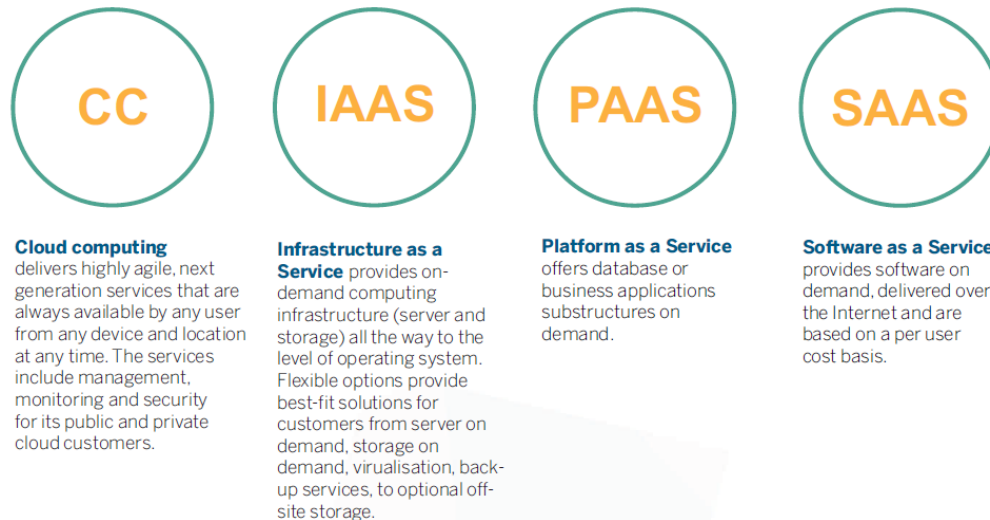
B: Allow of resource sharing

C: Upgrade of software

D: On-demand services

All this features have a common properties is scalability, reliability and efficiency (Huth & Cebula, 2011). To give benefit for each organization to have the latest technology, reduce cost, high level of services and availability at all time (Chard & Caton, 2010).

Another potential benefit is that personal information may be better protected in the cloud. Specifically, cloud computing may improve efforts to build privacy protection into technology from the start and the use of better security mechanisms ensuring that the information is safe and allow only people who need to access to it (Borgmann & Hahn, 2012).



In terms of social networking, using cloud-based communications provides click-to-call capabilities from social networking sites. Access to Instant Messaging systems and video communications will broaden the interlinking of people within the social circle (Dialogic, 2010).

Social networking sites can be expanded for more than just communication between friends, to become a social cloud, containing information and resources. This expansion will provide an incentive for the exchange of information and the construction of several integrated applications.

Infrastructure of the cloud in the area of social networks is a way to manage and control the user by documenting his/her knowledge, and experience, as well as being an important source of help in reducing crime and maintaining security.

This method helps to provide the environment with many requirements and

projects, bringing many efforts to take advantage of social concepts through these networks and extract information in many different areas by building applications check which maintain specific check questions and clear goal.

## 1.7 Big Data

The use of relational databases was the basis on which it adopted the organizations and individuals in solutions and new ideas. Structured database is a rule based on the relations between their information through specific primary key links between the tables.

The rapid development of technology and the adoption of users by creating great opportunities with the biggest challenges was the cloud computing. It is one of the distinctive solutions that left a clear impact in solving many issues while in addition creating many new ideas.

The cloud storage provides high available storage space consistently for many organizations and individuals. The size of the stored data increases with increasing storage space. The relational database is not the only data storage; today organizations have different types of data format (structured and unstructured) to be stored in the database. Combining structured data and unstructured data gives a new perspective to the data so-called big data.

The question that arises is: is big data with unstructured data achieving any commercial value?

Social networking sites are producing large amounts of unstructured data to reach the size of terabytes or even peta bytes. Social media interactions offer an in-depth view of customer's manner, priorities, and their purchasing moods... etc. as well as all sorts of information that aid business understand their customers preferences. In fact, customer care moves to a whole new grade by allowing business to address customer questions on social media, sometimes in real time. It supports businesses helping their customers more preferences, but also helps businesses put up a brand community concerning their social interactions. It is paving the way for preferable collaboration.

We instruct that social media is a majestic channel for customer interactions. It also offers businesses great insight into the mind of its customers. Cloud solutions like unified communications and video conferencing offer leading workspace collaboration. Social CRM is one tool that has grown from the popular model of social media. Social tools, like Social CRM, have emerged as the enterprise-grade solution for embracing a superior communication and collaboration within the framework of an organization. Mobile, Data and Cloud make every one of these solutions infinitely more accessible all while helping

businesses comprehend how they are utilized in order to measure and predict outcomes [Das and Kumar, 2013].

Big data demands additional new skills. There are a host of developed technologies and new platforms to understand how to be effective with these numbers of data, and the IT branches [Dion, 2010].

This thesis concerned of social big data analysis with different category to achieve security and reduce crime to take preventive measures before they occurred.

## 1.8 The Statement of the Problem

Facebook is a big data database host by cloud infrastructure, and Facebook has different data channels interactions with Facebook users; these interactions come from the users trust of Facebook sites and have moved them from real world to the virtual world. Facebook become a user life where people can do all life activities through it, that's has led to making Facebook probably a reason for crimes, by when we found different opinions between the users, with the readiness to make different events in different areas, maybe these can be considered the introduction of criminal activity, and benefit from the Facebook big data to predict crime locations.

We suggest this idea because through Facebook we can know what the users think about something and what they want to do before doing anything, even thought through Facebook are public for all to see.



## **1.9 Contribution**

In recent years, social media has become the users life, and it is the first platform provide the treasure to government, organization, and researchers by data that social media can collect because this time data is the language time and the capital of all success sector.

For that, we need to protect this treasure and increase the benefits from this data by stopping any criminal activity that's threatens an individual safety.

## **1.10 Thesis Organization**

In this thesis, the introduction discusses Social Media Sites, Big data and the Cloud, literature reviews are discussed in chapter two. The proposed system design for predicting crime locations based on the analysis of social media data is presented in chapter three. Chapter four contains the implementation of the proposed prediction system. The result of the implemented system is stated in chapter four. The conclusion and future works is detailed in chapter five.

# Chapter Two

## Literature Review

### 1.11 Introduction

Online, previously meant the users are connected to the internet. Today the internet connection is available from any location. Online means that data interaction such as photo, video, status and sharing any content with others within social media networks such as Facebook. Facebook is one of the most used sites in social media, which has number of users, and this number is growing gradually. Facebook based on connecting people to these they know and connect to friends who share the same interests.

Facebook succeeded at features friends list and information that sharing between them, and all this data is stored at big data so this data is increasing and Facebook features are increased like geo-social.

## 1.12 Facebook Properties

Geo-social (Wiki, 2015) in this type of networking a geographic services and capabilities that is used to enable additional social dynamics. Location data or geolocation techniques can be submitted to allow social networks to connect and coordinate users with local people. Geolocation can be IP-based or use hotspot trilateration. Texted location information tracking, for mobile social networks, can enable location-based services to enrich social networking.

When Facebook users use geo-tagging the real world is reflected within virtual world, because it can help in the transfer of the real user's movements patterns in the virtual world The first control of the user's life and different type of activity and in data interaction is life protection.

This is the main factor we should focus on and develop to preserve people life style.

Facebook information treasure is using this data by marketing and decision-making, e-commerce, education and much more in additional to a number of studies depend on Facebook data and Facebook users opinion to achieve number of goals, all previous result have succeed and adopted social media in the life style so it is the backbone for all sector in our lives.

### 1.13 Facebook and Crime

Facebook is one of the biggest social media websites and through the using Facebook and the using of its features of Facebook takes an important side in all areas of our lives. It takes the first rank in the fields of entertainment, sports, social and health. So issues Facebook is the main source of information through the presentation of information and interaction with them, whether this information was personal and family related or outside the family and friends to be in the community with other issues. Today we look to Facebook as a mean used by individuals for dating and entertainment without any benefit, and the people share their details precisely through Facebook, because it controls a large area of their time and thinking so our role is to make the interaction of individuals with Facebook as the basis for achieving the security aspect, which is the most important aspect to be focused on and supported by large way through Facebook, this information will teach us how to realize security and safety, how to be closer to the control of the crime and how it will contribute to raise the level of services that is provided to individuals to see the results of satisfaction and safety reach to The highest level.

Crime involves killing, drugs, stealing, or any other activity. When the crime world uses it this means criminal activity and is all the criteria reason for crime occurrence.

When, why, how, are three questions to ask for each crime that happens, and we can use archives crime data, evidence from crime location and forensic evidence. All these features will help to answer the questions and add to the results to crimes archives at police department. However, can we control the crime and expect it before it occurs?

The answer is yes, and that is what we want to achieve from this thesis by adopting Facebook data to predict the crime location, which will lead to preventing a number of criminal activities.

## 1.14 Literature Reviews

A Number of researchers benefited from local individual information, individual data and opinion from social media sites, and different online tools to analyze crime, one concerned to predict criminal behavior, another one predict criminal activity and crime is suggesting the location.

**Sujatha and Ezhilmaran** (Sujatha & Ezhilmaran, 2014): This research provides a new method for crime location prediction and protection by using ABC and ARM algorithms based on UCI data. This method depends on socio-economic features. These features are extracting from the database to get a new attribute database. Using AMABC algorithms to cluster this database, based on the result of this algorithm, using ARM algorithm to mine the high crime location. Mined patterns get the high crime location volume, and on the future which is high volume location.

This result depends on specific database size [1994,127] and all this data related to the crime. This method cannot predict any new type of crime like when and where it happened.

**Sukanya. M, T. Kalaikumar and Dr. S. Karthik** (Sukanya & et al, 2012):

The authors explain that criminal activities have increased tremendously. So security has become a major issue to people. The police should take control actions department. By analyzing the criminals' data, clustering and classification techniques are used. This data will be stored in the criminals' database. Spatial clustering algorithm and structured crime classification are used to classify the criminal activities. More number of criminal activities happened in the location. This location determined the hotspot of this type of crime. Similar types of criminal activities grouped together by using GIS of this information to determine the hotspot. Determining hotspot depends on three types of crime classification:

1. Classification of crime place
2. Classification of crime types
3. Classification of crime time

This paper presents the method to identify the hotspot of crime. Based on the type of crime the police department can easily identify the hotspot of the burglary crime.

This paper determines crime location and each location of crime activity type based on analysis of police department data. By this, it can achieve the protection from specific type of crime that has happened before at this location. This method cannot determine any new criminal activity and any new location.

**Yu Zheng** (Zheng, 2012): location-acquisition empower people to use location data with existing online social networks in a variety of ways. People can upload Location-tagged photos to Flickr, comment on an event at the exact place where the event is happening in Twitter, share their present locations on Foursquare, record travel routes with GPS trajectories to share travel experiences in Gelfie. The dimension of location helps bridge the gap between the physical world and online social networking services. A Location-Based Social Network (LBSN) does not only mean adding a location to an existing social network. Therefore, people in the social structure can share location-embedded information. The new social structure is made up of individuals connected by the interdependency derived from their locations in the physical world as well as their location-tagged media content such as photos, video, and texts. Perspective of users and locations analyzes the unique features of LBSNs beyond traditional social networks.



Users create their own location-related stories in the physical world and browse other people's information as well. An online social site becomes a platform for facilitating the sharing and understanding of people's experiences. People in an existing social network can expand their social structure with the new interdependency derived from their locations. As location is one of the most important components of user context, extensive knowledge about an individual's interests, behaviors, and relationships with others can be learned from his/her locations. For instance, people who enjoy the same location can connect with each other. Individuals constantly hiking the same mountain can be put into contact with each other to share their travel experiences. All of this will help to estimate and know information about users from locations they visit and information they share, then understand the location with different views, one of them is the number of users mobility to this location and users opinion.

**Chung-Hsien, Max W** (Chung & Max, 2012): Database gathering from police department, and from this database prepared approach to dataset designing, leads this field is crime records and crime related event, location and time embedded with this data. Data classified depend on spatial and temporal features to forecast the hotspots of specific type of crime which effect of any location.

This paper divides each city into a number of cells with same space, each cell aggregate categories of different types of crimes.

First, define some of attribute related the crime, location, event, and other factors. Then start of binary classification to make relation between input data and predefined class label of data, the input data classification with the predefined category label is counted for each category on the maps and set it as hotspots label, and if the count for this category is increased it begins heating up.

This paper depends on specific database and attribute that can predict specific type of crime before it happens.

**Mande, Srinivas, and Murthy** (Mande & et al, 2012): This paper identifies the criminal depending on some of data in the absence of clue or any witness by the forensic experts. Using data mining technique to identify the criminal mapping based on the database and variable of each new crime, helps to identify the criminal using the Auto correlation. The way in which the incident has taken location, time, modus operandi, criminal attribute and criminal psychological is store in a database. All these data need analysis to determine the criminal, this is the big benefit when there is no witness and clues.

This method has a database from the police department containing specific data and criteria analysis that depends on it.

**Das and Kumar** (Das & Kumar, 2013): This paper is builds on approach to achieve the analysis of big data depending on the twitter. At the beginning, the authors establish the connection to follow public tweets from the twitter, then using Hadoop platform to store the data by Hbase, which depend on it to build interfaces that make interaction with these tweets. This paper proposes approach to analysis big data but not using any algorithms on this data to extract any information.

**Nan and Chen** (Nan & Chen, 2011): This paper is proposes a method of how can benefit from social media data. The idea is to keep track of users who share their location in the real world and each user can see where their friends and the content tagged with this location, all this data collect by GPS with mobiles and LSN with laptops.

This paper has a dataset from Bright kite provides a database contain user ID, activity type like text, photo, and time of it activity. By the database, we define the mobility patterns to evaluate the location based information sharing.

The patterns depend on calculating each users movement path into clusters, analyzes it to define miles clustering to obtain three type of user behavior result; home users, home-work users, home-vacation users, and other users.

**Gerber** (Gerber, 2014): this method uses tweets from twitter to predict crime activity. By collecting the crime type and frequency for each type from police department, then determine study city and collects tweets tagged with GPS coordinates falling within the city limits and with the same period of police department data and tweets data.

Analysis of this data by calculate the probability occurrence of each crime type depends on some features defined before, users tweets is the source derived this features, while in addition the density of crime type from police department, these two value estimate the threat point.

Deep analysis of tweets data and application of numbers of data modeling gives more accurate features for crime type and to predict crime point.

**Weiyi S, Zhen M, and Jiangy** (Weiyi & et al, 2013): this paper proposes a lightweight approach for uncovering differences between pseudo and large-scale cloud deployments. The proposed approach makes use of the readily-available yet rarely used execution logs from these platforms. The approach abstracts the execution logs, recovers the execution sequences, and compares

the sequences between the pseudo and cloud deployments. Through a case study on three representative Hadoop-based BDA Apps, show that proposed approach can rapidly direct the attention of BDA App developers to the major differences between the two deployments. Knowledge of such differences is essential in verifying BDA Apps when analyzing big data in the cloud. Using injected deployment faults show that approach not only significantly reduces the deployment verification effort, but also provides very few false positives when identifying deployment failures.

**Groff, Elizabeth R. Nancy G. and La Vigne** (Groff & et al, 2003): this paper gives the true promise of crime mapping lies in its ability to identify early warning signs across time and space, as well as to inform a proactive approach to police problem solving and crime prevention. Recently, attempts to develop predictive models of crime have increased, and while many of these efforts are still in the early stages, enough new knowledge has been built to merit a review of the range of methods employed to date.

**Uttam Mande, Y. Srinivas, and J.V.R. Murthy** (Uttam & et al, 2012), The latest technological developments contributed significantly towards modernization, while at the same time increasing the concern about the security issues. These technologies have hindered the effective analysis about criminals. Application of data mining concepts have improved, too. They Yield better results in this direction. In this paper, binary clustering and classification techniques have been used to analyze the criminal data. .This paper aims to potentially identify a criminal based on the witness/clue at the crime spot an auto correlation model is further used to ratify the criminal.

**Tony H. Grubestic** (Tony, 2014): One of the more popular approaches for the detection of crime hot spots is cluster analysis. Implemented in a wide variety of software packages, including Crime Stat, SPSS, SAS, and SPLUS, cluster analysis can be an effective method for determining areas exhibiting elevated concentrations of crime. However, it remains a particularly challenging task to detect hot spots using clustering techniques because of the uncertainty associated with the appropriate number of clusters to generate as well as establishing the significance of individual clusters identified. This paper highlights the potential utility of several diagnostics for resolving such issues.

**Google Analytics** (Analytics Guide, 2014): it is a web analytics service offered by Google that tracks and reports website traffic.

Users create and configure their account and connect your site to Analytics to begin online review campaigns by tracking landing page quality and conversions.

Goals might include sales, lead generation, viewing a specific page, or downloading a particular file, additional to tracking visitors, create goals so you can see how your visitors are converting into subscriptions, donations, and revenue, while enabling ecommerce tracking for those who store within their site.

Google Analytics' approach is to show high-level, dashboard-type data for the casual user and more in-depth data further into the report set. Google Analytics analysis can identify poorly performing pages with techniques such as funnel visualization, where visitors came from (referrers), how long they stayed and their geographical position. It also provides more advanced feature including custom visitor segmentation.

Google Analytics e-commerce reporting can track sales activity and performance. The e-commerce reports show a site's transactions, revenue, and many other commerce-related metrics.

Google Analytics Cohort analysis feature helps understand the behavior of component groups of users apart from their user population. It is very much beneficial to marketers and analysts for successful implementation of Marketing Strategy.

**Foursquare** (Foursquare, 2015): location-based social network where you “check in” to your favorite locations and let your friends know where you are. Foursquare has deeper purposes than just location sharing with friends, though. Many restaurants and stores will reward loyalty and frequent check-ins with specials and deals.

To use Foursquare, you must have a GPS or Internet enabled phone. Once you are in a location - for example, library or restaurant you can “check in” to Foursquare and let your friends know where you are.

Every time you check into a location, you receive points. If you are checking into a place for the first time, you earn more points. If you check in with a friend, you earn even more points. The points don’t actually amount to anything except a friendly competition among your friends and the bragging rights of being on top. Foursquare useful from these point to analysis these locations and answer to any question asked for or about any location.



Your foursquare check-ins is completely private and only will be shared with people that you approve. And, as an extra measure, each time you check-in to a location you are given the option to share the check-in with friends on Foursquare. Those who do want to share also have the option of broadcasting the check-in on Facebook and Twitter.

**Brightkite** (Brightkite, 2015): Brightkite is a location-based social domain network which allows users to 'check in' by using text messages or the applications of mobile the users were able to see who is close and who has been there before. Brightkite authorized registered users to connect with their present friends and also meet new people according to the places where they go. Users "checked in" at a place, they can permit to add post notes and photos to a location and other users can put their comments on these posts .The Brightkite is considered a visualization tool which is used by the users to show real-time updates from a place, user or keyword. Users also have the choice to show posts from Twitter containing a specific hash tag. The combination with Facebook is a great strategy for Brightkite What it means at an essential level is that Brightkite users will be able to share their location with their friends and publish the information as a status message or news feed on Facebook...

**Shaomei and Adamic** (Shaomei & Adamic, 2015): This paper presents the empirical study of how visually impaired people use online social networks, specifically Facebook. We identify a sample of visually impaired users, and study the activities they perform, the content they produce, and the friendship networks they build on Facebook. We find that visually impaired users participate on Facebook (e.g. status updates, comments, likes) as much as the general population, and receive more feedback (i.e., comments and likes) on average on their content. By analyzing the content produced by visually impaired users, we find that they share their experiences and issues relates to vision impairment. We also identify distinctive patterns in their language and technology use. We also show that, compared to other users, visually impaired users have smaller social networks, but such differences have decreased over time.

Our findings have implications for improving the utility and usability of online social networks for visually impaired users.

**Chen, Kaafar, and Boreli** (Chen & et al, 2013) The extended availability of positioning technologies such as GPS in smartphones and other smartphones tools has enhanced the use of real-time location updates in the apps of smartphones and location based services. Location-Based Social Networks

(LBSNs) like Foursquare, Facebook Place, supply a platform for updating one's location that is viewable by friends, by checking into a set of venues in the geographical proximity of the user's present location. The check-in- and content sharing-based LBSN have been studied by extended way with a focus on various topics, including users' behavior. This paper, investigates a new LBSD mobile network by studying the dataset collected from an increasingly public social discovery application, "Momo"<sup>1</sup>, Momo supplies two OSN related functions: social discovery, that enables a user to discover surrounding people based on the geographical distance between them, and instant messaging that allows users to (subsequently) communicate. By default, the application updates the user's location to the server (unless users explicitly opt-out of status updates); hence, a rich set of spatio-temporal information about the users is captured in our dataset.

**Shaomei and Adamic** (Shaomei & Adamic, 2014) this paper presents the first large-scale empirical study of how visually impaired people use online social media, more specifically Facebook. We have a piece of 50K visually impaired users, and we study the activities which they perform the content they make and the friendship networks they build on Facebook. We find that visually impaired users participate on Facebook (status updates, comments, likes) as much as the

common population, and receive more feedback (comments and likes) on average on their content. By analyzing the content made by visually impaired users, we find that they share their experience and cases that related to vision impairment. The authors also identify characteristic patterns in their language and technology use. Authors also show that, compared to other users, visually impaired users have smaller social networks but such variances have reduced through time. The findings have implications for improving the utility and usability of online social media for visually impaired users.

**Irena and Florian** (Irena & Florian, 2014) the general nature of user created content on social media platforms offers the possibility for trend controlling as an insight into the topics that attract the attention of a large portion of users. When the scholars and practitioners have previously knew Google and Twitter as a valuable source of trend information, then there are no actual implementations or research efforts in the field of trend detection over Facebook public posts. In this paper, authors give two contributions towards trends controlling through the public posts of Facebook. In addition, the suggestion to evaluate a system for trend detection based on the Features of the posts, which shared on Facebook. Based on our results we suggest three categories of trending topics: 'daily routines', 'general topics' and 'disruptive events'.

Analyze and compare the features of the proposed categories in terms of distribution and information diffusion in order to increase the comprehension of emerging trends on Facebook. In the end, draw conclusions can perform our findings in terms of challenges and opportunities for future work in this direction. This paper present two contributions towards trend monitoring over Facebook public posts. Propose and evaluate a system for trend monitoring based on the Features of the posts shared on Facebook. Based on our results we distinguish between three categories of trending topics: (1) ‘disruptive events’, (2) ‘popular topics’ and (3) ‘daily routines’. We analyze and compare the features of the proposed categories in terms of distribution and information diffusion in order to increase the understanding on emerging trends on Facebook.

# Chapter Three

## The Proposed System Design

### 1.15 Introduction

Security and safety are the most important needs of any individual in this life so the government has to ensure those needs are met for the citizens in every way and by all means available.

The world today is the world of technology and must be based on the information in all fields. The information specified is no longer limited in the particular field and purpose of a specific target is the required information. We are now seeking to gather information in various forms for many fields, collect the structured and unstructured information that has built many applications and large data that provided the ability to store this data and work on them.

Facebook is an ideal source for data used in decision support and thus it is possible to discuss through Facebook posting of the events and emotions openly. In addition, we share emotions and events in its time, which mean in its real-time. Facebook can be used to organize many events and meetings and this helps to know events before the full details of where and when they occur as well as the participants of it.

This thesis provides a new method to take advantage of information and social networking sites in the achievement of security and safety while controlling crimes and to be a starting line for many developments.

The prediction of criminal activity and controlling it must be through the control of the criminal sites and criminal activities. This will help police to rearrange and distributing forces in the places to commensurate with the results of the analysis to reduce the incidence of crimes.

Crime Prediction Application is a set of application to collect, normalize and analyze Facebook data, the first application is a web application designed to collect Facebook posts and store this data in the project Database, the second application developed to normalize these posts and get them ready for analysis, then the normalized data is analyzed based on the certain criteria to find out suspected/outlier posts and then compare it with the actual cybercrime database to conclude the feasibility of this project.

## 1.16 Research Overview

This research is motivated to add new study of the most critical problem; that is the crime; we should define a new method for predicting the crime by the following:

1. Determine the industry of research based on the nature of city areas and popular locations, in this project we concentrate on the restaurants.
2. Determine the areas/locations (city-blocks) of interest, we took the most famous areas that considered to be major centers for restaurants in Amman (Abdoun, Abdallah Ghosheh Street, AlMadina Almonawara Steet. Mecca Street, Swaifieh and Shemeisani)
3. Determine the targeted restaurants in these locations, we covered the most important and popular restaurants (KFC, McDonalds,...etc)
4. Develop a web based application to collect the related posts from Facebook using Facebook APIs and by creating pages and using Facebook Like option to view and collect restaurants pages comments and reviews.

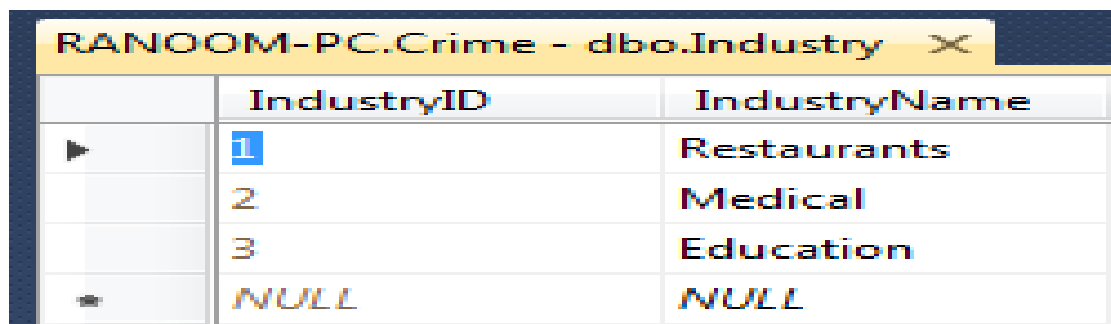


5. Build the industry dictionary which includes (the main keywords and its attributes) and build the common used key words that will be considered during the post analysis (adverbs and negative words)
6. Develop the normalization .Net class which execute a cleansing process to take out undesired letters and make it consistent with the dictionary keywords (remove "ال", replace "ة" with "ه", remove repeated characters .. etc.)
7. Develop the Classifier .Net class that will get the normalized posts, split them based on the industry dictionary and compute a rate for each post.
8. The results then reformatted and analyzed while comparing them with Facebook restaurant rate to check out any outliers.
9. Compare the outliers with the actual Cyber Crime and CID Database to prove the feasibility of the project.

## 1.17 Industry of Research

The industry that will be covered in this thesis is the restaurants, one of the major businesses in Amman which is almost crowded by customers and restaurant management mostly depends on the internet and other media to promote their products.

The application design provide the capability to define many industries, the industry key is considered in the detailed dictionary key words and attributes to have more accurate result, and to make it ready to process other industries posts.



The screenshot shows a window titled "RANOOM-PC.Crime - dbo.Industry" with a close button. It displays a table with two columns: "IndustryID" and "IndustryName". The table contains four rows: a row with "1" and "Restaurants", a row with "2" and "Medical", a row with "3" and "Education", and a row with "NULL" and "NULL".

	IndustryID	IndustryName
▶	1	Restaurants
	2	Medical
	3	Education
⊞	NULL	NULL

**Figure 10: Industry lookup**

## 1.18 Areas of Study

The thesis will study the most popular restaurant clusters in Amman where many restaurants have branches in these locations; the posts collection process will get also the location of the restaurant to be used when analyzing the posts, the targeted locations are listed in Figure 11.

RANOOM-PC.Crime - dbo.Locations		
	LocationID	LocationName
▶	1	Tla'a Al-ali, Al-Gardenz, Al-Madina Almonawara
	2	Al-Madina AlRyadiyah, Al-Shmeisani
	3	Abo-Nsair, Al-Jbaiha
	4	Abdoun,Dair Ghbar, Dahiet Al-Yasmeen
	5	Mecca St, Khalda, Al- Rabieh, Um-thainah, Abdallah Ghosheh

**Figure 11: Locations lookup**

## 1.19 Restaurants of Study

Most well-known restaurants are considered in this thesis, especially those that Facebook rated them based on customers rates, the thesis will study the restaurants listed in Figure 12.

RANOOM-PC.Crime - dbo.Entity ×				
	EntityID	EntityName	IndustryId	FB_Rate
▶	1	KFC	1	4
	2	McDonalds	1	3
	3	Zooka	1	4
	4	Lebnani Snack	1	4
	5	SubWay	1	4
	6	Deir El Qamar	1	4.2
	7	Chili House	1	4
	8	Salad House	1	2.5
	10	The Cake shop	1	4.5
	11	Pizza Hut	1	3
*	NULL	NULL	NULL	NULL

**Figure 12: Entity lookup (restaurants)**

## **1.20 Facebook Data Collection Application**

### **1.20.1 Facebook Application Programming Interface (API)**

The target data source is Facebook; we are able to have the Public Posts by using Facebook API that can make representational State Transfer (REST) calls to Facebook, Facebook Query Language (FQL) which is SQL-styled query language and Facebook JavaScript language. By releasing this platform, Facebook built an apparatus that allows developers to create external applications and to empower Facebook users to interact with one another in new and exciting ways—ways that you, as a developer, get to invent. Facebook has also opened up its platform to Internet-connected desktop applications with its Java client library (Graham, 2008).

To get customers posts, we built a Facebook Competition application under Facebook platform to gain access to users public data, the purpose of this application is to motivate users to login by Facebook username and password, it is a simple competition application that asks users some simple questions to enter the competition.

We then used customer user ID to get their comments. Additionally, we can extract the public posts and comments in general. All these information used to study user behavior and analyze user opinion for a certain study area.

### **1.20.2 Data Collection Algorithm**

These steps show the collection process of social media data. Facebook is the data source, gathering the information from it using the developed Facebook app.

1. Login at Facebook account.
2. Request build Facebook app.
3. Fill required information.
4. IF valid then start app.
5. Else retype required information, go to step 4.
6. Request data from API.
7. Retrieve JSON file.
8. Read JSON files from .NET.
9. Store at Database.

### **1.20.3 Facebook Page Sharing**

As Facebook has restricted the access by API to a certain number of posts and comments length, we have created a page to communicate/share with some Facebook pages in different locations and places using Facebook Like option, in this way we can get data about review, post, number of likes, and number of visit for these pages as they are developed and administered by the restaurants

to interact with customers and collect their feedback, opinions and needs. Based on these information, the restaurants management make the required decisions to enhance their products and services and keep the business running and competitive.

We also include the thesis many of statistics on crimes incidence of crime in Jordan from Criminal Investigation Department (CID, 2015) and from Cyber-crime unit in Jordan. The reason for adding these statistics is to support or to refute the results of this thesis study areas.

#### **1.20.4 Competition Application Design**

We build the Facebook application by using .Net environment, SQL server database, and establish the connection with Facebook API. This application uses number of questions where each question has number of possible answers and one of these answers is correct.

The ability of reaching the data is from the API. This app takes access token that is achieved when request is used to build new application from Facebook and through Facebook login. Users can access this app and Facebook login will be available through app developer Facebook page.

Facebook provides multiple client libraries for different programming language to give the ability to access information.

### Competition Application Steps:

- 1) Prepare web application under .net environment with Facebook library to achieve connection with Facebook API.
- 2) Build database schema using SQL server to achieve application target goal and to store JASON and XML files by retrieving from API.
- 3) Build user interface for application, competition app need design for view questions and possible answers as well as design for the control panel that shows the data which retrieve from API.
- 4) Programming this app for users so they can access it through Facebook login.
- 5) Login authentication conveys a number of data related to it like ID, Name and public personal information.
- 6) From this ID it is possible to access the user's information, with multiple layers, with code side define what data want to retrieve from API, read it from JASON and XML, and store this data at app database.

Another method for get information by the following:

- 1) Create account admin at Facebook.
- 2) Collect different locations and pages under this account by like option.



- 3) By ID for each page can access to all information, need at store at database like post, comment, post of pages, reviews, number of like, number of visit.

#### **1.20.4.1 Application Process**

This process shows applications configuration, at Facebook level and .Net environment level, to gain accessibility to Facebook data by API.

##### **1. Facebook, Create application level Process.**

###### **A. Facebook, Competition app.**

- 1) Open Facebook website.
- 2) If you have FB account then enter email and password then login
- 3) If not, Sign in by entering required information, after successfully creating an account, enter email and password then login to Facebook.
- 4) Request to build new Facebook app, and then name this app and get access token ID.

###### **B. Facebook, .Net environment Process.**

- 1) From (.net) it is possible to create new web app.
- 2) Included API Facebook libraries need.
- 3) Build design for app user interface.

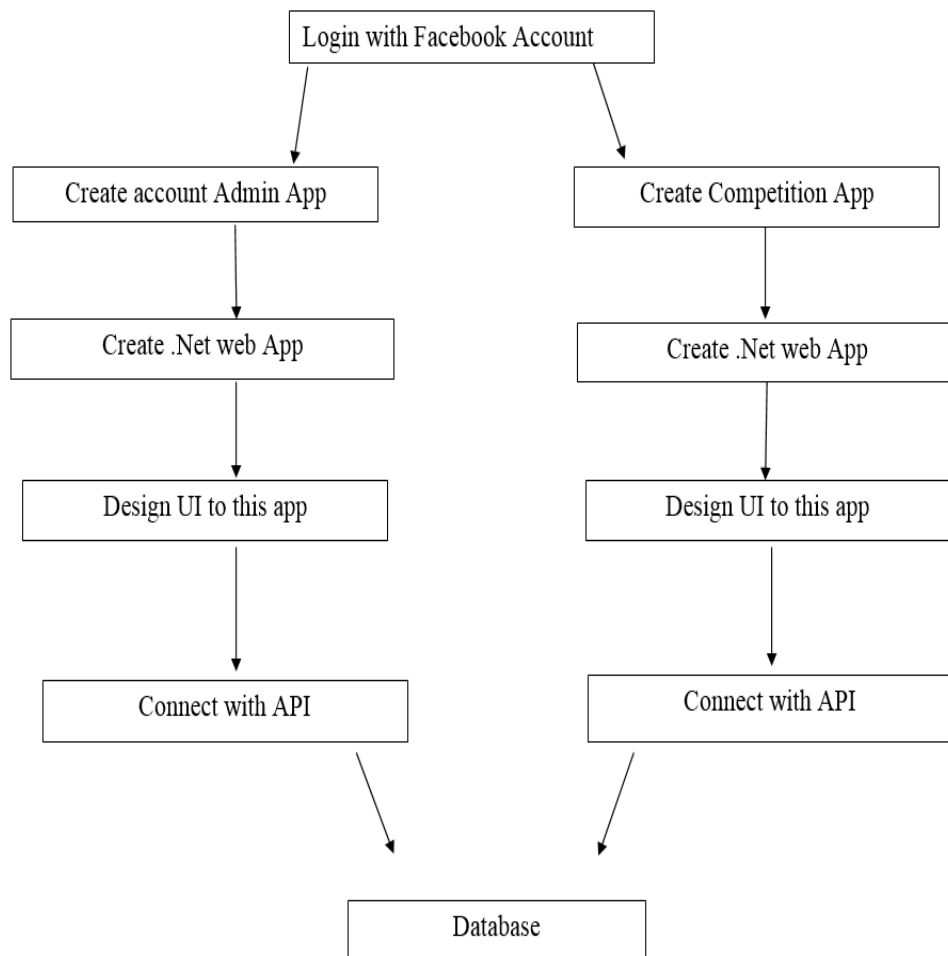
- 4) Programming main app features which depend on Facebook API or not and define app id to get connection with Facebook.
- 5) Login to this app by Facebook login achieve from Facebook developer page.
- 6) From this login achieve access token, object ID, which that a method for Facebook data access.
- 7) API data retrieve as JASON file, read this data and store it at DB.

## **2. Facebook, Admin account Process.**

- 1) Open Facebook website.
- 2) If user has FB account then enters email, password and then login.
- 3) If there is no account then signs in by entering the required information.  
After successfully create account, enter email and password then login to Facebook.
- 4) Request to build new Facebook app, naming this app and get app ID.
- 5) Using .Net environment to build new web app to define user interface to be able to login by Facebook login and get data from it.
- 6) Store this data at database there was defined before.

### 1.20.4.2 Facebook .Net Chart:

This chart shows steps to create Facebook app integrated with .Net web app and finally to retrieve data that stored into database as shown in Figure 13.



**Figure 13: Facebook, .Net app**

## 1.21 Industry Dictionary

For each industry there are common key words that people almost used to express their opinions, and we organized these key words in a well-designed database structure that facilitate and increase the efficiency of data processing, this dictionary contains four main tables that are considered the industry of research, also this database can be enriched with new key words and attributes by using its configuration interface that allow the user to enhance the accuracy as more key words added to the dictionary, the dictionary database has the below structure:

RANOOM-PC.Crime - dbo.Negative		
	NegativeId	NegativeWord
▶	1	هش
	2	هو
	3	لا
	4	ليس
	5	شيير
	6	ما

RANOOM-PC.Crime - dbo.Adverbs		
	AdverbID	AdverbName
▶	1	كثير
	2	كثير
	3	قليل
	4	صغير
	5	كثير
	6	إجمالاً
*	NULL	NULL

RANOOM-PC.Crime - dbo.Classes		
	ClassID	ClassName
▶	1	Food
	2	Service
	3	Price
	4	Place
	5	Overall
*	NULL	NULL

RANOOM-PC.Crime - dbo.Dictionary				
	IndustryId	ClassID	MainKeyWord	DicId
▶	1	1	الاكل	24
	1	1	الطعام	25
	1	1	الوجبه	26
	1	1	العصير	27
	1	1	الغداء	28
	1	1	الفلطور	29
	1	1	العشاء	30
	1	2	الخدمه	31
	1	2	الاستجابة	32
	1	2	الطلب	33
	1	2	التعامل	34
	1	2	التعاون	35
	1	3	السعر	36
	1	3	النحن	37
	1	3	القيمه	38
	1	3	الكلفه	39
	1	4	المكان	40
	1	4	التجهده	41
	1	4	الترتيب	42
	1	5	الوضع	43
	1	5	المجمل	44
	1	5	الشعور	45
	1	1	أكلو	46

RANOOM-PC.Crime - dbo.DicAttribute			
	IndustryID	ClassID	Attribute
▶	1	1	زاكي
	1	1	طازه
	1	1	طازج
	1	1	طيبه
	1	1	صحي
	1	1	لذيذ
	1	1	رائح
	1	1	بجن
	1	1	مترف
	1	1	باعتد
	1	1	نظيف
	1	1	وسخ
	1	1	عاديه
	1	2	مليحه
	1	2	سريته
	1	2	كويسه
	1	2	بطيئه
	1	2	رهيبه
	1	2	ممتازة
	1	2	زفت
	1	2	محترمه
	1	2	سبيله
	1	2	عاديه

Figure 14: Industry data dictionary

## 1.22 Data Normalization

After completing the data collection process, all posts are processed to get them ready for analysis and classification, this is achieved by running the normalization process that will scan the post and remove/clean it from undesired words such as repeated words and letter that are commonly used in posts writing, also to increase the efficiency of word matching and searching, the common Arabic characters that users always do not distinguish between them when writing the post are uniformed, English letters, punctuation ,numbers are all removed from the post before saving it in a new database field to get it ready for analysis and classification, below are the steps for the normalization process:

### Start.

- GetPost (PostID)
  - a. Remove stop word
  - b. Remove number
  - c. Remove prefix
  - d. Remove non Arabic letter
  - e. Remove duplicate words

- f. Remove duplicate letters
  - g. Replace common used Arabic letters
  - h. Remove punctuation characters
- Update master Posts table with normalized post.

الطعام كثير ممتاز وطيب بس في ملاحظة اكثر من زيارة في الاعياد تفاجئت انكم حرامية ونصابين عيني عينك مثلا اشترت اي شي؟ يضاف مبلغ معين على الفاتوره او بتشوف المنيو وتتفاجئ بسعر اخر على الفاتوره لما اتراجعه بيقولك هذه سعر قديم وما طبعنا المنيو الجديد ... اذا ما دقت راحت عليك وبصراحة ايام عجقه والاعياد ما حد بيدقق يعني مضمونه لكم مضمونه و المكان وسخ

---

Modify

---

طعام كثير ممتاز طيب بس ملاحظه اكثر زياره اعياد تفاجئت انكم حراميه نصاب عيني عينك مثلا اشترت اي شي يضاف مبلغ معين فاتوره او بتشوف منيو بتتفاجئ سعر اخر فاتوره لما اتراجع بيقول هذه سعر قديم ما طبعنا منيو جديد اذا ما دقت راحت عليك صراحه ايام عجقه الاعياد ما حد بيدقق يعني مضمونه لكم مضمونه مكان وسخ

**Figure 15: Data normalization process**

### **1.23 Database Schema:**

Database schema is designed to include a set of tables to store Facebook data that is collected from competition and Facebook pages described above, tables related to the competition itself and this will be out of the analysis process, the industry dictionary data tables are part of the project database schema, all these data table are the input for the classification process to compute the rate at the level of post, restaurant and location to calculate actual rate indirectly based on people behavior and impression instead of directly rating the request on Facebook. Figure 16 below describe the main data base tables and the relation between them.



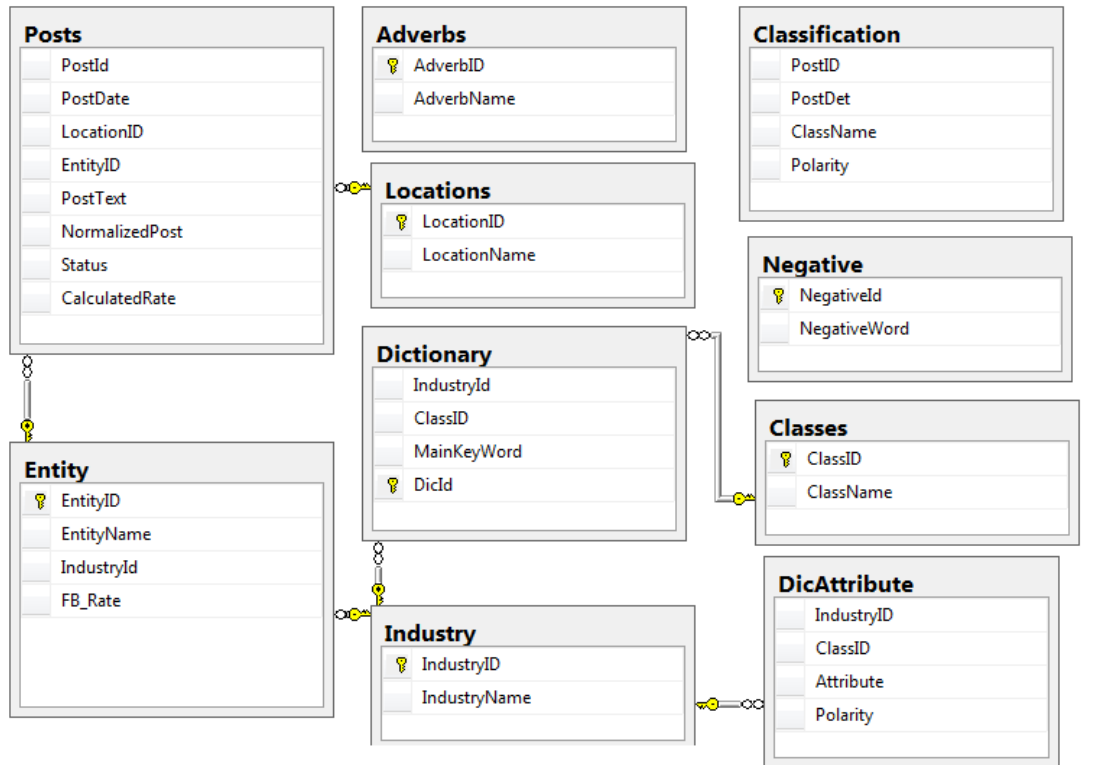


Figure 16: Database schema

## 1.24 Data Analysis and Classification

This section presents the core process of the thesis, the target of this process is to evaluate the post and compute a rate for it, this rate will be used as an indicator to the customer opinion/ satisfaction and his/her expected behavior accordingly, if the post rate is below the overall Facebook restaurant rate (outlier), we conclude that this customer behaved improperly and became a reason for a problem or crime, also the restaurant overall rate and location overall rate will be computed to measure the customers overall satisfaction, low rates indicates that these locations are good environment for crimes and the government should pay more attention and re-plan its existence to prevent any crimes.

The results of this process will be then compared with CID database for a certain period to check the relation and the feasibility of this thesis study results.

The classification process main steps include the following:

- Scan the post word by word till finding a key word based on the industry dictionary and fill it in with a new variable.
- Get the second word and check if it is negative word, if yes add it to the variable
- Check If it is an adverb, if yes add it to the new variable

- Check the attributes related to this key word based on the industry and the class of the key word and add it to the variable
- Get the polarity of the attribute and check if a negative word is found in the text , if yes get the opposite polarity (negative became positive and positive became negative)
- Create a new record for the split text of the post in the classification table and fill the section polarity and value (polarity value is 1 for positive and 0 for negative and 0.5 for neutral)
- Get the next word and repeat the process until the post is completed
- If the post did not contain the keywords for certain class , a new record will be created with neutral polarity (as the user did not complain or complement this class ), so each post will have the polarity for each class
- After that the sum of the polarities (1,0, 0.5) for a certain post will be computed to get the post rate.
- Update the post rate in the master posts table and get the next post.

At the end of this process we will have a rate for each post, this rate will be the subject of comparison to find out the outliers and to conclude the results.

## 1.25 Project Limitations

The biggest challenge in this thesis is a compilation of information through Facebook although we resolved some difficulties but we could not solve them permanently. The following are some of the limitations that we faced:

1. The public posts were collected by Facebook API so it is difficult to get any permission from Facebook to choose the whole information even after contacting Facebook and providing them with reports about the reasons for having this information.
2. Facebook API allows the retrieval of limited number of posts.
3. The working on the analysis of Arabic language was so difficult because there are few sources that used for the Arabic language in addition to the absence of a standard, which combines all the words and covers all cases.
4. The inability of processing the chat language that use sentences with Arabic characters and numbers.
5. Facebook posts do not have a fixed number of characters (such as twitter) and this caused a problem of losing part of these posts while retrieving the data.

6. Because of the existence of non-limit number of characters in Facebook post, this led to the problem that provided information through the post could not be written to expressive writing and there were data formats for different ages.

7. Additional sources of information from Division of cyber-crime at Jordan and Jordan CID provided information in the form of statistics and Excel files, which were process manually.

## 1.26 Project Environments

This project was built under Facebook, by create app depending on Facebook API to get data from Facebook. **API** is an interface used to get data from Facebook by create get request. API based on http to query data, manage ads, upload photos, post new stories and a variety of other tasks that an app might need to do. API is a social graph that represents Facebook information that includes the following:

1. Nodes (basically "things" such as a User, a Photo, a Page, a Comment).
2. Edges (the connections between those "things", such as a Page's Photos, or a Photo's Comments).
3. Fields (info about those "things", such as the birthday of a User, or the name of a Page).

Via standard HTTP, request methods. The data itself is usually wrapped in some standard format (such as JSON or XML) to make it easy to handle. Most Graph API requests will require the use of access tokens, which your app can generate by implementing Facebook Login, but in general, you can read APIs by making HTTP GET requests to nodes or edges on those nodes and each node has a unique **ID**, which is use to access it via the Graph API (Facebook, 2015).

## **ASP.NET Web Applications**

ASP.NET is a framework for building great Web sites and Web applications using HTML, CSS and JavaScript, which that runs in a web browser. It is created in a browser-supported programming language (such as the combination of JavaScript, HTML and CSS) and relies on a web browser to render the application.

This framework is very useful to build the suggested competition app, which read data from Facebook API. This app contains number of questions and users should answer them after Facebook login. By this login, we can get data from API while users enjoying with answering the competition questions to achieve prize.

## **SQL Server**

It is a relational database management system, and it is a software product with the primary function of storing and retrieving data as requested by other software applications which may run either on the same computer or on another computer across a network (including the Internet).

Competition app connect with SQL server database to store each users answer and check which one is winner, additional Facebook API data retrieve by JSON, read this JSON file and store it within app database.

## **C# Language**

It is an object-oriented programming language from Microsoft. It is designed to work with Microsoft's .Net platform. Microsoft's aim is to facilitate the exchange of information and services over the Web, and to enable developers to build highly portable applications. C# simplifies programming through its use of Extensible Markup Language (XML) and Simple Object Access Protocol (SOAP), we depend on it for analyzing Facebook API data.



# Chapter Four

## Result and Discussions

### 1.27 Introduction

In this chapter, we will present the output of each application/ process using the test data, then we will present the whole system output based on the actual data and after that we will find out the outliers match them with CID database to prove the feasibility of this application.

### 1.28 Competition Facebook Application

The main function of this application is to get the users or customer posts with its related information such as postdate, locations, these posts collected and stored in the posts master table, this table initially is filled with the PostIDd, PostDate, LocationID, EntityID (which represents the restaurant id) and the PostText, below is a sample of the collected data using this application as shown in Figure 17.



## **1.29 Normalization and Classification Application**

This application has two main functions, normalization and classification which described in the following sections

### **1.29.1 Normalization**

The Normalization process as described in chapter 3 , is a data cleansing process to standardize the post words by removing duplicated letters in the same word, duplicated words, English and number character, and replacing some Arabic letters to uniform and align it with the dictionary key words and attribute, the purpose of this process to increase the efficiency of post processing and splitting as well as to standardize the data dictionary

The input for this function is the Posts master table where the Post Text field is scanned and normalized and then stored in Normalized Post in the same table, below is a sample of this output as shown in Figure 18.



## 1.29.2 Classification

The main data manipulation process is performed in this part of the code, this function gets the normalized posts one by one and perform the following main steps to calculate the post rate:

- Scan the post word by word to find the key words based on the industry dictionary and attribute considering the negative and adverb word and the industry classes, the classes of a certain industry represent the main evaluation criteria for it (for restaurant industry we have five classes which are Food, Service, Price, Place and Overall )
- Split the post depending on the industry dictionary, attributes and classes, the output of this process is the classification table which have the Post ID, PostDet (the splitting output), ClassName and Polarity.
- At the end of the process, the function checks if the post generates the five sections of the classes or not, if not the remaining sections are filled and assigned Neutral Polarity value. For example if the customer did not complain or complement on a certain class, it will be considered acceptable (Neutral)

- The function is then computed the post rate based on the following weight:

- 1 for Positive Polarity classes
- 0.5 for Neutral Polarity classes
- 0 for Negative Polarity classes

Figure 19 presents sample of the testing process for these posts with the generated classes and the value assigned based on the polarity for each split (PostDet)

	PostID	PostDet	ClassName	Polarity	Rate
1	2	قعدة حلوه	Place	Positive	1
2	5	اكل طيب	Food	Positive	1
3	5	اسعار مناسبه	Price	Neutral	0.5
4	8	طعام كثير ممتاز	Food	Positive	1
5	13	مكان هاديء	Place	Positive	1
6	17	طعام ممتاز	Food	Positive	1
7	17	مقيلات ممتازة	Food	Positive	1
8	20	فطور ممتاز	Food	Positive	1
9	20	خدمة رائعه	Service	Positive	1
10	27	مطعم حلوه	Place	Positive	1
11	28	مطعم كثير ك	Place	Positive	1
12	31	مطعم رائعه	Place	Positive	1
13	32	مكان هاديء	Place	Positive	1
14	35	قعدة حلوه	Place	Positive	1
15	35	اكل يچتن	Food	Positive	1
16	35	طلب تاخر	Service	Negative	0
17	62	اكل طيب	Food	Positive	1
18	64	اكل ممتاز	Food	Positive	1
19	72	قعدة حلوه	Place	Positive	1
20	72	اكل يچتن	Food	Positive	1
21	72	طلب تاخر	Service	Negative	0
22	77	اكل طيب	Food	Positive	1
23	77	اسعار مناسبه	Price	Neutral	0.5
24	78	اكل ممتاز	Food	Positive	1
25	80	مطعم رائعه	Place	Positive	1
26	86	مطعم رائعه	Place	Positive	1
27	90	مطعم كثير ك	Place	Positive	1

Figure 19: Classification process results

The master Posts table then updated with the post rate bases on the classification process as listed in Figure 20.

	PostId	PostDate	L...	Ent...	PostText	NormalizedPost	Sta...	Rate
▶	1	2014-04-15	1	2	...المكان ها كان نظيبينيف بانعة!!!!!! بس الأكل كان زا	مكان ما نظيف اكل زاكى جدا خدمه طيبه	P	2
	2	2014-04-16	2	3	...القهقهه حلوه بس كان كتنبير مزين مزين	قعهه حلوه كتنبير مكان مزيج اكل يجنب طلب تاخر	P	1
	3	2014-04-17	3	4	...الأكل صحيح طازه وصحي بس سعره كثير عالي!!! والخب	اكل طازه صحي سعره عالي خدمه عاديه ما طيبه	P	1.5
	4	2014-04-15	1	5	...اليوم اكلت صندكم وجهه و كانت كتنبير رررررررررررررررررررر	يوم اكل صندكم وجهه كتنبير طيبه خدمه ما سريعه و مكان ضجه	P	1
	5	2014-04-15	1	4	...الأكل طيب بس الديفري مشكلته يتا اخر اسعار مناسبه و	اكل طيب بس ديفري مشكلته يتاخر اسعار مناسبه المكان نظيف	P	1.5
	6	2014-04-15	1	4	... بس الديفري بطيء حتى لو طلبت ع ساعه ١٢ الأكل راج	بس الطلب بطيء حتى لو طلبت ع ساعه اكل راج يوصل بل	P	0
	7	2014-07-15	1	4	مطعم جميل جدا بس عالي	مطعم جميل جدا بس عالي	P	1
	8	2014-06-15	1	4	...الطعام كتنبير ممتاز وطيب بس في ملاحظه اكثر من زيارة ف	... طعام كتنبير ممتاز طيب بس ملاحظه اكثر زياره ايجاد تناجي و حرايه نصاب	P	1
	10	2014-05-15	1	4	اسوء خدمه عنى الاطلاق	خدمه سيء اطلاق	P	2
	11	2014-04-16	1	4	أفضل وانصب مطعم سناك	أفضل مطعم انصب سناك	P	1.5
	12	2014-08-16	1	4	...خدمه سبتة جدا وخاصة الكاشير وبالنسبه للأكل كان مقر	خدمه سيء جدا خاصه كاشير نسبه اكل مقر ف جدا طول لها وصل	P	2.5
	13	2014-04-16	1	6	مكان هاديء جدا وعالي بنش الوقت	مكان هاديء جدا السعر عالي بنش قف	P	1
	14	2014-04-16	1	6	...المهم المطعم طيب الأكل زاكى ممتاز بس السعر عالي مقار	... مهم مطعم طيب اكل زاكى ممتاز بس سعر عالي مقارنه بمطعم رقيه مثل تل	P	3
	15	2014-09-16	1	6	...المكان كتنبير مرتب وراقي من كل النواحي التعامل والعوظف	... مكان كتنبير مرتب راقى كل نواحي تعامل موظفين موظفين اكل ممتاز تجربه	P	1
	16	2014-05-16	1	6	...من تجربتي ارض دير فمر مطعم أفضل عثمان يضافي ارقى رقى مطاعم عما	... من تجربتي ارض دير فمر مطعم أفضل عثمان يضافي ارقى رقى مطاعم عما	P	3.5
	17	2014-07-17	1	6	... جودة الطعام ممتازة واعتبره من الأفضل في عمان وانصح	... جودة طعام ممتازة اعتبره أفضل عمان النصح كبه نيه شره حصص صموما كا	P	2
	18	2014-07-17	1	6	...السعر عالي جدا وتوازي أسعار المطاعم داخل فنادق الخم	...سعر عالي جدا توازي اسعار مطاعم داخل فنادق خمسه نجوم اكثر لبعضا	P	2
	19	2014-07-17	1	6	...السعر عالي جدا وتوازي أسعار المطاعم داخل فنادق الخم	...سعر عالي جدا توازي اسعار مطاعم داخل فنادق خمسه نجوم اكثر لبعضا	P	2.5
	20	2014-04-17	1	6	...الطور ممتاز خدمه ممتازة ونظافة رائعه السعر جيد و بد	...طور ممتاز خدمه ممتازة نظافة رائعه سعر جيد لديهم تنويع في قائمه الأ	P	2
	21	2014-04-15	1	6	...المكان يجنب وحلو ومرتب والخدمه رائعه وينصح الكل بروج	مكان يجنب حلو مرتب خدمه خدمه رائعه وينصح كل بروج	P	3.5
	22	2014-04-16	1	6	...المحل مرتب وحلو جربنا بعيد ام ماضي كانت خدمه اكثر من ممتازة و ب	...محل مرتب حلو رحنا جربنا بعيد ام ماضي كانت خدمه اكثر من ممتازة و ب	P	4
	23	2014-06-16	1	6	...المكان حلو بس المشكله اناو اكلو ما تجبني ولا الخدمه عنى	...مكان حلو مشكله اكلوا ما تجبني خدمه مبيحه ينصل اتنادي ساعه عثمان ح	P	5
	23	2014-04-17	1	6	...أنا بعب محل عالي شوي مشاوي مقبلات ممتاز نظيف كتنبير قعهه ممتازة منا	...أنا بعب محل عالي شوي مشاوي مقبلات ممتاز نظيف كتنبير قعهه ممتازة منا	P	4.5

Figure 20: Calculated post rate

### 1.29.3 Test Data

The test data process was conducted for about 20 actual post to make sure that the developed application is ready to deal with these posts and generate the expected results, the test process was repeated many times and the code maintained to cover all possible cases, the dictionary and attributes were also updated by adding new words, update some of them based on the testing results for both normalization and classification functionalities, new methods and classes also developed to process new cases.

### 1.30 Results and Interpretations

After completing the testing process, we ran the application to process all collected posts in the period between April to October 2014, which were about 600 post processed in this final run , the output of this run is 3000 classifications (5 for each post) for the targeted locations and restaurants, the rate data field is computed based on these classifications.

Now by running some SQL statements on the posts master table we came up with the following results:

- 1- The outliers (Restaurants with computed rate less than Facebook rate )  
are:



- a. KFC Facebook rate is 4 while the computed rate is
- i. 3 for location number 3 (Abo Nsair and Al-Jbaiha )
  - ii. 2 for location number 5 (Mecca St. Khalda , Um Alsommaq ,Abdallah Ghosheh ,Um Othaina and Al-Rabiah)

**And KFC computed rate for all locations is 4**

- b. McDonalds Facebook rate is 4.5 while the computed rate is
- i. 3 for location number 5 (Mecca St. Khalda , Um Alsommaq ,Abdallah Ghosheh ,Um Othaina and Al-Rabiah)
  - ii. 4 for location number 1 (Tla'a Al-Ali, Al-Gardenz and Al-Madina Almonawara)

**And McDonalds computed rate for all locations is 4**

- c. Pizza Hut Facebook rate is 5 while the computed rate is
- i. 3 for location number 3 (Abo Nsair and Al-Jbaiha)

**And Pizza Hut computed rate for all locations is 4**

- d. Chilli House Facebook Rate is 3 while the computed rate is

- i. 1 for Location number 1 (1 (Tla'a Al-Ali, Al-Gardenz and Al-Madina Almonawara)
- ii. 2 for location number 5 (Mecca St. Khalda , Um Alsommaq ,Abdallah Ghosheh ,Um Othaina and Al-Rabiah)

**And Chilli house computed rate for all locations is 3**

- e. Lebnani Snack Facebook Rate is 4.5 while the computed rate is
  - i. 3 for Location number 4 (1 (Abdoun, Dair Ghbar and Alyasamin)
  - ii. 3 for location number 5 (Mecca St. Khalda , Um Alsommaq ,Abdallah Ghosheh ,Um Othaina and Al-Rabiah)

**And Lebnani Snack computed rate for all locations is 3**

2- We computed the location rate based on the customers posts (regardless of the restaurant) and got the following results:

- a. Location number 1 rate is 2
- b. Location number 2 rate is 4
- c. Location number 3 rate is 2
- d. Location number 4 rate is 4
- e. Location number 1 rate is 3

We can conclude from these results that the level of satisfaction is good in locations number 2 and 4 (Al-Madina AlRyadiyah, Al-Shmeisani, Abdoun,Dair Ghbar, Dahiet Al-Yasmeen) , while it is not acceptable in Locations 1, 3 and 5 (Tla'a Al-ali, Al-Gardenz, Al-Madina Almonawara, Abo-Nsair, Al-Jbaiha, Mecca St, Khalda, Al- Rabieh, Um-thainah, Abdallah Ghosheh)

People always complain when there is something wrong or disturbing and they do not always complement or tell about good things, if you for example visit a restaurant to get your breakfast and you found that your car is scratched when you left, you will tell all your friends about this what happened to your car, and you might move the responsibility to the restaurant and start shouting and ready to make a fight, however when you visit a restaurant with valet service you will find it something normal and you will not inform your friends about it.

The level of satisfaction is the indicator that these areas are preferred by the customers and leads us to conclude that the whole environment in these areas such as good public services, good parking, and the most important its safety are acceptable and people feel happy and safe when visiting these areas.

On the other hand lower safe areas are always attractive for criminals and vandals and there existence in these areas will be noticed by people and people

will complain and tell their friend and will post their concerns and complains in the social media.

Comparing these result with Cyber Crime Reports which organized by grouping the areas of reported crimes into 5 locations similar to those areas considered in the application, and comparing the number of crimes in each location with the location rate generated from the application we can find low rate location was higher in number of crimes and vice versa.

Cyber Crimes per location			
Location Number	Location Areas	Number of Crimes	Computed Rate
1	Tla'a Al-ali, Al-Gardenz, Al-Madina Almonawara	15	2
2	Al-Madina AlRyadiyah, Al-Shmeisani	12	4
3	Abo-Nsair, Al-Jbaiha	19	3
4	Abdoun,Dair Ghbar, Dahiet Al-Yasmeen	11	4
5	Mecca St, Khalda, Al- Rabieh, Um-thainah, Abdallah Ghosheh	20	3

### **1.31 Conclusion:**

The implemented application provides the country more control to predict the level of security and safety at any location by analyzing social media contents that contain valuable information about people feelings and opinions regarding specific location or cluster, the government and CID can benefit from this application to monitor and strengthen their existence in the required locations (locations with low safety and satisfaction rate) to prevent the crime before it occur.

## Chapter Five

### Conclusion and future works

#### 1.32 Introduction

In this thesis, we tried to study one of the most important topics to increase the protection level of the country by predicting the crime locations. These objectives can be achieved by providing the concerned government agencies with the predicted locations based on people feeling and opinions to manage the process of its forces in these locations to prevent the crimes before they happen.

#### 1.33 Conclusion

Facebook is rich with different data categories and users adopted it in all life aspect and share all emotion and opinion through it.

Any research by Facebook is difficult but more accurate because Facebook has unlimited spaces given to any users to write and participate with each other's, that gives wider and more comprehensive view for any subject.

Crime prediction or crime location prediction is very useful to achieve the protection which will result in a higher service quality at different locations and provide new services to the population under high level of protection and safety.

The developed application proved that Facebook data can be analyzed in a way to provide good indicators about people feeling and opinions to the one of most important needs for human being which is the feeling of safety, these indicator can be the trigger to the security agencies to mobilize their resources in the predicted locations to prevent criminals and vandals from committing a crime and protect people from being a victim for these criminals.

### **1.34 Future Works**

The implemented application satisfied the desired requirements, and it can be improved by the following:

1. Deep study and practice for text analysis especially Arabic language and solve chat language issue to give the ability for analysis.
2. Develop data mining analysis algorithms to give wider view for any location and any crime, and help to provide high level of protection. All of this will help to get authentication need for any social media data.
3. Develop study to make ability to know which the criminal of any crime, by studying users behavior and habits.
4. Study emotion of all country news to gain the ability to control criminal activity.

## References

Analytics Guide (on-line) 2014 . Available: [googlegrantssupport@google.com](mailto:googlegrantssupport@google.com)

Apple. What is IPv6? (on-line ) 2015 . Available :

<https://support.apple.com/en-us/HT202236>

Arora, H. (2011). TCP/IP Protocol Fundamentals Explained with a Diagram

(on-line). Available: [http://www.thegeekstuff.com/2011/11/tcp-ip-](http://www.thegeekstuff.com/2011/11/tcp-ip-fundamentals/)

[fundamentals/](http://www.thegeekstuff.com/2011/11/tcp-ip-fundamentals/).

Bokhari, Z. (2012). Role of Communication in the Development of Human

Society (on-line). Available:

[https://www.academia.edu/5182078/Role of Communication in the](https://www.academia.edu/5182078/Role_of_Communication_in_the_Development_of_Human_Society)

[Development of Human Society.](https://www.academia.edu/5182078/Role_of_Communication_in_the_Development_of_Human_Society)

2015 Brightkite (on-line). Available:

[http://www.adweek.com/socialtimes/mobile-social-network-brightkite-](http://www.adweek.com/socialtimes/mobile-social-network-brightkite-launches-facebook-connect-integration/215948)

[launches-facebook-connect-integration/215948](http://www.adweek.com/socialtimes/mobile-social-network-brightkite-launches-facebook-connect-integration/215948)

Bucki, J. (2015). Definition of Mobile Computing (on-line). Available:

[http://operationstech.about.com/od/glossary/g/Definition-Of-Mobile-](http://operationstech.about.com/od/glossary/g/Definition-Of-Mobile-Computing.Htm)

[Computing. Htm](http://operationstech.about.com/od/glossary/g/Definition-Of-Mobile-Computing.Htm)

Chen, G., and Nan, L. (2011). Analysis of a Location-based Social Network.



Chen, T., Kaafar, M., and Boreli, R. (2013). **The Where and When of Finding New Friends: Analysis of a Location-based Social Discovery Network.**

Chung-Hsien, Y., Max, W., Ward, M., and Wei, D. (2012). **Crime Forecasting Using Data Mining Techniques, International Conference on Data Mining.**

CID (on-line) 2015 . Available: [www.cid.psd.gov.jo/](http://www.cid.psd.gov.jo/).

Cosma, D. (2013). Advantages and Disadvantages of Social Networking (on-line). Available:

[http://socialnetworking.lovetoknow.com/Advantages and Disadvantages of Social Networking.](http://socialnetworking.lovetoknow.com/Advantages_and_Disadvantages_of_Social_Networking)

Das. T., and Kumar, P. (2013). Big **data analytics a framework for unstructured data analysis**, International Journal of Engineering Science & Technology.

Dialogic, (2010). Introduction to Cloud Computing, 2010.

Dion, H. (2010). **the "Big Five" IT trends of the next half decade: Mobile, social, cloud, consumerization, and big data**

Duggan, M., Ellison, N., LampeCliff, L., And Madden, M. (2014). **Social Media Update while Facebook remains most popular site.**

ELI, (2015). **Educause learning initiative**, 7 things you should know about Facebook II (online). Available:

<https://net.educause.edu/ir/library/pdf/eli7025.pdf>.

[Enquiries@ukessays.com](mailto:Enquiries@ukessays.com), (2015). Scada Stands for Supervisory Control and Data Acquisition Information Technology Essay (on-line). Available:

<http://www.ukessays.com/essays/information-technology/scada-stands-for-supervisory-control-and-data-acquisition-information-technology-essay.php>

Facebook (2015). (online). Available:

<https://developers.facebook.com/docs/graph-api/overview>

Foursquare, (2015). (online). Available:

<http://www.nyu.edu/content/dam/nyu/studentAffairs/images/Explained/foursquare.pdf>

Gao, H., and Liu, H. (2014). **Data Analysis on Location-Based Social Networks**.

GC, Tutorial-Reports.com, (2013), Introduction to Wireless Networking (online). Available: <http://www.tutorial-reports.com/wireless/introduction.php>.

Gerber, S. (2014). Predicting Crime Using Twitter and Kernel Density Estimation.

Graham, W. (2008). Facebook API Developers Guide.

GrigorikIlya, (2013). High Performance Browser Networking, Chapter 5: Introduction to Wireless Networks (on-line). Available: <http://chimera.labs.oreilly.com/books/1230000000545/ch05.html>.

Groff Elizabeth, R., Nancy, G. Vigne F. (2003). Casting the Future Of Predictive Crime Mapping (online). Available:

[http://cdn2.hubspot.net/hub/53/blog/docs/facebook\\_for\\_business\\_ebook\\_hubspot.pdf](http://cdn2.hubspot.net/hub/53/blog/docs/facebook_for_business_ebook_hubspot.pdf). <http://www.internetworldstats.com/stats.htm>.

<http://www.windowsnetworking.com/articles-tutorials/wireless-networking/Introduction-Wireless-Networking-Part1.html>. <https://blog.udemy.com/advantages-and-disadvantages-of-social-networking/>

<https://www.ietf.org/>

<https://www.ietf.org/>

Hupspot. (2015). how to use Facebook for business

HuthAlexa and Cebula, J. (2011). The Basics of Cloud Computing.

Infoplease, (2015). Most Popular Internet Activities (on-line). Available:

<http://www.infoplease.com/ipa/A0921862.html>

Internet World State (2014). Internet Usage Statistics

Investintech.com. (2015). a brief Guide to the History of the Internet (on-line). Available: <http://www.investintech.com/content/historyinternet/>.

Irena, P., and Florian, M. (2014). **Monitoring Trends on Facebook.**

Jackson, G., Demand, M. (2015). The Advantages of Social Network Sites (on-line) Available: <http://smallbusiness.chron.com/advantages-social-network-sites-32069.html>

Janssen, C. (2015). **Computer Network**, Janalta Interactive Company (on-line). Available: <http://www.techopedia.com/definition/25597/computer-network>.

Mattern, F., and Floerkemeier, C. (2010). **Distributed Systems Group**, Institute for Pervasive Computing, ETH Zurich From the Internet of Computers to the Internet of Things.

Mikoluk, K. (2013). Advantages and Disadvantages of Social Networking: Should You Spend More Time Connecting Online?.

[Neo@macsimumnews.com](http://Neo@macsimumnews.com) (2015). Intel's Wireless Multi-Core Patent Applications (on-line). Available: <http://www.mactech.com/content/intels-wireless-multi-core-patent-applications-june-2006>.

Odessa, C. (2015). **Computer and Networks (on-line) Available:** <http://www.conceptdraw.com/samples/network-diagram>.

Ricci, F. (2012). **Internet Technologies On the Security of Cloud Storage Services.**

Ryun, N. (2015). **Facebook A Beginner's Guide.**

Saruumaan, (2013). "Social Media Essay." StudyMode.com. (on-line). Available: <http://www.studymode.com/essays/Social-Media-Essay-1399465.html>.

Shaomei W., and Adamic, L. (2015). Visually Impaired Users on an Online Social Network

Sujatha, R., and Ezhilmaran, D. (2014). **An Adaptive Method for Analyzing and Predicting the Crime Locations by Means of AMABC and ARM.**

Sukanya, M., Kalaikumaran, T. and Dr. Karthik, S. (2012). **Criminals and crime hotspot detection using data mining algorithms: clustering and classification.**

Tabona, A. (2004). **An Introduction to Wireless Networking.**

Tony H. (2014). **Detecting Hot Spots Using Cluster Analysis and GIS.**

[tutorialspoint.com](http://www.tutorialspoint.com) (2015). Data Communication and Computer Network (online). Available:

[http://www.tutorialspoint.com/data\\_communication\\_computer\\_network/data\\_communication\\_computer\\_network\\_tutorial.pdf](http://www.tutorialspoint.com/data_communication_computer_network/data_communication_computer_network_tutorial.pdf).

Uttam, M., Srinivas, Y., and Murthy, J.V.R. (2012). **An Intelligent Analysis of Crime Data Using Data Mining and Auto Correlation Models.**

Vermesan, O., Sintef, N., Friess, P. and EU, B. (2013). **Internet of Things: Converging Technologies for Smart Environments and Integrated Ecosystems.**

Vyncke, E. (2009). **IPv6 over Social Networks.**

Weaver, J. and Tarjan, P. (2012). **Facebook Linked Data via the Graph API.**

Weiye, S., and Zhen, M. (2013), **Assisting Developers of Big Data Analytics Applications When Deploying on Hadoop Clouds.**

Wieber, J. (2015). **the Internet of Things and the Future of Social Commerce.**

Wike, R., and Horowitz, J. (2012). **Social Networking Popular Across Globe,** from Pew Research Center.

Wiki, (2015). (on-line). Available:  
[http://en.wikipedia.org/wiki/Geosocial\\_networking](http://en.wikipedia.org/wiki/Geosocial_networking).

Zegna, V., and Pepper, M. (2005). The Internet Explained (on-line).  
Available: <http://www.southbourne.com/articles/internet-explained>.

Zheng, Y. (2012). **Tutorial on Location-Based Social Networks.**