TRIBHUVAN UNIVERSITY
INSTITUTE OF ENGINEERING
PULCHOWK CAMPUS

Spatio-Temporal Crime Prediction Model in Kathmandu valley using GIS

By
Krishna Prasad Neupane
(069-MSCSKE-657)

A FINAL THESIS REPORT
SUBMITTED TO THE DEPARTMENT OF COMPUTER AND ELECTRONICS ENGINERING
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF MASTER OF SCIENCE IN COMPUTER SYSTEM AND KNOWLEDGE ENGINEERING

DEPARTMENT OF COMPUTER AND ELECTRONICS ENGINERING
LALITPUR, NEPAL

June 2015

COPYRIGHT

The author has agreed that the library, Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering may make this thesis freely available for inspection. Moreover, the author has agreed that permission for extensive copying of this thesis for scholarly purpose may be granted by the professor who supervised the work recorded herein or, in their absence, by the Head of the Department wherein the project report was done. It is understood that the recognition will be given to the authors of this thesis and to the Department of Electronics and Computer Engineering, Pulchowk Campus, and Institute of Engineering in any use of the material of this thesis. Copying or publication or the other use of this thesis for financial gain without approval of the Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering and author's written permission is prohibited.

Request for permission to copy or to make any other use of the material in this report in whole or in part should be addressed to:

Head

Department of Electronics and Computer Engineering

Pulchowk Campus, Institute of Engineering

Lalitpur, Kathmandu

Nepal

RECOMMENDATION

The undersigned certify that it has been read and recommended to the Department of Electronics and Computer Engineering for acceptance, a final-term draft of thesis entitled "**Spatio-Temporal Crime Prediction Model in Kathmandu valley using GIS"**, submitted by "**Mr. Krishna Prasad Neupane"**, in partial fulfillment of requirement for the award of the degree of "**Master of Science in Computer and Knowledge Engineering".**

_____

Supervisor: Dr. Dibakar Raj Pant

Head of Department

Department of Electronics and Computer Engineering

Institute of Engineering,

Pulchowk Campus

_____

External examiner: ……

_____

Committee Chairperson: ……..

Department of Electronics and Computer Engineering

Date

_____

ABSTRACT

Crime is an illegal act deviating from normal violation of the norms giving losses and harms for people. Social, psychological, economical and environmental factors are to be considered in crime issue. All these concepts affect occurrence of crime in different ways. Peoples who have role in crime prediction are police, local governments, law enforcement agencies and people exposed to crime and offenders. The spatio and temporal model is generated by using crime data for the year 2070 in Kathmandu police Headquarters. Methodology starts with obtaining clusters with K-mean, Nearest and Neighborhood(Nnh) and Spatial and Temporal Analysis clustering(STAC) algorithms. Above discussed clustering methods are compared in terms of number of crimes andland-use to select the most appropriate clustering algorithm. Crime data is divided into daily epoch, to observe spatio and temporal distribution of crime over the Kathmandu valley.To predict crime in time dimension a time series model (ARIMA) is fitted for each week day. Thespatial and temporal model of this thesis can give crime prediction in both space and time.

Keywords:Spatial, temporal, ellipse, clustering, Spatial and Temporal Analysis of Crime (STAC),Euclidean distance, Geographical Information System (GIS), Autoregressive and Integrated Moving Average (ARIMA).

ACKNOWLEDGEMENT

Firstly, I want to express my deepest gratitude to my supervisor Assist. Prof. Dr. DibakarRaj Pant for his guidance and confidence through the development of my thesis.

I want to thank all my instructors Prof. Dr. Sashidhar Ram Joshi, Prof. Dr. SubarnaShakya Assist. Prof. Dr. SanjeebPandey for their guidance, advice, criticism, encouragements and insight throughout the studies in IOE and bring me to prepare this thesis.

I also grateful to the police headquarter for unhesitatingly supplying the sample data and adding their experiences and perspectives with spatial data.

I would like to thank Dr.Amar Deep Regmi and Dr. BhagwatRimal for contributing their support during my thesis period in Geographic Information Technologies.

At last, special thanks to my family supporting, encouraging me during my thesis and send their love from Bastari,Syangja.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

GIS	Geographical Information System
STAC	Spatio-Temporal Analysis of Crime
VDC	Village Development Committee
Nnh	Nearest neighborhood
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
AR	Auto Regressive
MA	Moving Average
ARMA	Auto Regressive and Moving Average
ARIMA	Auto Regressive Integrated Moving Average
SAR	Seasonal Auto-Regression
SMA	Seasonal Moving Average
DF	Degree of Freedom

LIST OF FIGURES

LIST OF TABLES

CHAPTER ONE

INTRODUCTION

1.1 Background

Crime can be defined as a disorder on behavior that is an integrated result of social, economical and environmental factors. Nowadays crime analysis is gaining significance and one of the most popular subjects is crime prediction. Persons of crime intend to forecast the place, time, number of crimes incidents and crime types to get precautions. With respect to these intentions, in this thesis a spatio-temporal crime prediction model is generated by using time series forecasting with simple spatial approach in Geographical Information Systems (GIS).

Crime is an integrated result of social, political, economical and various conditions that happen in a specific geography in a specific period of time. Why and where crime takes place is quite important to analyze the three main reasons of crime: A likely offender, a suitable target and an absence of guardian. Crime can be prevented or reduced by making people less likely to be offend, making targets less vulnerable, and by making guardians more available.

All categories of crime require different prevention techniques. The basis of environmental criminology is crime triangle in Figure 1. Crime triangle states that a criminal activity occurs when a vulnerable target and a motivated offender meet in a convenient environment.

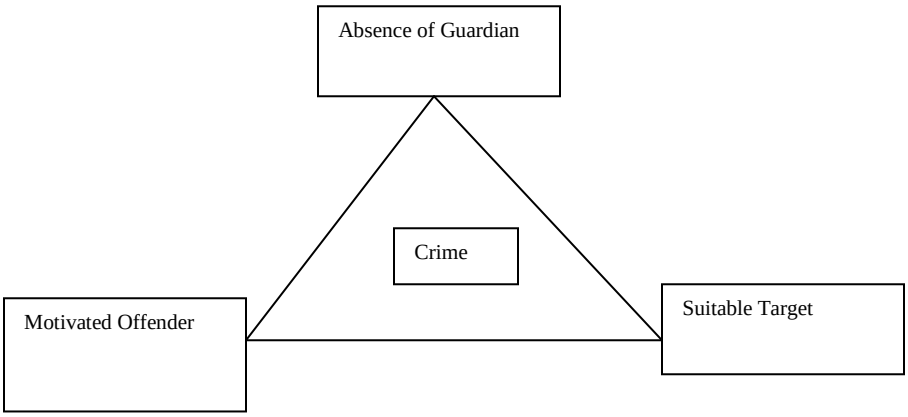


Figure 1:- Crime Triangle (source:H. Al-Madfai, 2007)

Street crimes, organized crimes, drug crimes, political crimes are main categories of crime events. All these appearing types are subdivided into different kinds of crime. For example, women and child, social, and theft are types of crime categorize under street crime. Stakeholders of crime mapping are police departments, politicians, nongovernmental organizations, local governments, law enforcement agencies and community. While planning crime prevention measures, stakeholders make use of crime theories to understand the spatial phenomenon in crime.

Spatial analysis is concerned with the geo-referencing data and temporal analysis is used to analyse the time series data. Clustering is common technique for arranging the data and is used to create the hotspots on this thesis. Clustering of the spatial data is carried out by the use of suitable clustering algorithms.

The spatio temporal modelis generated by utilizing crime data for the year 2070 BS in KathmanduValley. Methodology starts with obtaining clusters with different clustering algorithms. Then clustering methods are compared in terms of land-use and representation to select the most appropriate clustering algorithms. Later crime data is divided into daily epoch, to observe spatio temporal distribution of crime.

In order to predict crime in time dimension a time series model (ARIMA) is fitted for each week day [1]. Then the forecasted crime occurrences in time are shown according to spatial crime cluster patterns. Therefore, the model of this thesis can give crime prediction in both space and time.

1.2 Problem Definition

Crime is the abnormal behavior that caused losses in the each place of the society.Kathmandu valley is one of the insecureplaces of the Nepal regarding crime activities. Basically, crime activities were happened on the basis of deferent perspectives. There are different types of crimes happened in the Kathmandu valley, which are increasing day by day. This is the main problem currently faced by the police stations.

Another issue in crime analysis is crime forecasting. Gorr and Harries [2],indicated the purpose of crime forecast to directly support crime prevention and law enforcement. Developing highly reliable methods for forecasting future crime trends and problems is one of the most preferred ways to improve crime prevention and reduction measures. With the advance of crime forecasting,

It has been necessity to address the issue of crime by formulating a policy and implementing relevant programmes to minimize the existing effects and likely impacts in different ecological regions.Preventing crime is a necessity to make people live in more peaceful world. To achieve

more calm and secure life, police is the most responsible foundation for crime prevention by targeting of resources. Police use strategic, tactical and administrative policies that assist to take precautions before an occurrence of a criminal activity. To make effective policies and improve prevention techniques, police should make use of criminological theories and crime analysis.

In the scope of this thesis clustering analyses are used to identify hot spots. Cluster analysis aims to collect data into groups according to several algorithms. Two main groups of clustering occurring in spatial data analysis are hierarchical and non-hierarchical/partitioning approaches. Partitioning approaches use optimization procedures to divide data into meaningful groups. K-means is the example of partitioning approach with different algorithm [2]. Hierarchical approaches group data set according to the type of distance specified in the algorithm. Nearest neighborhood hierarchical clustering is a type of clustering approaches. Also, there are new clustering techniques generated for specific purposes. One of them is Spatio-Temporal Analysis of Crime (STAC) which is a powerful tool to identify crime patterns and detect crime hot clusters. The other specific tool, which differing from the other algorithms by considering the underlying population when generating hot spots.

With the analysis of data on the form of spatial and time series, spatial and temporal predictions of crime are used to make long and short term planning.

1.3 Objective of the Thesis

The main objective of this thesis is to develop a spatio-temporal crime prediction model based on geographical information systems coupled with spatial statistical methods.

Specific Objectives of the task

The specific objectives of the tasks are as follows:

- To detect the spatial pattern of crime.
- To identify appropriate clustering algorithm for mitigation of crimes in selected study areas regarding social, economic and environmental issues.
- To predict the number of crime in time.
- To measure the potential reductionof crime issues by forecasting in weekday basis.

1.4 Scope and Limitation of Work

The scopes of the study consist of following areas;

- To develop questionnaires for detail analysis of crime pattern in study areas.
- To have the field visit in selected district of Nepal for research purpose.

- To study the present crime status to have detail study through it.
- To find possible forecasting technologies for crime mitigation in the region using simple spatial and temporal system (including stakeholder and local people view).
- To recommend possible measures as one time effort in such vulnerable areas.

The limitations of the study consist of following areas;

- There is no digitized spatial data on the police headquarter.
- Inclusion of all the points is one of the main limitations of this approach.
- Spatial outliers are forced to be included to clusters, hence cluster orientation and sizes are deviating from the optimal.
- There is no option to predict the location of the crime.
- The limitation in the study is the size of data which only consisting one year.

1.5 Organization of the Thesis

To outline the thesis, study area and literature review are explained in the second chapter. Also, crime analysis with geographical information systems is in the scope of the second chapter.

In the third chapter, data and methodology are described, the methodology of thestudy and the information about methods and techniques are given.The fourth chapter involves the analysis and comparison of clustering algorithms.
The concept of fifth chapter is to generate a crime prediction model withstatistical model fitting approach. Box-JenkinsARIMA model is used for forecasting future crime occurrence in time. Then, crime prediction model is generated and applied to the studyarea to indicate the results.
The last chapter is the discussion and conclusion part evaluating the results of thisthesis.

CHAPTER TWO

STUDY AREA AND LITERATURE REVIEW

In this chapter, study area in terms of geographical locations is defined and the different literatures are adapted for spatio-temporal crime prediction model is explained.

2.1 Description of the Study Area

The study area is Kathmandu valley, which is the major area of Nepal. The populationof Kathmandu valley is crowded due to the most developed place of Nepal. All peoples are from different districts of Nepal and they are coming with their high expectation of life. Due to the large number of people and also lack of the opportunities some people are trying to make their profession as a criminal. Following figure shows the study area of the thesis:

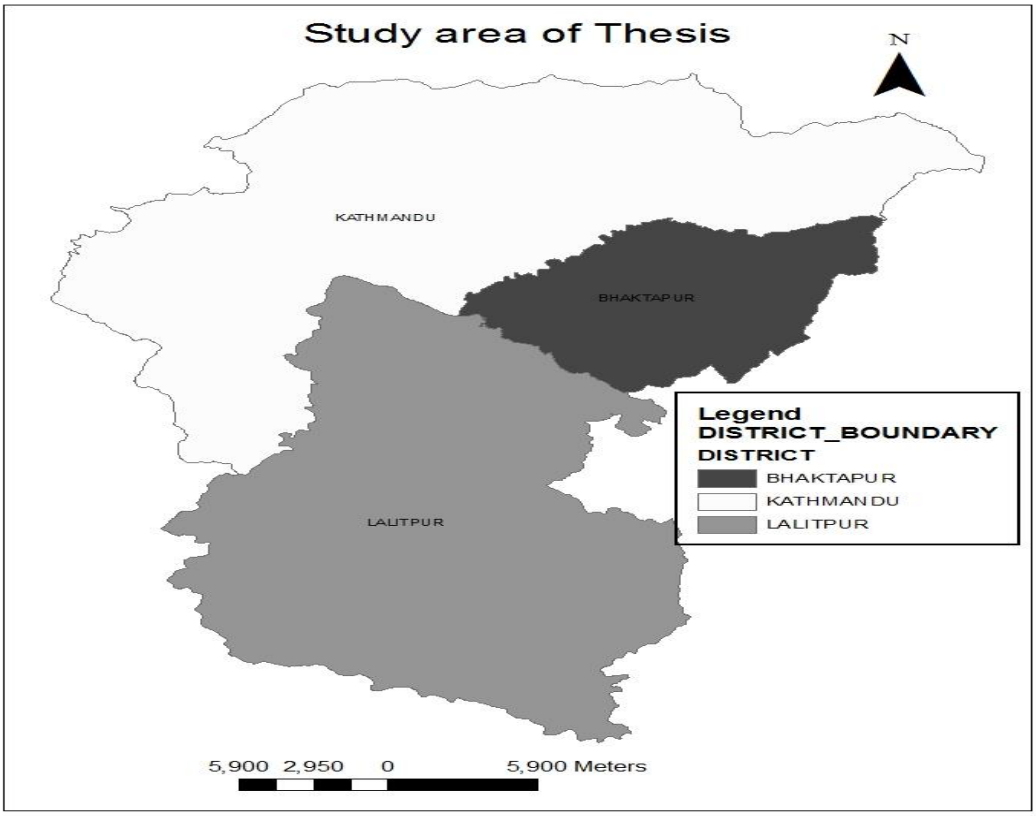


Figure 2:- Kathmandu Valley (study area)

Having both cultural and commercial facilities in Kathmandu, it is the most developed district in Kathmandu Valley. Evidently, nearly 60 percent of public associations in Nepal are in this district. Kathmandu has various types of land-use areas from residential, commercial, public buildings, parks, museums to military zones, etc. It is more probable to observe different types of crime in a mixed type of land-use.

As stated before, Kathmandu is an important place for cultural and sport activities. There are lots of cinemas, theatres, swimming pools and parks located in the district. Also, business centers, shopping malls are densely populating the district which contributes the economical development of the city. Therefore, Kathmandu is one of the most vital and active part of Kathmandu Valley attracting offenders to commit crime.

2.1.1 Kathmandu Valley

The Kathmandu Valley is the most developed and populated place in Nepal. The majority of offices and headquarters are located in the valley making it the economic hub of Nepal. It is popular with tourists for its unique, rich, culture and architecture; including the highest number of jatras in Nepal. Kathmandu valley consists of three districts namely: Kathmandu, Lalitpur and Bhaktapur. Kathmandu valley is surrounded by the hilly and most of the places of the Kathmandu valleys are hilly regions. These three districts are described one by one in following topics [3].

The GIS map of buildings and roads are shown in figure below:

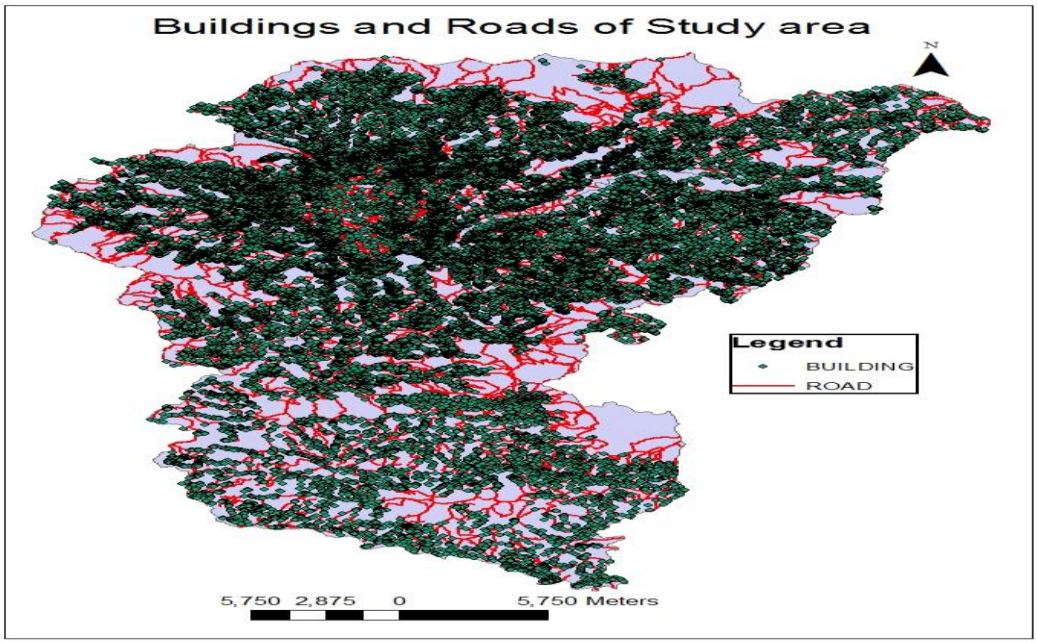


Figure 3:- GIS map of Buildings and Roads in Kathmandu valley

2.1.1.1 Kathmandu

Kathmandu district is one of the three districts located in Kathmandu valley; it covers an area of 395 km^2. Kathmandu is highly populated district of the Nepal about 1,744,240 in 2068 census report. In Kathmandu, major crimes are happen. Crimes are happened due to the large number of unmanaged population, bazaar area and also the poor security management.

In Kathmandu most of the people are living and fighting for their job.This is the main factor of creating a problem regarding generation of crime. The main business spot of the country is Kathmandu district. People are trying to achieve success by establishing their huge amount of money but what they think before they won't able to got the success which committed to do some crime or push them to accept any types of criminal activities.

Most of the homeless people and refuges are presence on this district. People from these sectors are illiterate and unemployed. Due to this reason they are accepting the different types of crime by involving in different places of the district.

Geographical map of the Kathmandu district is shown below with the all Village Development Committees (VDC):

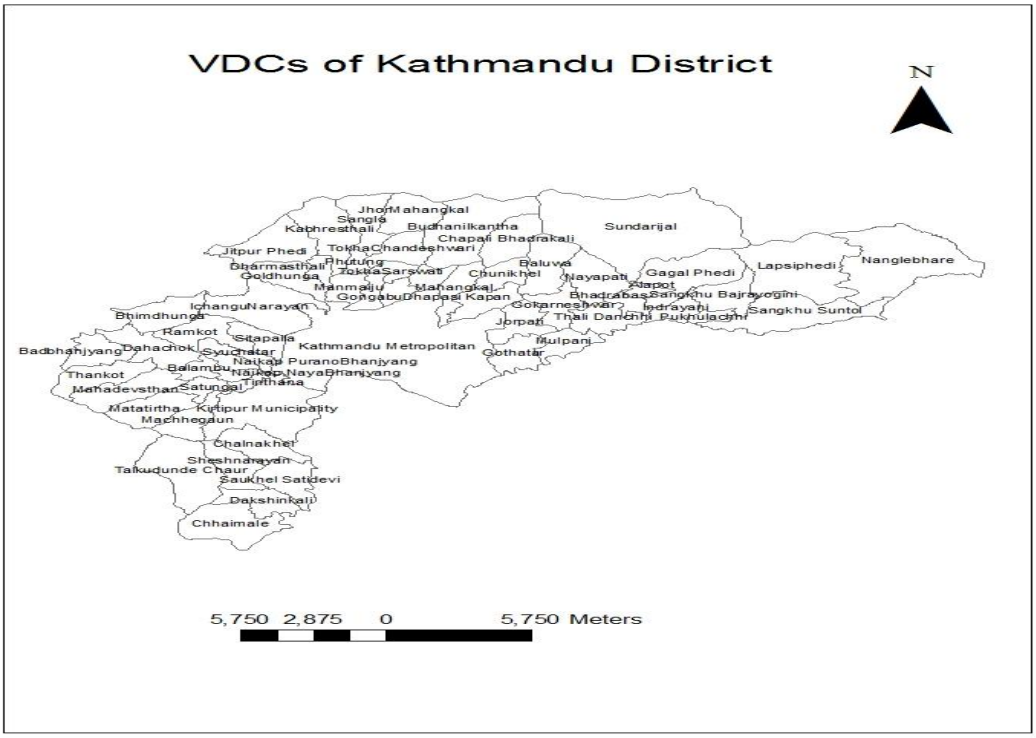


Figure 4:- GIS map of Kathmandu District VDCs.

2.1.1.2 Lalitpur

Lalitpur is a part of Bagmati Zone, is one of the Seventy-Five districts of Nepal, a landlocked countryof South Asia. The district, with Patan as its district headquarters, covers an area of 385 km². It is one of the three districts in the Kathmandu Valley, along with Kathmandu and Bhaktapur. Its population was 466,784 in the initial 2068 census tabulation.

Lalitpur district is a unique amalgamation of culture, history and agro-based economy. Expanding from urban areas to remote southern hills, this administrative district has 41 Villages. Urban areas in Lalitpur have strong trade based economy while majority of the households residing in countryside have their agriculture based economy.

It is the second most important district on the Kathmandu valley. Here also lots of people are living from the outside the district. They are involving with different occupation but due to the less opportunity some people are trying to join into the criminal.

Geographical map of the Lalitpur district is shown below with the all Village Development Committees (VDC):

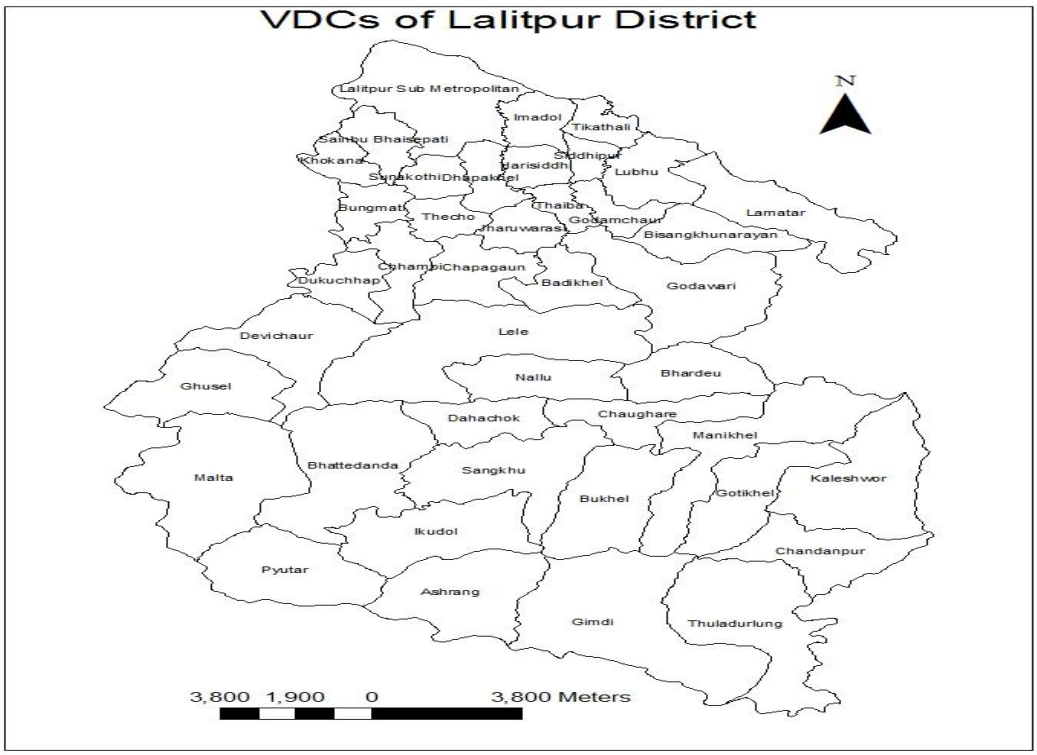


Figure 5:- GIS map of LalitpurDistrict VDCs

2.1.1.3 Bhaktapur

Bhaktapur is located in the eastern part of Kathmandu valley, is the smallest among the seventy-five districts of Nepal. The district, with Bhaktapur as its district headquarters, covers an area of 119 km^2 (46 sq mi)and in 2068 had a population of 304,651.

It is the smallest districts of Nepal. Most of the people of this location are engaged with the agricultural activities. People from outside the district are less. The people crowd is not more than that of Kathmandu and Lalitpur.

Bhaktapur and MadhyapurThimi are the only municipalities in the district, while there are the sixteen VDCs. Geographical map of the Bhaktapur district is shown below with the all Village Development Committees (VDC):

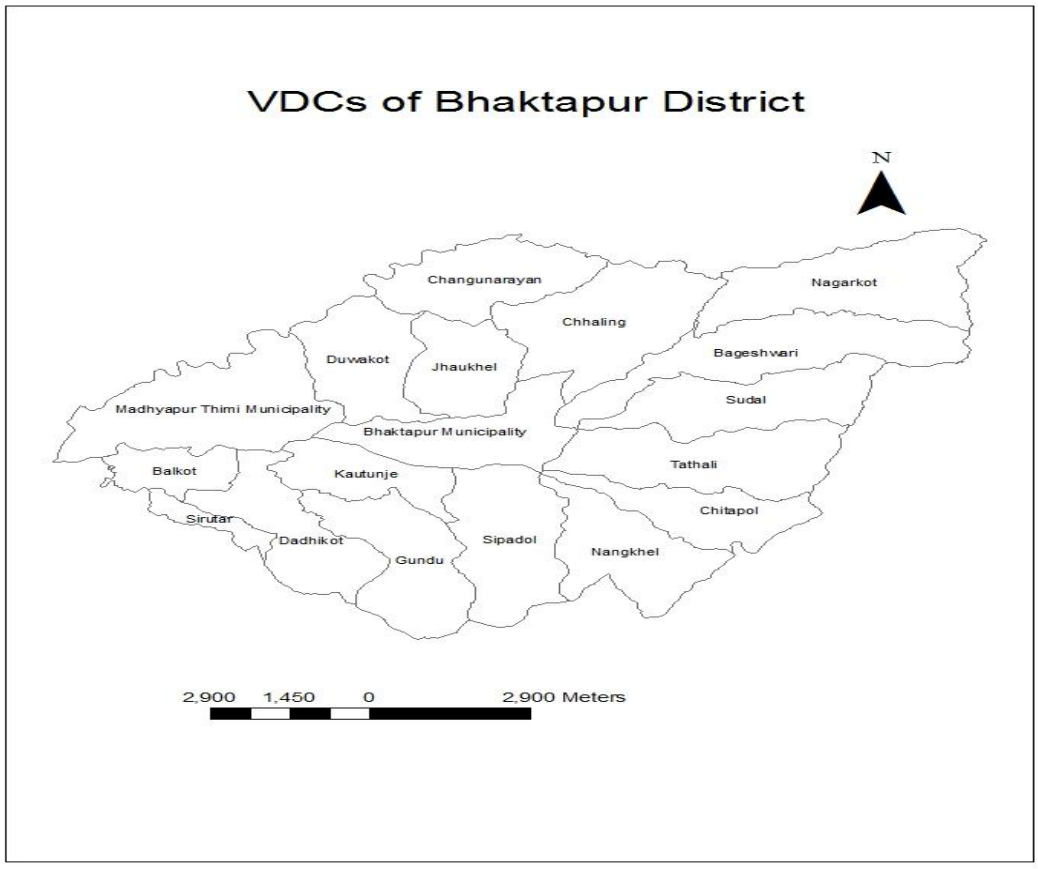


Figure 6:- GIS map of Bhaktupr District VDCs.

2.2Literature Review

Crime is a phenomenon which is universal in its varying forms in all cultures and societies, at all stages of organization. The alarming increase in the rate of criminal activities in Nepal, as reported daily in the local news and media is perhaps are flections of the nature of every society where goals are used to measure individuals status in society.

The Nepalese Police are not equipped with modern automated information system. This is one of the basic problems militating against the effective prevention, detection and control of crime. Even though, there are so many police stations and their outpost distributed around without enough equipment that Police cannot perform their noble role effectively and efficiently except when provided with adequate funding, equipment, infrastructural facilities, social amenities, and manpower. The level of effectiveness of the police in any country depends mainly on the level of manpower and equipment provided. The level of violent crimes in Kathmandu is on the increase looking at the state of the nation.

2.2.1 Criminology and Crime Theories

Traditional criminology depends upon the root causes of crimes.The main areas of criminology are crime patterns, opportunities for committing crime, prevention of victims in criminal activities and environment prone to crime [4]. stated that if an offender is willing to commit a crime; he/she should pass over some physical obstacles. The probability of occurrence of crime increases with the absence of physical barriers preventing crime.

Yusuf [5] describe the distribution of Police stations and Manpower in Kano Metropolis which shows Dala LGA had 2 police station, 4 police outpost, 8 senior officers, 123 junior officers with the population of 418,759 (2006 census) and is the least area with senior police officers, even though is the second most populous local government in the Kano Metropolis. He further explained that the distribution of the police station and manpower in Kano Metropolis are uneven. The population ratio of the police according to United Nations should be 1:450 recommended standards. However, Dambazau explained that there is a general belief that the Nigerian Police has been with a strength of personnel that is far below the capacity required to police the estimated Nigerian population of approximately 120 million, considering the minimum United Nations standard. That is why the study focuses on seeing how GIS would be useful for the Nepalese Police in crime detection, analysis and mapping.

Nowadays crime forecasting is used to predict repeated actions amongst offenders, types and rates of future crimes. Using demographic and economic factors, complex statistical modeling and crime mapping or combination of these are the main methodologies that can be met in the literature for crime forecasting [6].

Hot spot mapping is the simplest way of identifying future crime patterns [7]. Univariate and multivariate are short-term forecasting techniques. Univariate is an explorative forecast model which uses a previous value of one variable to predict crime. Multivariate is a leading indicator

forecast model which uses multiple variables that affect crime to predict crime patterns. The main difference between the two types of model is that extrapolative models can only continue or extrapolate existing crime patterns into the future.

2.2.2 Analysis of crime with GIS

Geographical information system (GIS) is a computer based technology that should be applied by a professional staff to obtain satisfactory results. Appropriate geographical information systems (GIS) background and related statistical software are requirements to utilize the high technology advances in crime.

Continual development in computer technology innovated geographical information systems (GIS) for the studies where the geography should be concerned. Many industries and organizations are the users of GIS. Crime maps started to be created with GIS to archive, manipulate and query the crime data; to update crime patterns; to make spatial analysis and to develop crime prediction and prevention models. Crime mapping plays an important role in pro-active policing and crime prevention in stages of data collection, data evaluation and data analysis. The application areas of crime mapping are recording and mapping crime activities, predicting crime, identifying crime hot spots and patterns, monitoring the impact of crime reduction measures and communicating with stakeholders [8].

2.2.3 Crime Forecasting

Nowadays crime forecasting is used to predict repeated actions amongst offenders, types and rates of future crimes. Using demographic and economic factors, complex statistical modeling and crime mapping or combination of these are the main methodologies that can be met in the literature for crime forecasting [9].

Forecasting techniques are divided into two categories in terms of predicted time period. Crime forecasting includes long-term forecast models for planning and policy applications in broader manner and short-term forecast models for tactical decision making [2].

CHAPTER THREE

DATA AND METHODOLOGY USED IN CRIME PREDICTION MODEL

3.1 Description of Data

Crime data of year 2070 in the study area is used in the analysis which is illustrated in figure. Data is taken from Kathmandu. Spatial and temporal information regarding to these incidents were obtained from Kathmandu Police Directorate. Crime data were recorded by police stations. Data include occurrence time, location and type.

Six types of data are available which are murder, suicide, theft, economical crime, social crime and women & child crime. However, in this study all the crime types are aggregated to have higher number of incidents for constructing reliable prediction model.

There are 6615 incidents recorded in Naxal Headquarter police precincts. Crime incidents are mapped with graduated symbols in which different sizes of features represent particular values of variables [7]. To understand the density of criminal activities in the area the best way is to apply graduated symbols as incidents are overlapping.

Obtaining accurate and reliable crime data that reflect the real incidents is hardly achievable. Staffs in police departments and people exposed to crime are responsible for giving and saving the data. As stated in previous chapter, one of the necessities of computer based crime mapping is qualitative staff. When staff is not expertise in computers and content of the work, errors may exist even in entering data. Data should be properly collected and stored to get more reliable results because analyzing crime incidents is difficult as crime is a mobile phenomenon. In addition to staff, the people are not always informing police when exposed to crime.

Commercial areas are larger in police precinct, including hotels, high schools, primary schools, and car parks. Kathmandu Valley has residential areas with high schools, primary schools, post offices, universities and retail centers. Land-marks were built according to the main structure of the area like hotels are mainly at Kathmandu, whereas schools are located at all over the valley.

There are six different types of crime distributed over the area. Number of each type observed is illustrated in Figure 7, where the social crime has the highest number of incidents. All the incidents per crime type are mapped to demonstrate the spatial distribution in the study area as shown in Figure 9.

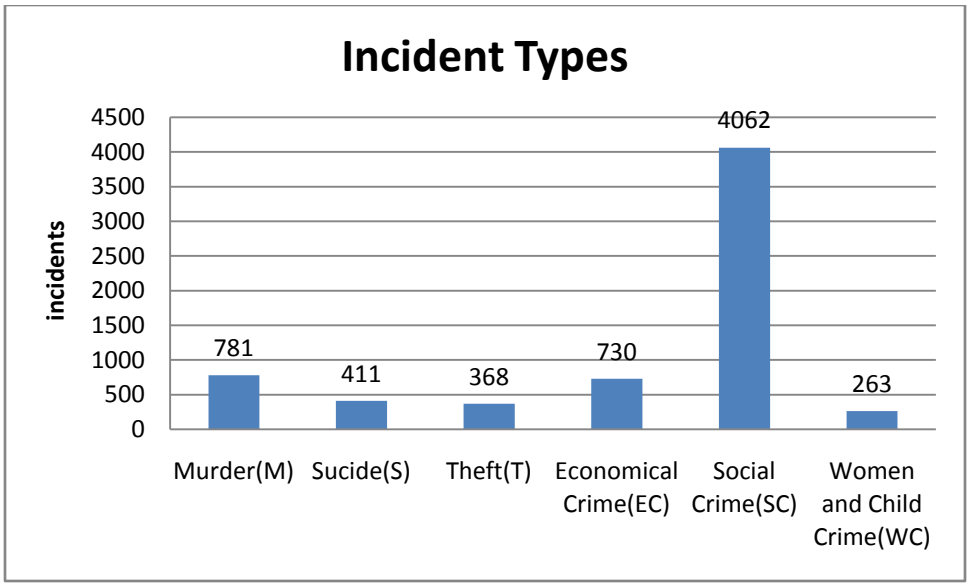
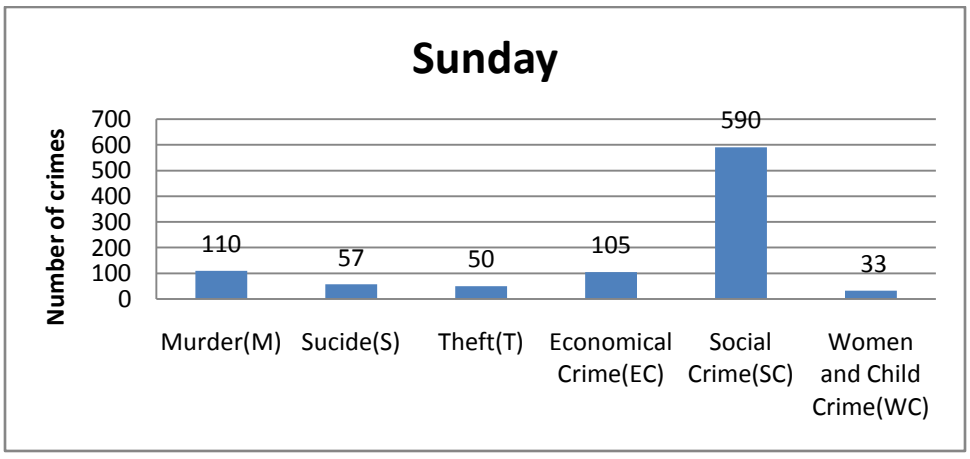


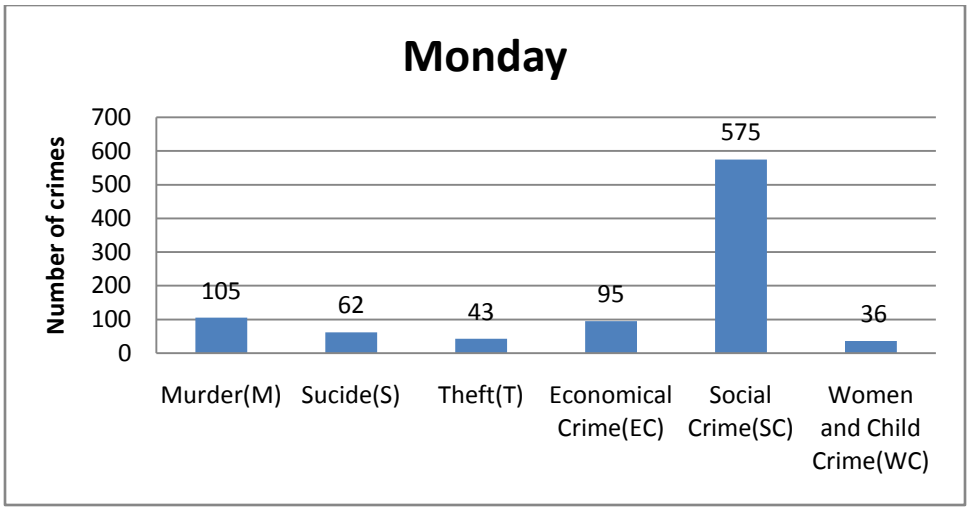
Figure 7:- Number of incidents with respect to crime types in the study area.

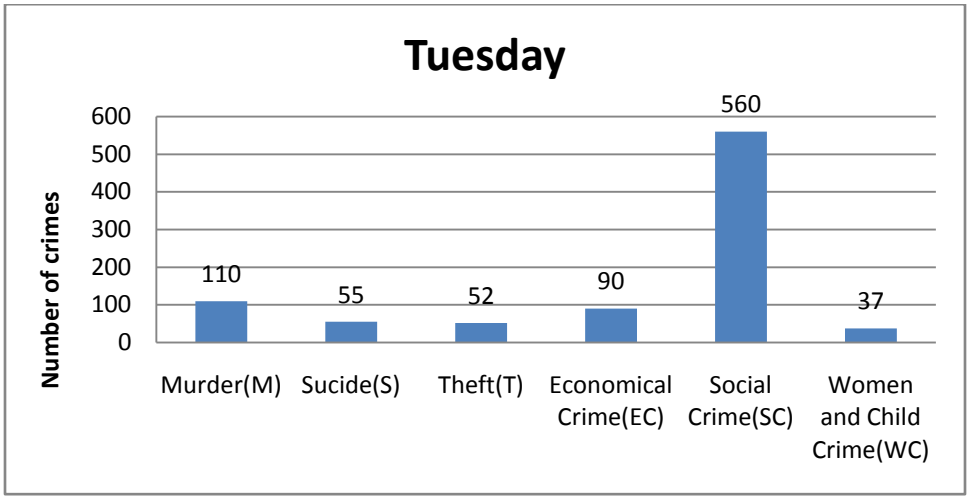
Regarding to Figure 7, the distribution of social crime and economical crime incidents are more randomly than the others. Economical crime incidents are generally observed near the center of the study area. Also, the number of social incidents in Kathmandu is significant.

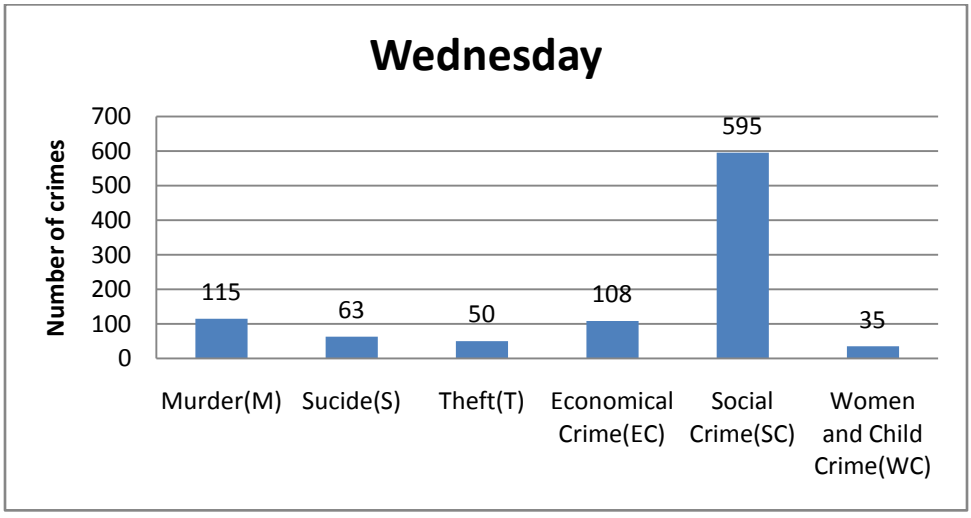
Murder is occurred mostly at Kathmandu region, while there is less theft incidents recorded in Bhaktapur is differing the other crime types in that although the number of incidents for murder and economical crimes are similar, the distribution of incidents are totally different. Most of the economical incidents are located in Kathmandu.

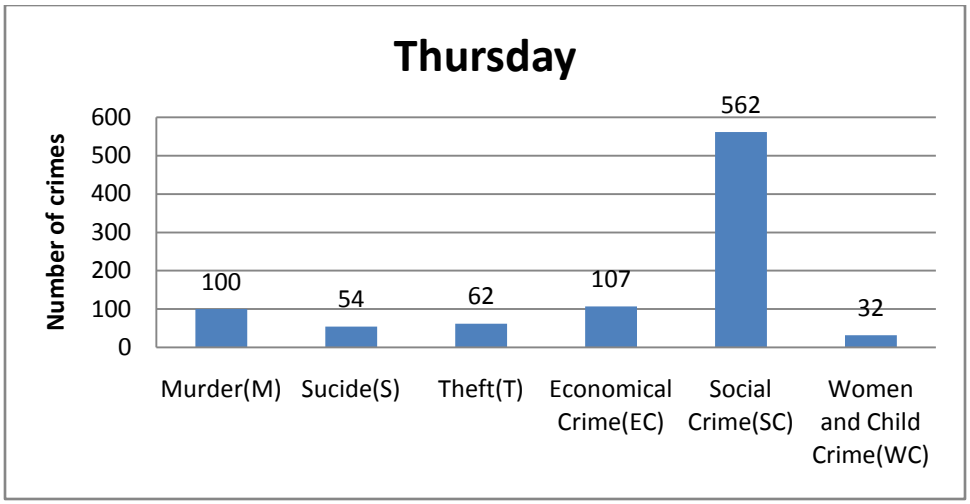
As the crime prediction model is based on weekday values, hence crime incidents for each day are mapped with the number of incidents per crime type per day in Figure 8.

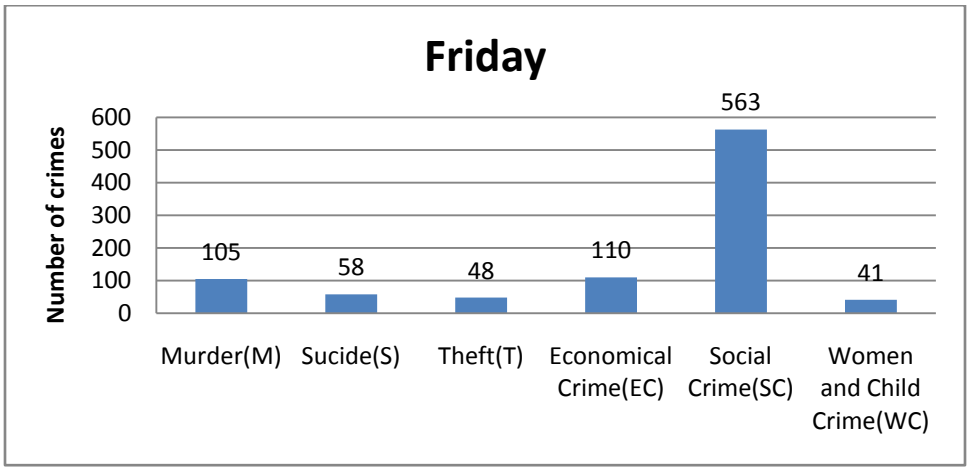



Monday
Number of crimes
700
600
500
400
300
200
100
0
105
62
43
95
575
36
Murder(M)
Sucide(S)
Theft(T)
Economical Crime(EC)
Social Crime(SC)
Women and Child Crime(WC)


Tuesday
Number of crimes
600
500
400
300
200
100
0
110
55
52
90
560
37
Murder(M)
Sucide(S)
Theft(T)
Economical Crime(EC)
Social Crime(SC)
Women and Child Crime(WC)


Wednesday
Number of crimes
700
600
500
400
300
200
100
0
115
63
50
108
595
35
Murder(M)
Sucide(S)
Theft(T)
Economical Crime(EC)
Social Crime(SC)
Women and Child Crime(WC)





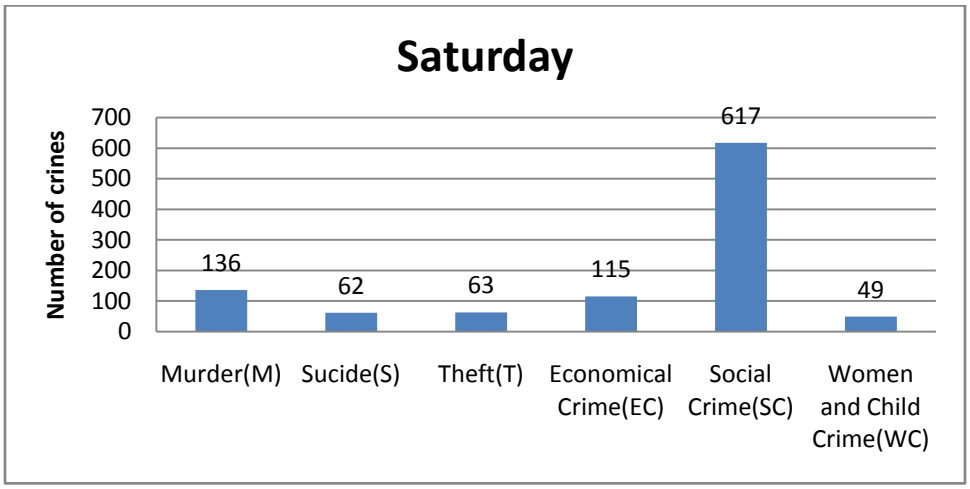


Figure 8:- Number of incidents per crime type per day.

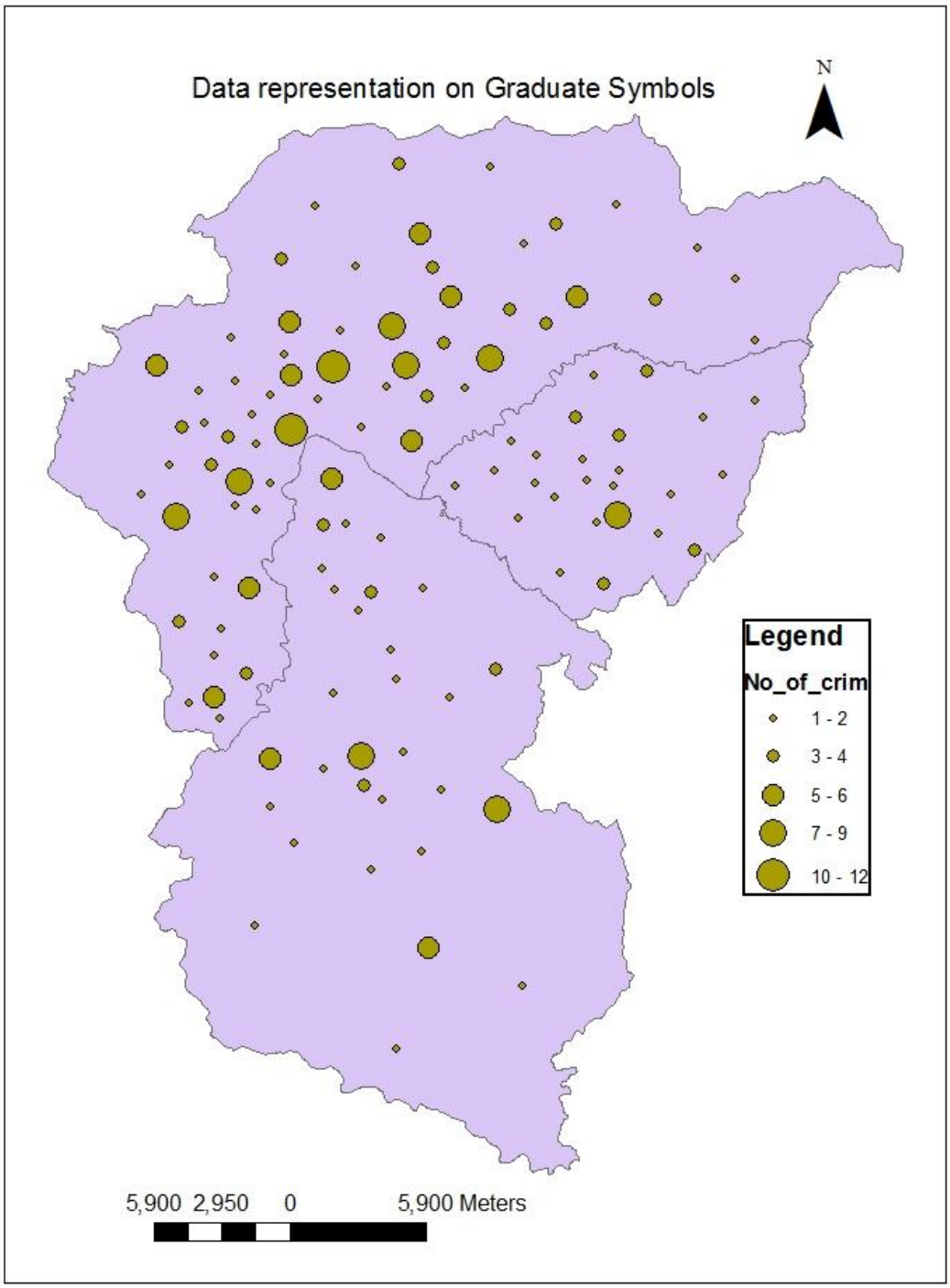


Figure 9:- Spatial Spread of crime in study area

3.2Model Development

The spatio-temporal crime prediction model in this study is adapted from model generated by Al Madfai et al., [3]. There are significant differences in two models as the number of crime incidents and police station used in the adapted study.

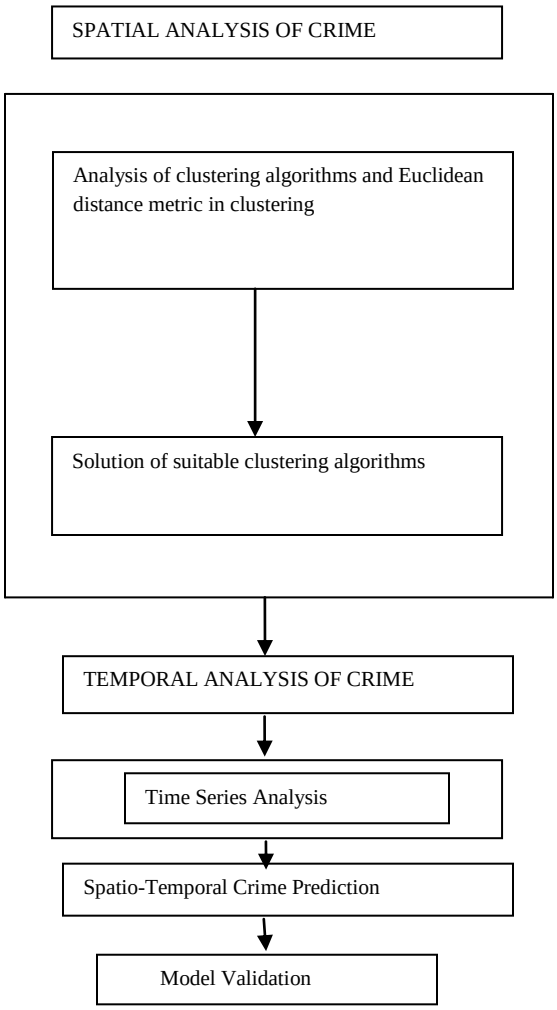


Figure 10:- Spatio-Temporal Crime Prediction Model

The methodology starts with analysis of various clustering algorithms. These algorithms are K-means, Nnh Hierarchical and STAC. Clustering algorithms are analyzed and compared with each other to get the most convenient algorithm to be involved in spatio-temporal crime prediction model.

The Euclidean distance metric, is considered in the scope of the study. This distance metric is applied to seek the difference between the orientations of the clusters. Distance metric is used with the selected clustering algorithm.

In spite of hierarchical profiling approach, time series analysis with ARIMA fitting is used to model the data in this thesis. As a result forecasted values of one year data is obtained.

3.3Algorithm Development

3.3.1 Spatial Analysis

This section covers the various spatial analysis techniques used in this thesis. Mainly three types of cluster analysis are computed here.

3.3.1.1 Cluster Analysis

The classification of objects into different groups sharing the same characteristics is termed as clustering. Clustering is a common technique for data mining, image analysis, biology and machine learning. Techniques which search for separating data in to convenient groups or clusters are termed as clustering analysis.

Cluster analyses can provide significant insight into identifying crime patterns. The technique in crime phenomena is generally used for hot spot detection. Hot spots in crime analysis are one of the major explanations of criminal activities and spatial trends. Clustering is one of the reliable and objective methods to identify hot spots.

Among the various types of clustering algorithms; in this thesis K-means, Nnh hierarchical and STAC are selected to be used. K-means is included in optimization-partitioning techniques, and Nnh hierarchical is a hierarchical clustering analysis technique. STAC is clustering algorithm generated for specific usage purposes. Spatio-temporal analysis of crime was invented by crime analysts to detect crime hot clusters.

3.3.1.1.1 K-Means Clustering

The K-means clustering method is a non-hierarchical clustering approach where data are divided into K groups. User is flexible to decide on the number K. The aim of this technique is to create K number of clusters so that the within group sum of squares are minimized. As iterating all the possible observations is enormous, the algorithm finds a local optimum. To reach the optima, algorithm is repeated several times and the best positioning K centers are found. Then theremaining observations to the nearest cluster to minimize the squared distance [4], are assigned. The formula of the algorithm is:

$$\text{MinV}=\sum_{i=1}\sum_{k} a_i y_{ik} d_{ik}^2 \quad \text{.....Eq(1)}$$

Where;
i = index of observations;
a_i = attribute weight of observation I;
k = index of clusters;
d_{ik} = distance between observation I and cluster k;
y_{ik} = binary value if observation I is assigned in cluster k;

In k-means clustering each observation should be assigned to only one group and all the observations have to be included by clusters. Taking squared Euclidean distance as a distance

measure is important because the model is becoming non-linear and can be solved by appropriate heuristic approaches.

In Crime Stat software the routine starts with an initial guess about the K locations. Initial guess is made by software randomly selecting initial seed locations or by giving the seed point by analyzer. It is an iterative procedure determining the initial seed and assigning the observations to the initial seed and then recalculates the center of the cluster and assigns observations again where the procedure stops [10].
K-means clustering routine outputs clusters graphically as either ellipses or convex hulls. For the standard deviational ellipses 1X, 1.5 X and 2X are the options to select. Here X represents the standard deviation from normal. The level of standard deviation is chosen by the user. Where, by standard deviational ellipses some of the data is abstracted, convex hull represents the cluster by drawing a bounding polygon outside the data [10].

3.3.1.1.2 Nnh Hierarchical Clustering

Levine [10] stated that hierarchical clustering is a clustering method which groups observations on the basis of defined criteria. Criteria are related with type of distance between observations. The clustering is repeated until all the observations are clustered according to the selected criteria and order. Based onthe criterion selected, hierarchical clustering is called nearestneighbor (Nnh). Nnhclustering is an agglomerative procedure, taking the observations individually and forms first order clusters based on a defined threshold distance and minimum number of observations in clusters. If two observations are nearer than the threshold value, a new cluster is generated. The second and higher order clusters are formed with the same manner until only one cluster is left or the threshold criteria fails. Two choices of defining threshold value are, random distance determined by the software itself and fixed distance defined by the user.
The general algorithm of hierarchical clustering is outlined as follows where the distance between two observations i and j is represented by d_{ij}'s and cluster I contains n_i objects. Let D be the remaining d_{ij}'s . Suppose there are N objects to be clustered.

• Find the smallest element remaining in **D.**
• Merge clusters i and j into a single new cluster, k.
• Calculate a new set of distances d_{km} using the following distance formula:

$$d_{km}=\alpha_i d_{im}+\alpha_j d_{jm} + \beta d_{ij}+\gamma |d_{im}-d_{jm}| \quad \text{Eq. (2)}$$

Where;
m represents any cluster other than k. These new distances replaced_{im} and d_{jm} in**D**. Also let, $n_k= n_i + n_j$
Note that the algorithms available represent choices for α_i, α_j, $\beta$and γ.
• Repeat steps 1 - 3 until **D** contains a single group made up off all objects.
This will require *N-1* iterations [11].

Controlling the size of grouping either with threshold distance and minimum number of observations in clusters give chance to identify dense small geographic environments is one of the advantages of hierarchical clustering.

3.3.1.1.3 STAC Clustering

The spatial and temporal analysis of crime (STAC) was developed by the Illinois Justice Information Authority as a clustering algorithm in which the number of points is counted in a circle laid over a grid. The STAC routine in Crime Stat identifies densest clusters by demonstrating either standard deviational ellipses or convex hulls [4].

STAC is adopted as the other clustering routines; however, it differs in the process that the overlapping clusters are combined into larger clusters until there are no overlapping clusters. Both partitioning and hierarchical clustering routines are included in the STAC. Search circles and aggregating small clusters into larger ones are some properties of the other clustering routines.

Algorithm will develop on the basis of cluster analysis of the crime data. There are verities of cluster analysis, where this model will accept the best fitted analysis to generate the algorithm.

3.3.2 Temporal Analysis

Temporal analysis basically consists of ARIMA forecasting. This is described in detail below:

3.3.2.1 Univariate Box-Jenkins(ARIMA) Forecasting

A time series is a set of values observed sequentially at regular intervals of time such as weekly traffic volume, daily crime rates, and monthly milk consumption. The main objective of the time series analysis are to understand the underlying and time-dependent structure of the single series univariate series and to figure out the leading, legging and feedback relationships [12].

Univariate Box-Jenkins is a time series modeling process which describes a single series as a function of its own past values. To find an appropriate equation that reduces a time series with underlying structure to white noise is the aim of the Box-Jenkins process.

Box-Jenkins Analysis refers to a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average (ARIMA) time series models. The method is suitable for time series that have at least 50 observations.

The model is generally referred to as an ARIMA (p,d,q) model where p, d, and q are integers greater than or equal to zero and refer to the order of the autoregressive, integrated, and moving average parts of the model, respectively. When d = 0, the model is turned to be an ARMA (p,q) model.Autoregressive integrated moving average (ARIMA) modeling is formed from two parts: the self-deterministic part and the disturbance component. The self deterministic part can be calculated using autoregressive (AR) model. This takes the past value to predict the future value. Following is the formula for the AR model:

$(1 - \Phi_1 B_1 - \Phi_2 B_2 - \ldots\ldots\ldots\ldots\ldots - \Phi_p B_p)$,

Where $\Phi_1 \ldots\ldots\ldots \Phi_p$ are the parameter values of the polynomial, and B is the backshift operator.

The disturbance component (the residuals from the autoregressive model) is modeled by a moving average (MA) model. Each moving average factor is a polynomial of the form:

$(1 - \theta_1 B_1 - \theta_2 B_2 - \ldots\ldots\ldots\ldots\ldots - \theta_q B_q)$,

Where θ_1 …….. θ_q are the parameter values of the polynomial and B is the backshift operator.

The backshift operator is a special notation used to simplify the representation of lag values. $B_J X_t$ is defined to be X_{t-J}. So, $(B_1)X_t = X_{t-1}$ which means a 1 period lag of X[13]. Hence the ARIMA (p, d, q) model is:

$\Phi(B)(1 - B)^d Y_t = \theta(B)\, \varepsilon_t$,…… Eq. (3)

Where ε_t is an error term generally assumed to be independent, identically distributed samples from a normal distribution, d is a positive integer that controls the level of [12].

CHAPTER FOUR
COMPARISON OF DIFFERENT CLUSTERING METHODS AND GENERATION OF HOTSPOTS IN THE STUDY AREA

In this chapter, the aim is to decide the suitable clustering technique to be used in spatio-temporal crime prediction model.Predicting crime through hot spotting is a new advance to police departments to make tactical, strategic and administrative policies and to get right prevention measures. Clustering is gaining importance as it is a reliable way of determining crime hot spots.

In order to make comparison between the clustering models, different clustering methods are applied to the study area.Both hierarchical and non-hierarchical/partitioning approaches are considered. K-means clustering algorithm is applied as partitioning based clustering approaches. This method is including optimization procedures to get final configurations. Hierarchical clustering algorithm represents the hierarchical approach while spatio-temporal analysis of crime and geographical analysis are generated specifically for cluster detection. CrimeStat 3.3, is run to carry out analysis and results are interpreted with ArcGIS 9.3and GISsoftwares.

4.1 Implementation of K-Means Clustering

"K" is the key part of K-means clustering, which represents the number of clusters.It is not easy to determine the number of clusters in K-means clustering. Freedom of defining the number of clusters can be an advantage or disadvantage according to the purpose of usage.

In order to determine the number of clusters, different K values are examined to determine the best configuration. CrimeStat 3.3 and ArcGIS 9.3 are employed in order to apply K-means clustering to the crime incident data in the study area.

Visualization of the results is another concern deciding the number of clusters. Two techniques are available in CrimeStat 3.3: standard deviational ellipses and convex hulls. In standard deviational ellipses, it is optional to decide on the size of the ellipses which are 1X, 1.5X and 2X. Here X represents the amount of standard deviation in algorithm. 1X is generally preferred as the other options gave an exaggerated view of the underlying clusters. For K=6, standard deviational ellipses are overlaid to indicate the clusters Figure 11.

K-Means Clustering:

Clusters .....: 6
Measurement type ..: Direct
Standard Deviations: 1.0
Start time .....: 12:35:31 PM, 04/16/2015

Iterations .....: 50

End time .....: 12:35:40 PM, 04/16/2015

Cluster	X-Axis(mi)	Y-Axis(mi.)	Area(sq mi.)	MeanSquareError	Points
1	0.57675	0.41470	0.75141	1.47945	22
2	0.57335	0.50145	0.90323	3.82697	26
3	1.09521	0.64487	2.21882	6.62296	20
4	0.53137	0.74941	1.25104	3.51658	15
5	0.61941	0.68761	1.33805	10.92010	20
6	0.71801	0.45185	1.01922	2.46358	13

Table 1:- Result of K-Means clustering

Total Area covered by a clusters =7.48177(sq mi.)

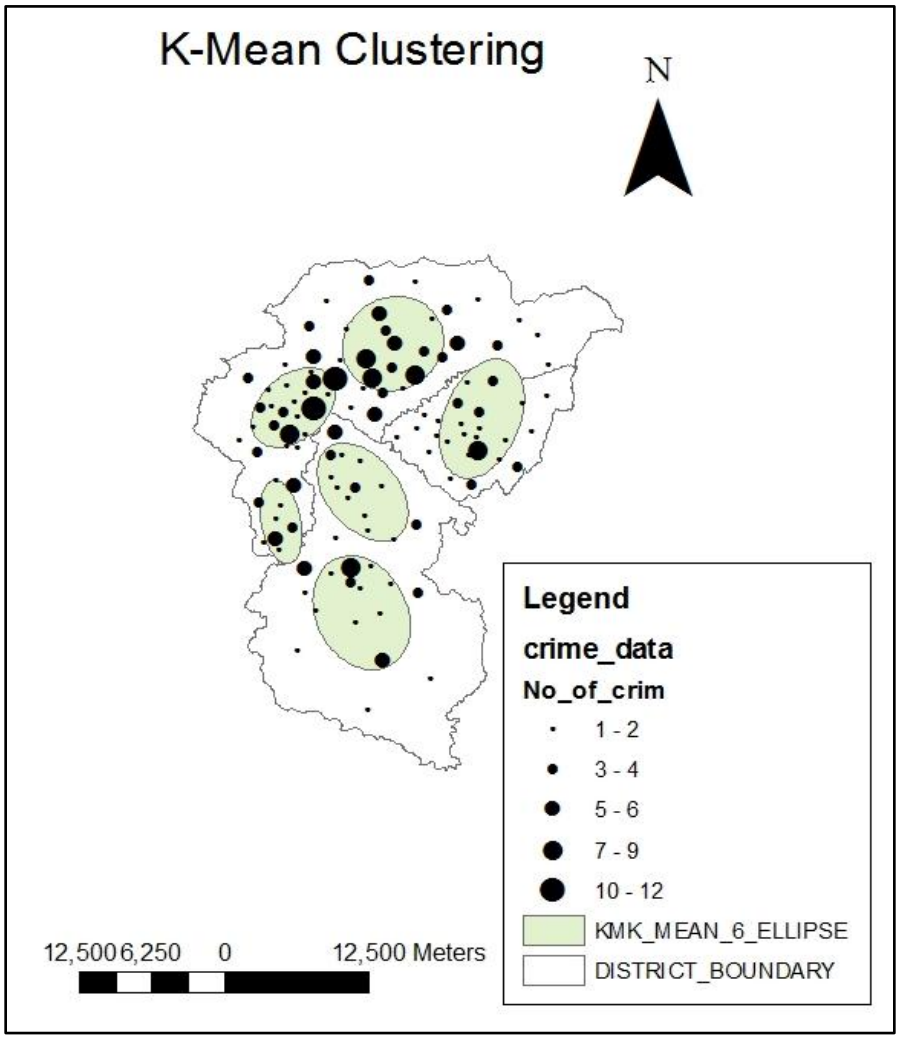


Figure 11:- K-mean clustering ellipse

4.2 Implementation of Nnh Hierarchical Clustering

Nnh clustering is an agglomerative procedure, taking the observations individually and forms first order clusters based on a defined threshold distance and minimum number of observations in clusters. If two observations are nearer than the threshold value, a new cluster is generated.

Two choices of defining threshold value are, random distance determined by the software itself and fixed distance defined by the user. When random distance option is selected, too many clusters are generated, so different fixed distance options are tried to get the best configuration. Levine [10] suggested to take the threshold value 0.5 miles or smaller to get feasible results. Also, after interpreting lots of "minimum number of points", 5 are selected visually. The resulting maps including first order clusters are shown in Figure 12.

Nearest Neighbor Hierarchical Clustering:

Likelihood of groupingpair of points by chance...: 0.50000 (50.000%)
Z-value for confidence interval.....: 0.000
Measurement type.....: Direct
Output units.....: Miles, Square Miles, Points per Square Miles
Standard Deviations .....: 1.0
Clusters found.....: 6
Simulation runs.....: 10

Displaying 6 ellipse(s) starting from 1

Cluster	X-Axis(mi)	Y-Axis(mi.)	Area (sq mi.)	Points	Density
1	0.51994	0.89641	1.46424	35	23.903262
2	1.05369	0.61470	2.03482	27	13.269019
3	0.47080	0.69706	1.03098	14	13.579267
4	0.71351	0.46213	1.03590	10	9.653452
5	0.21379	1.07560	0.72243	5	6.921098
6	0.15594	0.30592	0.14987	9	60.053105

Table 2:- Result of Nnh-Hierarchical clustering

Total Area covered by a clusters =6.4384 (sq mi.)

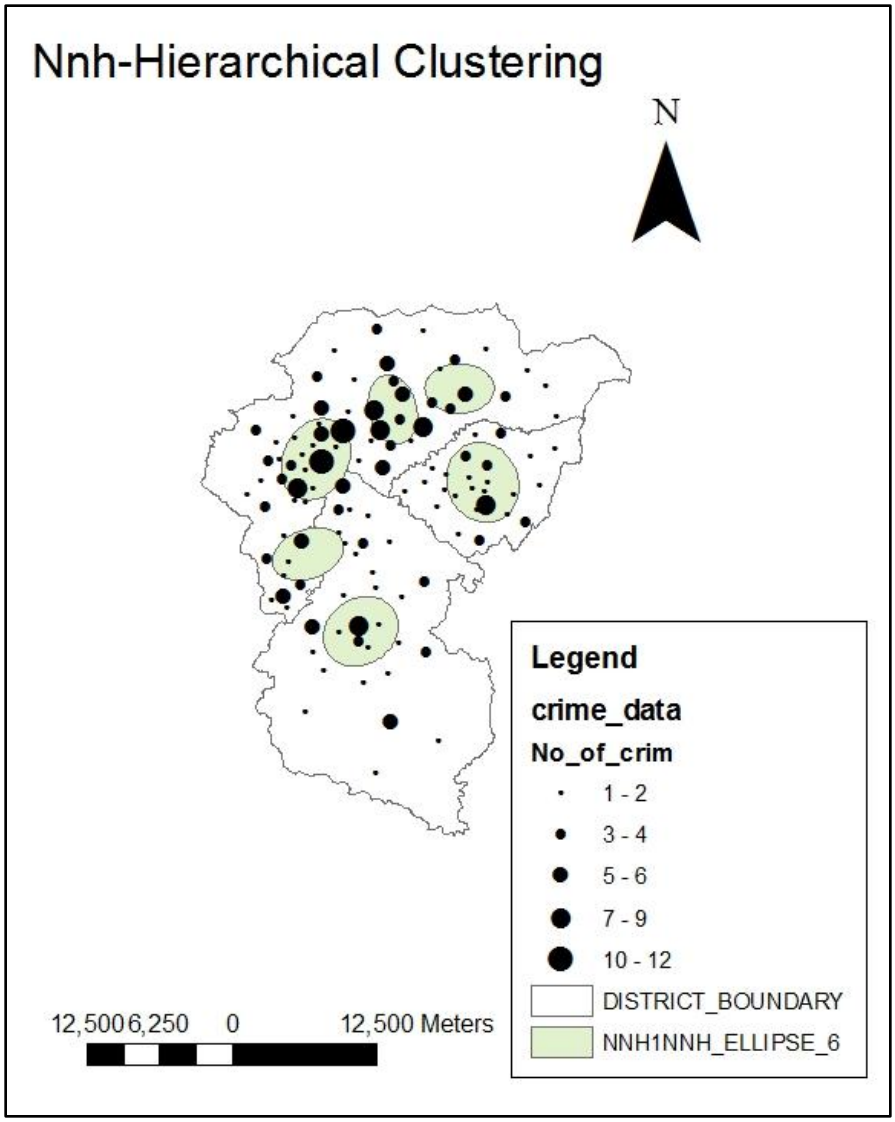


Figure 12:-Nnh-Hierarchical clustering ellipse

4.3 Implementation of STAC Clustering

STAC is another crime hot spot program which is quick, visual and easy to use [10]. STAC identifies the major concentrations of points for a given distribution. Circles are drawn and overlaid for points in a defined grid.

Several "minimum number of points" is tried for each fixed distance and 5 are selected to be used in the analysis. Fixed distance and minimum number 5 are an effective when compared to the others selected.

Spatial and Temporal Analysis of Crime:

Measurement type .....: Direct
Scan type... Rectangular
Input units... Feet
Output units ... Miles, Square Miles, Points per Square Miles
Standard Deviations ...: 1.0
Start time .....: 12:32:59 PM, 04/16/2015
Search radius.....: 1968.503937
Boundary.....: 621554.98240, 3037946.14777 to 648163.82158, 3076324.41343
Points inside boundary: 114

End time.....: 12:32:59 PM, 04/16/2015

Cluster	X-Axis(mi)	Y-Axis(mi.)	Area (sq mi.)	Points	Density
1	0.42827	0.37866	0.50947	17	33.368150
2	0.46522	0.19704	0.28799	11	38.196109
3	0.46966	0.07679	0.11330	8	70.607118
4	0.16236	0.28348	0.14459	6	41.496894
5	0.46450	0.13408	0.19566	6	30.665605
6	0.50571	0.25472	0.40468	5	12.355416

Table 3:-Result of STAC clustering

Total Area covered by a clusters =1.65559 (sq mi.)

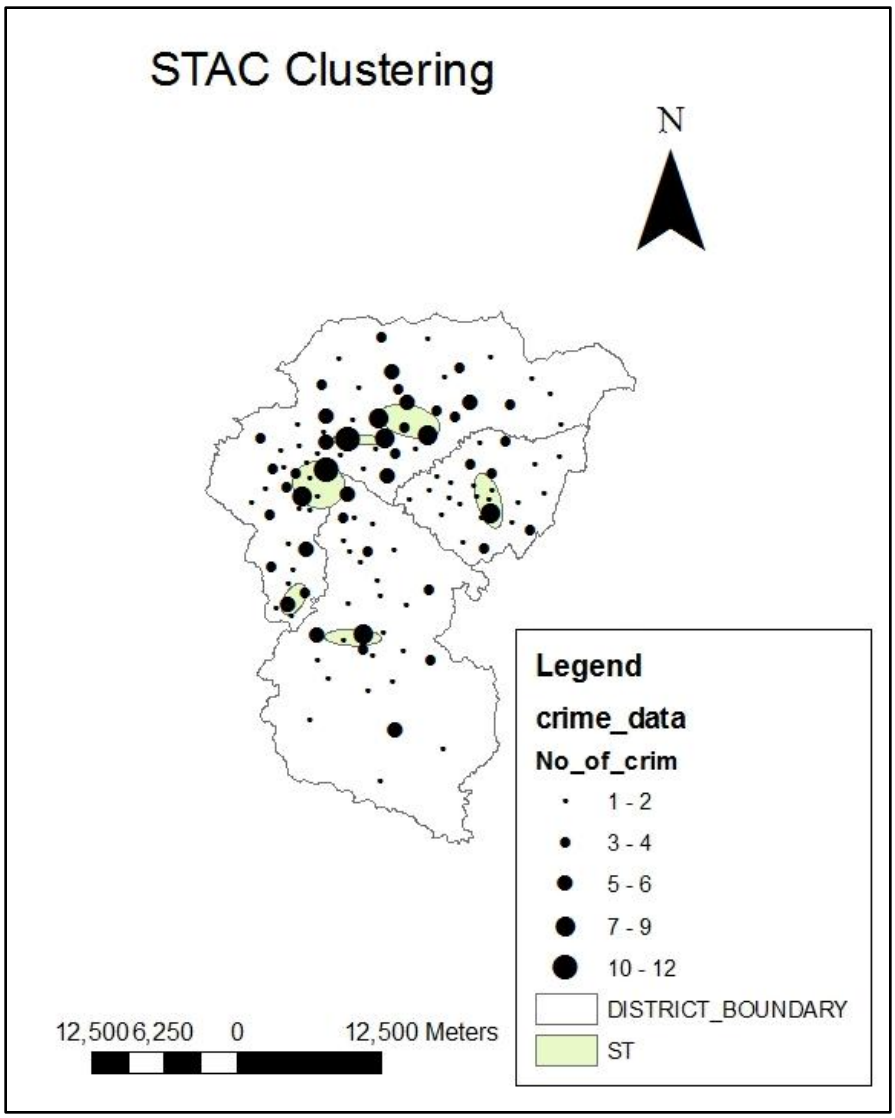


Figure 13:- STAC clustering ellipse

4.4 Comparison of the clustering methods

The aim to compare the clustering methods stated in this chapter is to choose the most appropriate for the spatio-temporal crime prediction model.Firstly, to make a general comparison between the clustering methods in the study area, ellipse maps are represented by Figure 11, 12 and 13. Briefly, K-means clustering methods is type of partitioning approach and cover all the observations in the area. Nnh hierarchical clustering is distance specific hierarchical approach and STAC is combination of two approaches, partitioning approach with search circles and hierarchical approach with aggregating smaller clusters into larger clusters.

From the above ellipse clusters we saw that K-Means clustering covers the more and appropriate region of the study area. Nnh-Hierarchical clustering also tries to fit on the data even.It is better than STAC clustering but not able to cover more region than K-Means clustering.

S.N	Clusters	Area Covered (sq mi.)
1.	K-Means Clustering	7.48177
2.	N-nh Hierarchical Clustering	6.4384
3.	STAC Clustering	1.65559

Table 4:- Area covered by each clusters

According to these discussions and figure of area covered, K-mean is selected to be used in the crime prediction model for several reasons:

1. Clusters of K-mean do include more homogenous areas than the other methods. The biggest cluster in Kathmandu, almost cover all the commercial area in Kathmandu region. Homogeneity of land use in the clusters is an advantage as the crime incidents happened is more typical.
2. Actually, it is an advantage in crime prediction as when number and place of crime incidents are forecasted also crime types will be predicted. Police for example, use the advantage to control the similar areas.
3. It has covered more area than other clusters and also computation efficiency is faster.

Many combinations of K-Means are tried. Crime prediction model is daily based model. Hence, data is divided into seven weekdays and clusters are generated foreach day. In crime prediction model, areas of clusters should not be so small or solarge to get more meaningful results. Area of cluster in Table 1 shows the detail of K-Means clustering and also Table 4 shows the comparison of area covered by all clustering algorithms.

4.5 Selection of Hotspot

Hotspot is the crime forecasting method in spatial data analysis. This shows the more victims or prone locations of crime. Hotspot is used to demonstrate the crime affected area by the police department. It is very useful to the police to disseminate the information about the crime and make the people aware and beware from crime.

K-means covers the larger number of crimes and also homogeneous areas than other clusters.And hence, we decided that the best hotspot is generated by using K-means algorithm. In this way we selected the K-means hotspots for this thesis analysis.

CHAPTER FIVE

SPATIO-TEMPORAL CRIME PREDICTION MODEL WITH ARIMA MODEL FITTING AND FORECASTING

In this chapter, a spatio-temporal crime prediction model is generated with ARIMA forecasting approach. A Box-Jenkins ARIMA model is commonly used in several sciences. The ARIMA model has four step iteration; identification, estimation, diagnostic checking and forecasting.

To predict the future values, Box-Jenkins ARIMA model is fit to daily data for the year 2070BS. All the steps are evaluated iteratively and forecasted values are gained. Minitab and Xlstat are employed during these processes and Microsoft Excel is used in statistical calculations.

5.1 Fitting Box-Jenkins ARIMA model

The first stage in the Box-Jenkins model is the identification stage. In order totentatively identify a model, first whether the time series is stationary or notshould be determined. A time series is stationary if the statistical properties likemean and variance are essentially constant over time [14]. The simplest way to understand this is to plot the values against time. If thevalues seem to fluctuate with constant variation around a constant mean, it isreasonable to believe that the time series is stationary. Plotting the number ofincidents of each day against time, time series plot of number of incidents in Figure 14 is gained. Although having some outliers especially in the second halfof the year, the graph seems stationary.

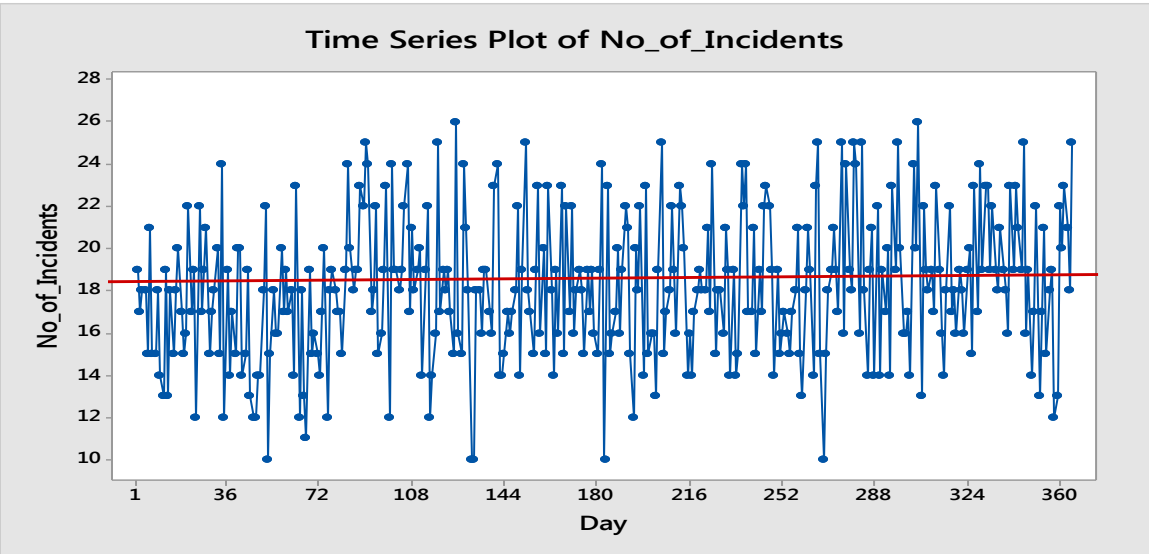


Figure 14:- Time series plot of crime

To detect the mean and variance movements, both of them are plotted with dividing the data into 8 lags. Movement of mean in 8 lags is not so volatile; the values are between 17.5 and 18.6 until the 8th lag is shown in Figure 15. Hence, the variation of the mean is not significant. Also, looking at the autocorrelation plot, stationarity can be evaluated.

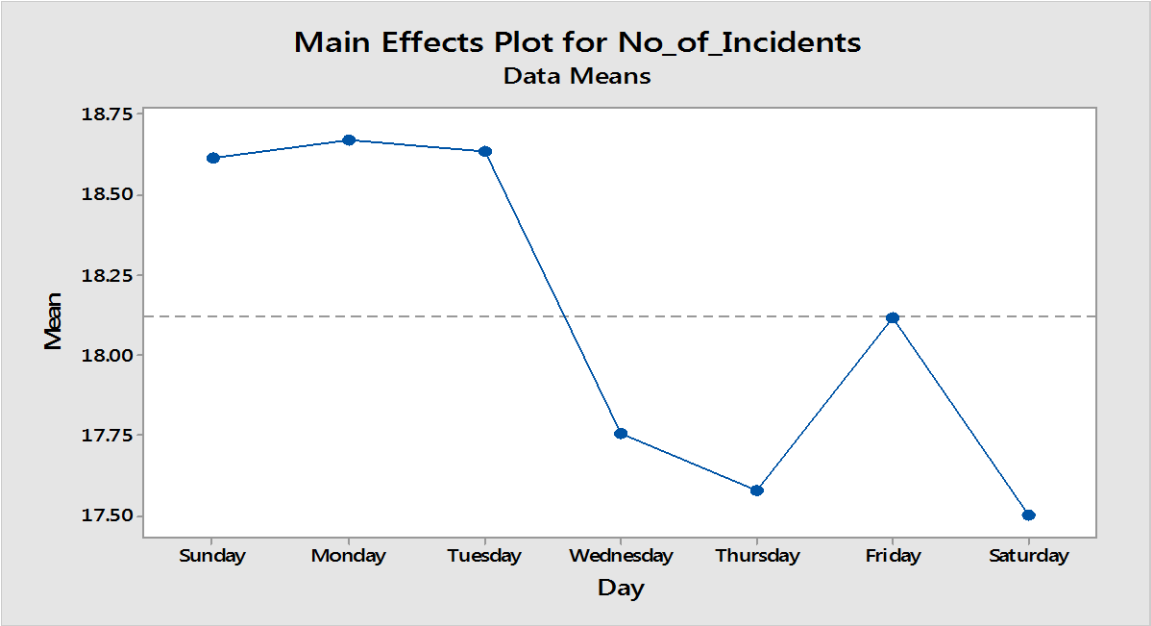


Figure 15:- Movement of Mean in 8 lags

To seek the movement, variation of variance is plotted for the time series data. Mean is 18.12 and maximum deviation of the mean is 0.62, which does not seem significant. Histogram of the data indicates that the data seem normally distributed and shown in Figure 16.

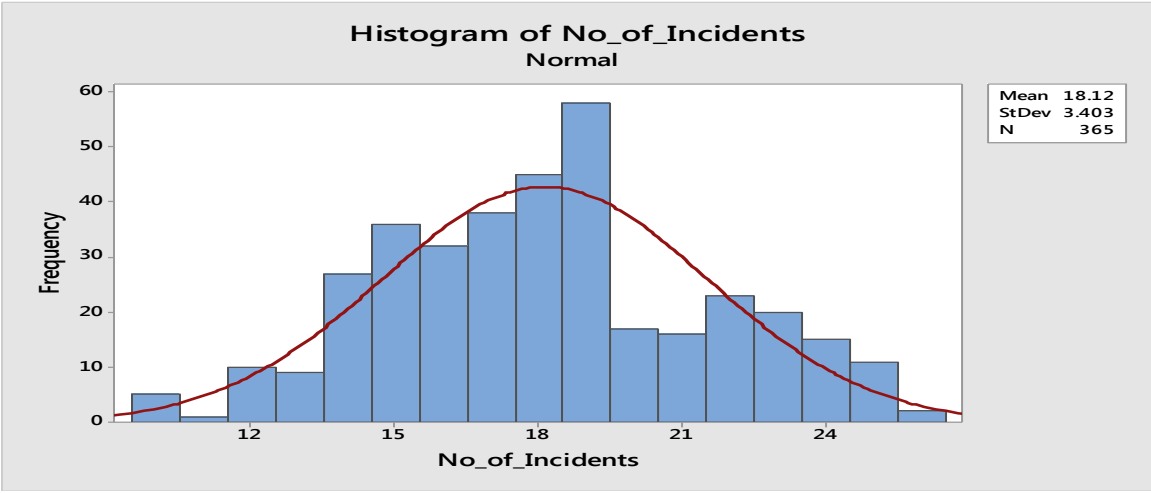


Figure 16:- Histogram of Incidents

After, confirming the stationarity of the data, the next step is to determine the levels of AR and MA values. As there is no need to difference the data, the model is turned to an ARMA model as I represent the amount of differencing.For the decision of the levels, autocorrelogram and partial autocorrelogram are plotted and as it is an iterative process, different combinations are tried to get the best result. As seen obviously from the Figure 17 that the lags are significant when the lag values pass the red line. Red line indicates the 5% significance level of autocorrelations.

Bowerman and O'Connell [14] stated that for lower lags (lag < 3); the spike exists if t value is greater than 1.6 and for higher lags, a spike is considered to exist if t is greater than 2. According to this statement, it is convenient to say according to Table 5 that lag 4 is significant.

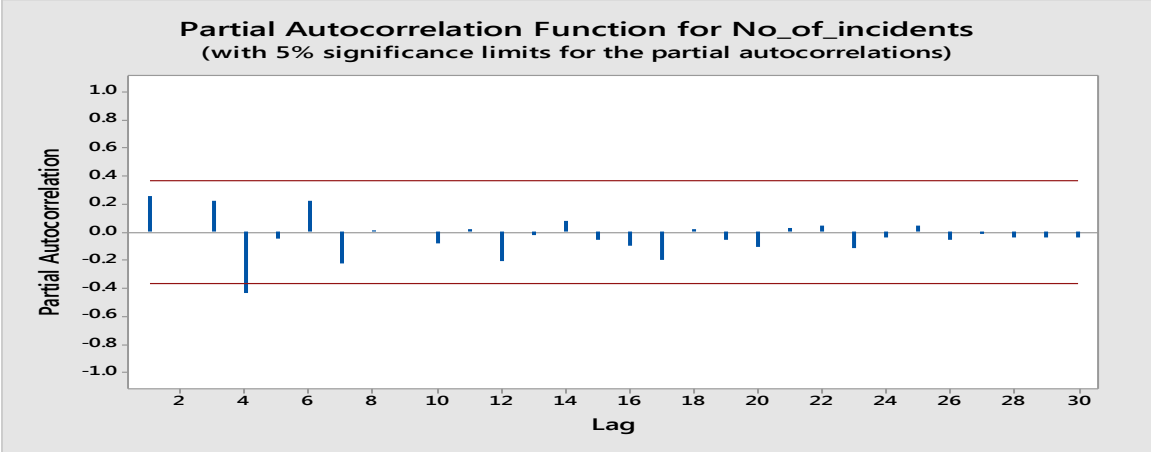


Figure 17:- Spike existing on 4th lag of partial autocorrelation plot

Lag	PACF	T
1	0.261950	**1.46**
2	0.000742	0.00
3	0.225874	1.26
4	-0.434599	**-2.42**
5	-0.043234	-0.24
6	0.224816	1.25
7	-0.220651	-1.23
8	0.010344	0.06
9	-0.006738	-0.04
10	-0.078711	-0.44
11	0.020948	0.12
12	-0.211131	-1.18
13	-0.017987	-0.10
14	0.077019	0.43
15	-0.056355	-0.31
16	-0.097327	-0.54
17	-0.196819	-1.10
18	0.017902	0.10
19	-0.057930	-0.32
20	-0.105809	-0.59
21	0.030675	0.17

22	0.042710	0.24
23	-0.115120	-0.64
24	-0.039276	-0.22
25	0.046450	0.26
26	-0.058825	-0.33
27	-0.012964	-0.07
28	-0.041005	-0.23
29	-0.036136	-0.20
30	-0.034842	-0.19

Table 5:-PACF and t-values of incidents

After detecting spike existing in graph, which will guide in trial period, several combinations of AR and MA levels is going to be evaluated to get the best result. As spike is detected at lag 4 for partial autocorrelation, the trial starts from AR(4) and MA(4).

At last in diagnostic checking part, modified Box-Pierce (Ljung-Box) Chi-Square statistic is going to be evaluated to analyze the residuals obtained from the model. If the probability value is near to the value 1, it is reasonable to say that the model is adequate [14]. Also, the adequacy of the model should be supported with normal probability plot and the autocorrelation and partial autocorrelation of the residuals. Final estimated parameters is shown in Table 6 and Box-Pierce Chi-square statistic is shown in Table 7 below.

Type	**Coef**	**SE Coef**	**t**	**P**
SAR 12	0.9830	0.0260	37.82	0.000
SMA 12	0.9320	0.0447	20.85	0.000
Constant	0.31304	0.01822	17.18	0.000
Mean	18.463			

Table 6:- Final estimates of parameters

Lag	12	24	36	48
Chi-Square	12.9	24.1	32.6	43.0
DF	9	21	33	45
P-Value	0.166	0.828	0.849	0.955

Table 7:- Modified Box-Pierce (Ljung-Box) Chi-Square statistic

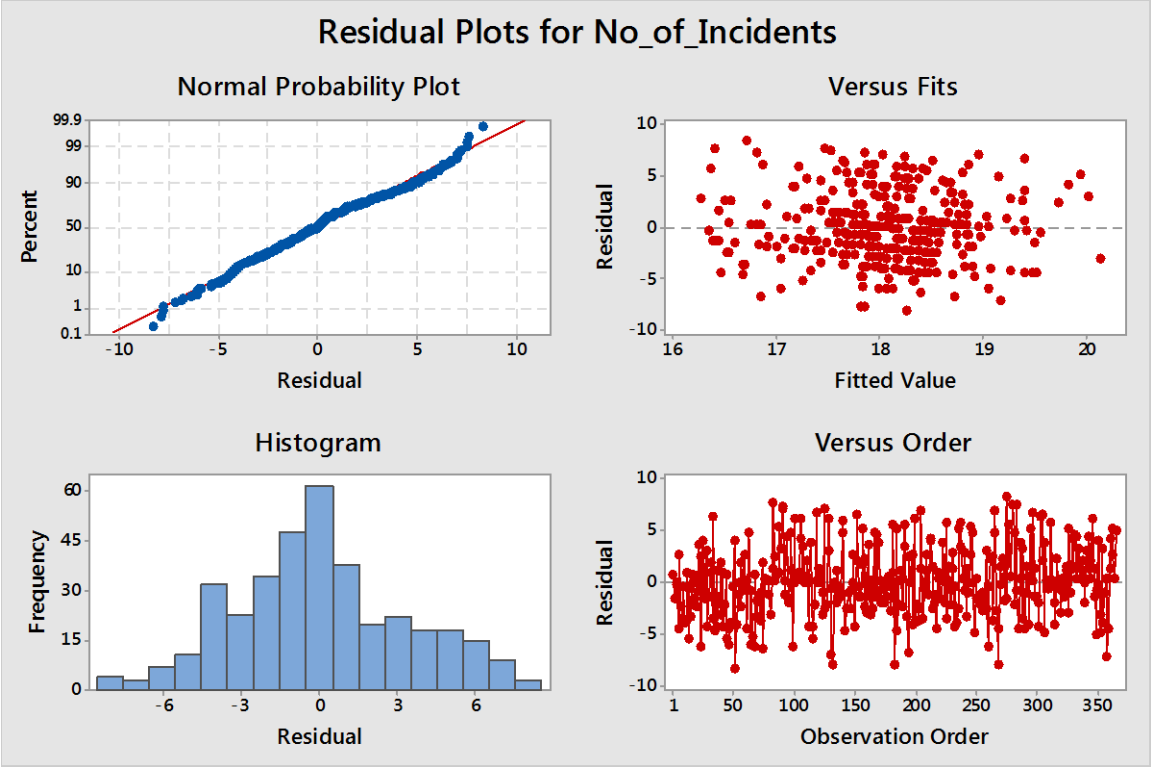


Figure 18:- Residual plots of AR (1) and MA (1)

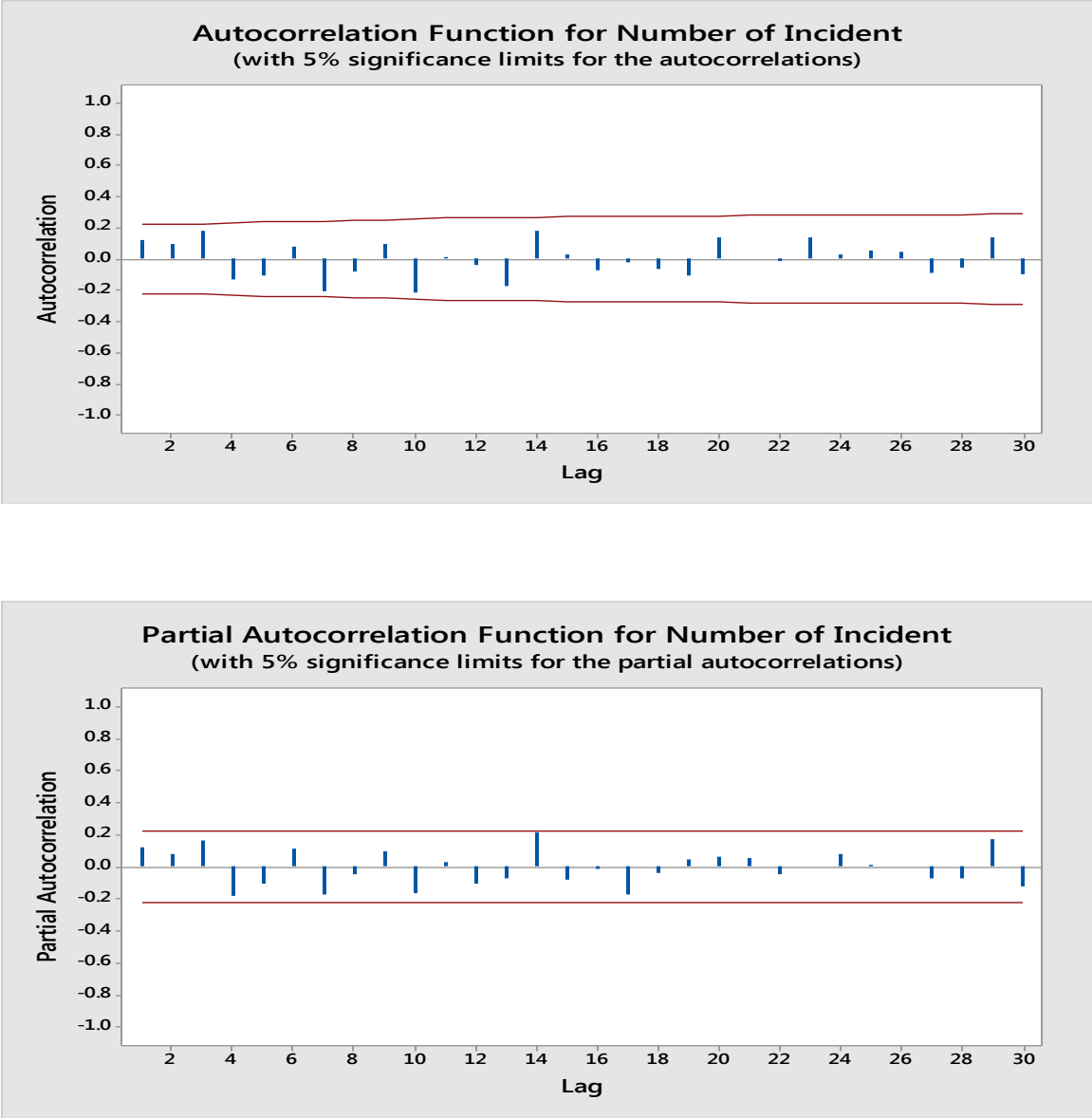


Figure 19:- ACF and PACF plots of Residue

Forecasting results are obtained for all original values and future values.However, the model gives all the future values approximating the mean value as 18.463. An example of forecasted values and their residuals of Baishak2070 BS are given in Table 8.

Month	Day	Number of Incidents	Forecasted Values	Residuals
1.Chaitra 2070	Saturday	18	18.02767637	-0.0276763
2.Chaitra 2070	Sunday	21	18.0092454	2.9907546
3.Chaitra 2070	Monday	19	18.2705349	0.7294651
4.Chaitra 2070	Tuesday	18	18.08644574	-0.0864457
5.Chaitra 2070	Wednesday	16	18.05773216	-2.0577321
6.Chaitra 2070	Thursday	23	18.09018837	4.90981163
7.Chaitra 2070	Friday	19	17.97483859	1.02516141
8.Chaitra 2070	Saturday	23	18.03318806	4.96681194
9.Chaitra 2070	Sunday	21	18.01556192	2.98443808
10.Chaitra 2070	Monday	19	18.26544158	0.73455842
11.Chaitra 2070	Tuesday	25	18.08939113	6.91060887
12.Chaitra 2070	Wednesday	16	18.06193139	-2.0619313
13.Chaitra 2070	Thursday	19	18.09297032	0.90702968
14.Chaitra 2070	Friday	14	17.98265757	-3.9826575
15.Chaitra 2070	Saturday	17	18.03845907	-1.0384590
16.Chaitra 2070	Sunday	22	18.02160262	3.97839738
17.Chaitra 2070	Monday	13	18.26057066	-5.2605706
18.Chaitra 2070	Tuesday	17	18.09220789	-1.0922078
19.Chaitra 2070	Wednesday	21	18.06594725	2.93405275
20.Chaitra 2070	Thursday	15	18.09563079	-3.0956307
21.Chaitra 2070	Friday	18	17.99013512	0.00986488
22.Chaitra 2070	Saturday	19	18.0434999	0.9565001
23.Chaitra 2070	Sunday	12	18.02737954	-6.0273795

24.Chaitra 2070	Monday	13	18.25591245	-5.2559124
25.Chaitra 2070	Tuesday	22	18.09490165	3.90509835
26.Chaitra 2070	Wednesday	20	18.06978774	1.93021226
27.Chaitra 2070	Thursday	23	18.09817509	4.90182491
28.Chaitra 2070	Friday	21	17.99728614	3.00271386
29.Chaitra 2070	Saturday	18	18.04832062	-0.0483206
30.Chaitra 2070	Sunday	25	18.03290419	6.96709581

Table 8:- Forecasted and Residuals of crime incidents for month of Chaitra 2070BS

5.2 Forecasting of future data

Future values are forecasted using the AR (1) and MA (1) model. The future values for Baishak 2071 are shown in Table 9 below:

Month	**Day**	**Forecasted Values**
1.Baishak 2071	Monday	17.9218
2.Baishak 2071	Tuesday	18.6328
3.Baishak 2071	Wednesday	18.9170
4.Baishak 2071	Thursday	18.6388
5.Baishak 2071	Friday	18.7939
6.Baishak 2071	Saturday	17.1369
7.Baishak 2071	Sunday	17.9755
8.Baishak 2071	Monday	18.7344
9.Baishak 2071	Tuesday	18.0836
10.Baishak 2071	Wednesday	18.4204

11.Baishak 2071	Thursday	17.6381
12.Baishak 2071	Friday	20.1864
13.Baishak 2071	Saturday	17.9310
14.Baishak 2071	Sunday	18.6299
15.Baishak 2071	Monday	18.9093
16.Baishak 2071	Tuesday	18.6358
17.Baishak 2071	Wednesday	18.7882
18.Baishak 2071	Thursday	17.1594
19.Baishak 2071	Friday	17.9838
20.Baishak 2071	Saturday	18.7298
21.Baishak 2071	Sunday	18.0900
22.Baishak 2071	Monday	18.4212
23.Baishak 2071	Tuesday	17.6521
24.Baishak 2071	Wednesday	20.1572
25.Baishak 2071	Thursday	17.9401
26.Baishak 2071	Friday	18.6271
27.Baishak 2071	Saturday	18.9017
28.Baishak 2071	Sunday	18.6329
29.Baishak 2071	Monday	18.7827
30.Baishak 2071	Tuesday	17.1815
31.Baishak 2071	Wednesday	17.9218

Table 9:- Forecasted crime values of Baishak 2071BS

5.3 Model validation

Model is validated by calculating the deviation between original data and forecasted data. For this purpose we have taken Chaitra 2070BS data. Error on the data due to forecasting is calculated below.

Spatio-Temporal Error (STE) = $\frac{Actual\ \ value - Forecasted\ \ value}{Actual\ \ value}$ x100% ……..Eq. (4)

STE = $\frac{567-542.67}{567} x100\%$

= 4.3%

From this STE figure we can conclude that model is validated with the error of 4.3%.

CHAPTER SIX

DISCUSSION AND CONCLUSION

6.1 Discussion of the clustering algorithms

The first part of the methodology of this thesis is to generate clusters according to different approaches and compare the clusters with respect to land-use, algorithms, covered area, and suitability to a spatio-temporal crime prediction model. In order to determine the most suitable clustering algorithm, K-means, Nnh hierarchical and spatio-temporal analysis of crime (STAC) are analyzed. After analyzing, we came to decide the appropriate cluster algorithm for this model and hence K-mean algorithm has been selected.

6.2 Discussion of the ARIMA andforecasting

After appropriate selection of the algorithms, we came to analyze the methodology in temporal part. In temporal analysis we had use ARIMA analysis. In ARIMA analysis, spike is seemed on 4^{th} lag and trial is begins from AR (4) and MA (4). After long analysis the best fitted value is found. This model is AR (1) and MA (1), which fits our model and also fulfills the Box Jenkins p-value.

After selection of ARIMA model, we forecasted the two months values. One is for month chaitra of 2070BS to validate the model and another is baishak 2071BS as a future forecasting of the model.

6.3 Conclusion

In this thesis, a spatio-temporal crime prediction model is created. After ananalysis of the spatio and temporal model, clusters and the number of incidents per week day arepredicted.The results of this study are to determine the affective clusterareasand the level of influence in terms of time.

The clustering algorithm based onspatial data and ARIMA modelis implemented in this study. In which, spatial analysis is used to create a suitable algorithm and temporal analysis is used to forecast the future value of the crime. Three different algorithms have been implemented in this thesis where K-mean is selected as an appropriate algorithm on the basis of larger area and more number of incidents covered by it. The ARIMA model is fitted with crime data to predict the future value in the time domain.

Hence, the spatio and temporal model of the thesis is generated to predict the future value of a crime in the Kathmandu Valley.

REFERENCES

[1] X. Wang and D. E. Brown, "The spatio-temporal generalized additive model for criminal incidents," in Intelligence and Security Informatics (ISI), 2070 IEEE International Conference on. IEEE, 2070, pp. 42–47.

[2]W.L. Gorrand R. Harries,"Introduction to Crime Forecasting," 2003, International Journal of Forecasting, Special Section on Crime Forecasting, Vol.19, pp. 551-555.

[3] H. Al-Madfai, C. Ivaha, G. Higgs, A.Ware, J.Corcoran "The Spatial Dissaggregation Approach to Spatio-Temporal Crime Forecasting," 2007, International Journal of Innovative Computing, Information and Control, Vol. 3, Number 3.

[4] M.Felson and R.V. Clarke, "Opportunity Makes the Thief" Crime Detection and Prevention Series,Paper 98. Police research Group. London: Home Office.

[5] L.Kaufman, and P. Rousseeeuw, Finding Groups in Data:"An Introduction to Cluster Analysis", John Wiley,New York, 1990.

[6]E. Polate "Spatio-temporal crime prediction model based on analysis of crime clusters" university, Turkey, 2007.

[7] R. Boba,"Crime Analysis and Crime Mapping", Sage, USA, 2005.

[8] S. Chainey, and J. Ratcliffe, "GIS and Crime Mapping", John Wiley, England, 2005.

[9] Web1: http://aic.gov.au/publications/crm/crm030.pdf, Australian Institute of Criminology. "Is crime predictable?"

[10] N. Levine, Crimestat: "A Spatial Statistics Program for the Analysis of Crime Incident Locations", Ned Levine and Associates and the National Institute of Justice, Washington, DC, 2002.

[11] Web2: http://www.ncss.com/download.html#Manuals in Box, Time Series Analysis Forecasting and Control, San Francisco.

[12] D. Pena, G.C. Tiaoand R.S. Tsay, "A Course in Time Series Analysis", 2001, John Wiley, Canada.

[13] M. Felson and E. Poulsen,"Simple indicators of crime by time of day", 2003 International Journal of Forecasting, Vol.19, pp.595-602.

[14] B. L. Bowerman and R. T. O'Connell, "Forecasting and Time Series",1993,an Applied Approach, Pacific Grove, CA: Duxbury

INTERVIEWS

Interview with a police officer (2071BS) in Naxal Police Headquarter.

Interview with a police officer (2071BS) in HanumanDhoka Police Station.