CHAPTER 3

LITRATURE SURVEY

3.1 INTRODUCTION

Crime could be an abuse of law, associated with criminals and victims. It is a problem which might be identified and acknowledged by the majority who expense if directly and indirectly. It adds stress to people’s lives and impairs the standard of a lifetime of people and communities. Crime has been studied for nearly 2 centuries by a conceit of various educational disciplines, as well as criminology, sociology. There are hypothetical and practical advances that have allowed crime to be studied from a variety of various geographical views.

This Chapter starts with the dimensions of the crime analysis in section 3.1, followed by the crime forecasting techniques and geographic models used in the crime forecasting in section 3.2. Crime classifications on each temporal and spatial thing are in section 3.3. Hot spotting techniques and the various varieties of spatial clustering algorithms utilized in hot spotting in section 3.4 and 3.5 respectively. Forecasting simulations and geographic approaches are in section 3.6. While collecting the present things during this research design scope of the research is in section 3.7. The concluding comments have been presented in section 3.8.
### 3.2 DIMENSIONS OF CRIME ANALYSIS

Cohen et al (1979) has proposed routine activity theory for environmental criminology and criminologists are interested in analyzing crime patterns, because understanding behavior can help crime reduction and prevention. The Underlying theories help explain crime behavior.

Skogan (1990) has proposed crimes that are also known to vary with time and location. They analyzed criminal behavior in space and time as spatial choice models and showed that they provided efficient and accurate predictions of future crime patterns.

Clark (1995) showed that situational crime prevention can reduce crime by altering the environment. It aims to stop crimes before they occur. Situational crime prevention can also mean improving street lighting, adding video surveillance cameras or just getting more pedestrians on the streets.

Hagan (2000) has proposed Crime analyses that have few quantitative techniques. A technique to generalize from a sample of one is to argue that cluster knowledge overlooks or shapes the importance of individual success or failure. Homothetic (group) styles merely add up the totals and appearance at averages. Idiographic (single subject) styles have the advantage of rescuing individual knowledge from the pile of averages. This argument works best if the individual in question falls into some extreme class (successful at crime or an entire failure at it). Scientists talk over with these cases as "outliers" and it is most likely to the higher, to use somebody successful than a failure. Studies of so-called successful, or able, criminals are particularly helpful to check out how most offenders try and avoid detection by law enforcement.
Bob a (2001) has proposed the central sort of information analyzed is crime, which specializes in crime information like arrests, offenders, victims, property and proof. Additionally to crime, law enforcement agencies address several alternative problems and therefore collect several alternative sorts of information, for instance, entails service (e.g. noise complaints, burglar alarms and suspicious activity), traffic info (e.g. accidents and citations), voters perceptions (e.g. worry of crime, crime prevention behavior, satisfaction with the police), victimization, probation records and parole info.

John Eck et al (2001) has observed that 10% of the offenders committing the most crimes are involved in about 50% of the offenses; that 10% of the most victimized people are involved in about 40% of the crimes; and that 10% of the places with the most crime, accounts for about 60% of crime. As a result, repeat address incidents dominate police work.

Weisel (2003) has advised throughout crime analysis that patterns are identified and relationships of crime and law enforcement information with different sorts of info are studied. Such information embodies Socio-demographic and Spatial (location).

Osborne et al (2003) has proposed effective crime analysis employs data mining, crime mapping, statistics, fresh methods, charting and a solid understanding of criminal behavior. In this sense, a crime analyst serves as a combination of an information systems specialist, a statistician, a researcher and a planner for law enforcement agency.

Amarnathan et al (2003) has proposed that the criminals have become technologically refined in committing crimes. Therefore, police desires such a criminal offense analysis tool to catch criminals and to stay ahead within the eternal race between the criminals and also the law
enforcement. The police ought to use the present technologies to allow themselves the much-needed edge.

Erika et al (2004) has proposed the macro-level analysis to verify the “collective guesswork” that may be used to divide the crime incidents within the space. Census and land use knowledge will show the densities of the buildings, the quantity of units in every building and whether or not they are located in an exceedingly residential or business space or mixed. At this time, we will verify the rates primarily based on the quantity of buildings and therefore the number of burglaries.

Chen et al (2004) has proposed that once knowledge has been ready for mining, the modeling stage will begin. Selecting and developing models involve domain information, the results of that are validated against known or expected results and either deployed or refined. Gupta et al (2006) in any intelligent system crime analysis tool for police; it is needed to know Indian Police structure, responsibilities of the police, key changes and challenges the police is facing.

Manish Gupta et al (2007) has proposed an interactive interface as a crime analysis tool for CCIS, primarily based on the present call support and information mining techniques, so as to make police activities efficiently. The proposed crime analysis tool, can give a higher edge and with the utilization of the crime analysis tool policing will be made effective, quick and accountable in their operation. The effectiveness of the proposed crime analysis tool has conjointly been illustrated in crime hot spot analyses. The proposed crime analysis tool for CCIS is quicker to implement and easier to use. The crime analysis tool will be integrated with the latest visualization techniques like Geographical data System for enhancing the understanding of the results and patterns.
Haifeng Zhang et al (2007) has proposed that the use the Location Quotients of crime (LQC) is a method of mapping the prevalent types of crime across urban neighborhood and adopting crime density to investigate the relationships between neighborhood characteristics and crime. Their research used three aspects (1) the rationale of the employment of official crime rates for neighborhood. (2) An enhanced understanding of the effectiveness and statistical properties and (3) Models of crime measured by density, crime rates and LQCS.

Brantingham and Brantingham (1984; 1993; 2008) have proposed the places that are crime attractors and alternative places that are crime generators. Crime engaging places tend to be locations that draw plenty of individuals for work, shopping, or recreation and where there is a relative lack of effective or consistent guardianship. Against the law attractor, may be a target-rich atmosphere, which will exist solely at sure times of the day when businesses alternative attractions are open to draw the focus to that location. Such places are generally characterized by a high density of economic land use and an occasional density of residential land use. People, as well as offenders, usually move to such locations.

Cummings (2008) has proposed data mining and predictive analytics for a spread of law enforcement and applications of intelligence, together with tactical crime analysis, risk and threat assessment, behavioral analysis of violent crime and proactive deployment ways are currently being employed by police as call support systems.

McCue and Parker (2009) have said the Richmond Police Department has found data mining and predictive analytics to be the most effective approach to addressing the so-called volume challenge associated with this massive influx of information and is pioneering the use of these tools in policing.
3.3 CRIME FORECASTING BASED ON GEOGRAPHICAL MODELS

Chong and Hendry et al (1986) have suggested that the analysis of forecast accuracy uses 1-step ahead forecast. We could calculate sequences of multi-step ahead forecasts, but this would require modeling the explanatory variables to obtain true multi-step forecasts, or using the actual values of the explanatory variables. Neither strategy appears very satisfactory: the results of the former will be influenced by the quality of the models for consumption and unemployment, whereas our focus is on modeling crime and the latter is open to critiques of the sort leveled.

Wieczorek and Hanson (1997) the data derived through geographic data systems provides an easy geographic inventory (e.g., Indicate the placement of all events that occurred inside a bound time amount in a specific location) or show totally different spatial patterns (e.g., the locations of burglaries might vary by the time of day – late-night incident locations might differ greatly from midday incident locations) or presents the results of a lot of complicated analyses.

Kulldorff M et al (1997) has proposed that crime detection and forecasting methods require a coarse aggregation of cases (e.g. by month, by square mile), due to both computational considerations and the relatively small number of serious crimes. These limitations reduce the spatial and temporal precision with which departments can pinpoint clusters of crime, as well as their ability to rapidly respond to these clusters. The use of expectation-based spatial scan statistic methods originally developed for the bio surveillance domain, which can use a finer aggregation of data and can efficiently search for emerging space time clusters of varying size and duration.
Lab (2000) has steered the police department to determine locations and times that are additionally prone to criminal activity. The initial step of crime prevention is to analyze the current standing of incidents such as determining the density or pattern of the incidents.

Weiss (2002) has proposed that exploratory knowledge Analysis lies at the core of Crime Forecasting System. Exploratory knowledge analysis (EDA) is used to investigate knowledge for the aim of formulating hypotheses for value testing. The objectives of the EDA are: recommending hypotheses regarding the causes of observed phenomena and assess the assumptions made on the statistical inference on which these are mostly based. He prompts the longer term graphical analysis techniques.

Gorr et al (2003) forecasting techniques are divided into two categories in terms of the predicted time period. Crime forecasting includes long-term forecast models for policy planning and applications in broader manner and short-term forecast models for tactical decision making.

Wilpen et al (2003) has discussed the origin of crime forecasting in the year 1998, when the US National Institute of Justice (NIJ) awarded five grants to study crime forecasting for police use as an extension of crime mapping. Instead of mapping only the recent crimes and assuming that observed patterns would persist, the objective was to forecast crime one period ahead, with results displayed as maps.

Hamilton-Smith (2004) has proposed the police service suffers from poor forecasting of crimes and as a result there are recommendations to still develop regionally owned and tailored forecasting skills amongst practitioners.
Jacqueline Cohen et al (2004) has shown that the result of estimated models have coefficients with expected positive signs of time lagged independent variables and a mixture of positive and negative coefficients for time and space lagged independent variables, reflecting crime attractor and displacement crime theories. The end results of forecast validations are that the leading indicator models produce useful forecasts that are significantly better than the extrapolation method. For this reason the leading indicator model passes our form of Granger causality test. The regression model is the best for forecasting large crime decreases, but the neural network is the best in forecasting large increases, all by wide margins.

Ratcliffe (2006) has proposed in Space-time interactions by considering how movement patterns were constrained by time in conjunction with spatial constraints. This points to geography approach, “that temporal constraints, in conjunction with the locations of offending nodes, are a serious determinant in spatial-temporal patterns of property crime”.

Michael Abramowicz et al (2007) has suggested that the challenges involved in predicting crime rates or the impact of different crime policies are very similar to those in other forecasting domains. Classic examples include predicting sales of a product, changes in interest rates, the likelihood of a terrorist attack, or the outcome of political elections. In each of these cases, the inputs needed to generate a reliable forecast may be tilted by a variety of factors, some of which might begin but troublesome, and some of which may be selfish and opportunistic.

Cohen et al (2007) has developed highly reliable methods for forecasting future crime trends and problems, which are the most preferred ways to improve crime prevention and reduction. With the advance of crime forecasting, spatial and temporal predictions of crimes are used to make long and short term planning. In the situation of getting
accurate predictions, it is possible to manage security resources efficiently. Police give attention on highlighted areas, target patrols, allocate resources and carry out other police interventions to prevent crimes.

Christian Ivaha et al (2007) has proposed Dynamic Spatial Distribution Approach (DSDA) which is a modeling approach that gives spatial-temporal forecasts that incorporate the influences of salient weather that have an impact on crime dynamics at each of the temporal and spatial levels.

Brantingham et al (2008) has proposed mathematical modeling that has been applied in many social science analysis initiatives. Recent applications within the field of criminology have included agent primarily based on, cellular automata process and system dynamics models. These kinds of models concentrate on the particular movement of agents (offenders), their interaction with the surroundings and therefore the development of patterns which are assuming an underlying set of rules for the model. The present project differs from these previous approaches by applying mathematical structures to outline the particular rules by that the higher than mentioned models might behave. Using directional and distance parameters outlined by offenders' home locations and therefore the locations of their crimes, the influence of activity nodes in the spatial distribution of crimes in crime neutral areas is investigated.

Diansheng Guo (2009) has proposed that spatial data mining and knowledge discovery has emerged as a lively analysis field that focuses on the event of theory, methodology and observe for the extraction of helpful info and information from large and complicated spatial databases. Spatial information mining continues to be at an awfully early stage and its bounds and potentials are however to be defined. There are opportunities and challenges facing each spatial information mining analysis.
Richard Berk (2010) has proposed that asymmetric costs can be introduced into loss functions to affect the forecasts produced. Sometimes different cost ratios can lead to very different forecasts.

3.4 CRIME CLASSIFICATION ON BOTH TEMPORAL AND SPATIAL

Jiawei Han et al (1992) has proposed an approach that applies an attribute-oriented concept tree ascension technique in generalization. It integrates the machine learning methodology to set-oriented database operations and extracts generalized information from actual data in databases. Our methodology substantially reduces the computational complexity of the database learning processes. Totally different information rules as well as characteristic rules, discrimination rules, quantitative rules and information evolution regularities are often discovered efficiently using the attribute oriented approach. In addition to learning in relational databases, these approaches are often applied to information discovery in mixed-relational and deductive databases. The discovered rules are often employed in intelligent question answering and semantic question optimization.

Oskari Heinonen et al (1996) has made a successful data mining technique of attribute oriented induction which is viewed as conceptual clustering. They supply the AOI algorithm for conceptual clustering. It performs two operations (1) Generalizing an attribute: it executes by choosing the closest rows and then selecting the attribute to make the final order. (2) Selecting the attribute for general comparisons: it performs the generalization step of an attribute and causes an update of all the rows promptly.

Douglas et al (1997) has proposed classified incidents of murder, rape and arson from the macro- to micro- level and used several principles
that would assist analysts in developing classifications for different crimes. Three forms of classifications highlighted that this might embrace decision kind, crime and uniform crime reports.

Ester et al (1997) has proposed a Classification method which consists of two parts. First, the user defines the number of classes. To test, whether the number of classes has been chosen correctly, a set of training data is selected and the classification is performed on it. Subsequently, classification rules are derived from the training dataset. Next, those rules are applied to the test dataset. Classification is considered as predictive spatial data mining, because we first create a model according to which the whole dataset is analyzed.

Qingshuang Jiang et al (1999) has proposed a Remarkable and economical rule induction strategy that ensures the generation of complicated and high-level decision rules for un-annotated datasets. Rule granularity has been considered as a very important thing about the comprehensibility of the discovered data. Longer rules and a lot of specific rules do not essentially give higher classification accuracy as compared to shorter and generalized rules. Hence, the requirement for a lot of concise rules and small set of rules. The rule induction strategy presented here permits for the generalization of rules, while maintaining classification accuracy via the incorporation of attribute oriented induction that integrates machine learning methodology with relational database operations with rough sets primarily based on rule induction ways.

Shekhar et al (2003) has proposed that spatial data are generally distributed over a continuous space, as opposed to traditional datasets that are usually discrete. The patterns found in traditional data tend to be global in nature, whereas spatial patterns tend to be local. Spatial regions tend to display a higher level of dependence between them, while traditional data
have a higher degree of independence from one sample to another. Although it is possible to apply some traditional data mining techniques to spatial data, ignoring these features of spatial data can often lead to poor performance in terms of both computational efficiency and pattern discovery.

Ahmadi (2003) has classified crimes into Security, Economic and Social crimes. Security crime focuses on security against someone and examples embody impediment, burglary, shoplifting, theft and motor theft. Economic crime focuses on crimes that are against cash or property of someone. Examples are auto theft, shoplifting, burglary and Fraud.

Michelangelo Ceci et al (2004) has presented a spatial associative classifier that mixes spatial association rule discovery with naïve Bayes classification. The domain specific information could also be outlined as a group of rules that creates doable the qualitative spatial reasoning. Additionally, hierarchies of spatial objects are expressed by collections of ground atoms and are exploited to mine classification models at completely different granularity levels. Such constraints are used to bias the spatial association rules discovery so as to satisfy user expectations. Constraints are used to partially fix the structure of extracting rules so as to get spatial association rules that contain solely the category label within the head. Finally, for every granularity level, extracted rules concur in building the spatial classification model by exploiting a multi-relational naïve Bayesian classifier integrated with the SDB.

Vera Karasov (2005) has proposed every data object stored in a database is characterized by its attributes. Classification is a technique, the aim of which is to find rules that describe the partition of the database into an explicitly given set of classes. Objects with similar attribute values are integrated into the same class. In spatial classification the attribute values of neighboring objects may also be relevant to the membership of objects in a
certain group. Therefore, we have to include the neighborhood factor in the calculation.

Laxman and Sastry (2006) have proposed that temporal classification involves predicting the class of a temporal sequence. This is done by looking at previously classified temporal data, building a predictive model based on that data and then using it to predict the category of unclassified sequences in the data.

Home Office (2006) has noticed Changes in the way in which police record crime also affects figures. The police station officers counting the rules for the count and classifying of notable offenses recorded by the police forces. The classification was created a number of censes variables and classifies every output area.

Chun-Che Huang et al (2009) has proposed that methodology in crime classification examines the generalized rules based mostly on credible classification. Furthermore, a rule extraction algorithm was developed for the induction of rules from knowledge sets. The comparatively small variations of rules introduced, allowed for an inexpensive domain interpretation. The testing effort involved forty utterly totally different sets obtained from randomly generated data. The computational results indicate that the algorithm may handle large-scale knowledge sets.

3.5 HOT SPOTTING TECHNIQUES AND VISUALIZATION

Tobler (1970) has proposed that a hot spot is a geographical area, with higher than average incidence of certain disordered events, or an area where people have a higher than average risk of victimization. This is interesting because everything is related to everything else but nearby things are more related than distant things.
Bailey et al (1995) has proposed that the methodology for determining the hotspot areas is kernel estimation that is performed by density analysis in ArcGIS. Kernel estimation was originally developed to get a quick estimate of a univariate or multivariate likelihood density from an observed sample of observations.

Curry et al (2000) has found out that the regional crime analysis program is capable of collecting statewide data and allowing statewide access to this data. It also analyzes the data, plots the analyses, reports the analyses and maps the data using GIS (graphical information systems) software.

Chainey et al (2005) has proposed that one consistent criticism with the new hot spot analysis of incident-level crime is that several of the parameters used to see a hot spot are subjective. Widely differing results might come back from even little alterations within the defined parameters of the analysis. When looking forward to thematic mapping techniques, the problem is compounded by the modifiable areal unit problem (MAUP), which ends from spatial analyses conducted on a true unit of varying size outlined by administrative boundaries instead of the physical geography.

Xiaobai Yao et al (2005) has proposed four visualization strategies to depict the query results of a qualitative location characterized by multiplicity and uncertainty. To decide which strategy is best for a certain scenario, it is necessary to find which characteristics are presented in the query results. Spatial data visualization strategies for text-based location information can contribute tremendously to the provision of geographic information services via cell phones and PDAs, as well as to other GIS applications in which intuitive computer–user communication is important. Furthermore, QL analysis and visualization can also contribute to exploratory data analysis and knowledge discovery in databases where text-based location information is significant.
Grubesic (2006) has proposed fuzzy clustering method, which is a part of partitioning clustering methods. This differs when fractional memberships of observations are taking values less than 1. The goal of the fuzzy clustering method is to minimize the aggregate distance between observations and clusters, including the dissimilarity of not weighting observations.

Michael et al (2006) has proposed that an immediate benefit of the spatial component in the spatial database is the ability to use GIS to visually explore spatial distributions of crime hotspots and relationships to other spatial data layers. Spatial analysis approaches, such as the nearest neighbor analysis, can reveal whether patterning, clustering, or heterogeneous relationships exist. More advanced techniques such as the quadratic kernel density and quadrate analyses can reveal information about the spatial intensity of crime. Visual explorations enhanced by analytical approaches such as nearest neighbor analysis, kernel density estimates and quadrate analysis can confirm or reject observed patterns.

Isaac Van Patten et al (2008) has proposed that PAI and RRI are to be immensely helpful within the analysis of a convergent hot spot analysis of theft. While applying this PAI and PRI methodology many limitations should be taken into consideration. First of all, like the other studies based mostly on official crime statistics, this study was conducted on an analog of the true theft scenario above all space.

Michael Townsley (2008) has proposed a method for visualizing spatial and temporal patterns of crime at different levels of spatial and temporal resolution. The Trellis-like trait of vertical and horizontal alignment of plots allows efficient decoding of information by the reader, specifically: comparing the long term levels of crime within each hot spot, comparing the intra-day time signature of each hot spot, the relative spatial and numerical
magnitude of each hot spot, and the production of a baseline to use against the observed hot spot patterns.

Spencer Chainey et al (2008) has proposed KDE is increasingly becoming the technique of choice by those that generate hot spot maps, partly based on the findings from previous reviews that have profiled the technique's ability to outperform others in accurately identifying the location, size, orientation and spatial distribution of the underlying point data and the visual appeal in the output that it generates. KDE is also the best of the common hotspot mapping techniques for predicting spatial patterns of crime. Introducing the PAI has made it possible to make comparisons between hotspot mapping techniques and crime types in their ability to predict future crime events. Its calculation also allows the analyst to generate a statistic (the hit rate), that indicates the possible impact the targeted resource activity could have, if it was focused in those areas, defined as hotspots.

Giuseppe Borruso (2008) have proposed Network Density Estimation (NDE) and is impressed by KDE as a way for analyzing the local spatial distribution of event processes, placed on a network, during the study of a region. The important characteristic of the performance is that of considering shortest-path trees instead of circular search functions for a density analysis, thus computing solely events which will be reached along the network segments. The strategies are often implemented using commonplace GIS and spatial statistics functions.

Michael Townsley (2008) has proposed that “Hotspot plot” may be a systematic, intuitive approach for displaying crime patterns in two dimensions (space and time) simultaneously. The hotspot plot demonstrates that, for the information used here, there exists considerable variation within the temporal distribution between hot spots, at totally different levels of
resolution: within the year and within the day. These variations are located effectively owing to the careful style of the plot.

Chainey et al (2008) has proposed a metric for evaluating the effectiveness of various hot spotting techniques that shows nice promise. The Predictive Accuracy Index (PAI) is specifically designed to think about the localized hot spot against the larger study area in an exceedingly good type that enables direct comparison of hot spot prospective accuracy between completely different study regions. The PAI springs from a ratio of the hit rate (the variety of crimes within the hot spot compared to the region) to the world share (proportion of the region represented by the recent spot). Their analysis was reported on the analysis of three completely different general techniques of hot spot maps, viz. Thematic mapping, nearest-neighbor derived ellipses and kernel density estimation. Using baseline information, they predicted the accuracy of the various strategies in predicting “future” crime in those areas identified as hotspots.

3.6 SPATIAL CLUSTERING ALGORITHM FOR HOTSPOT IDENTIFICATION

Gutrell et al (1995) has proposed that the point events at different locations are called Kernel Density Estimation. This method is used to generate a continuous crime density surface, from crime point data. The analyst begins with a dot map of crime events. Kernel smoothing results in a continuous weather map that shows geographic variation in the density or intensity of crime. Peaks in the map represent areas of high crime hotspots and valleys represent areas of low crime.

Zhang et al (1996) has proposed that an effective hierarchical clustering is a new method to deal with the large data sets, BIRCH (Balanced and Iterative Reducing and Clustering using Hierarchies). BIRCH
summarizes an entire dataset into a CF-tree and then runs a hierarchical clustering algorithm on a multi level compression technique, CF-tree, to get the clustering result. Its linear scalability is good at clustering with a single scan and its quality can be further improved by a few additional scans. It is an efficient clustering method on arbitrarily shaped clusters. But BIRCH is sensitive to the input order of data objects and can also deal with numeric data only. This limits its stability of clustering and scalability in real world applications.

Wang W et al (1997) has proposed STING (STATistical INformation Grid). This is a method which relies on a hierarchical division of the data space into rectangular cells. Each cell at a higher level is partitioned to form a number of cells of the next lower level. STING can be used to answer efficiently different kinds of region-oriented queries. However, when the number of k is very large in high dimensions, both the quality of cell approximations of clusters and the runtime for finding them, deteriorate.

Rakesh Agrawal et al (1998) has proposed CLIQUE (CLustering In QUEst) for high-dimensional data, as a density-based clustering method. CLIQUE automatically finds subspaces (grids) with high-density clusters. CLIQUE produces identical results, irrespective of the order in which input records are presented and it does not presume any canonical distribution of the input data. Input parameters are the size of the grid and a global density threshold for clusters. CLIQUE scales linearly with the number of input records and have good scalability as the number of dimensions in the data.

According to Rakesh Agrawal (1998), the CLIQUE algorithm was one of the first subspace clustering algorithms. The algorithm combines density and grid based clustering and uses an APRIORI style search technique to find dense subspaces. Once the dense subspaces are found, they are sorted by coverage, defined as the fraction of the dataset the dense units in the
subspace representations. The subspaces with the greatest coverage are kept and the rest is pruned. The algorithm then finds adjacent dense grid units in each of the selected subspaces using a depth first search. Clusters are formed by combining these units using a greedy growth scheme. The algorithm starts with an arbitrary dense unit and greedily grows over a maximal region in each dimension until the union of all the regions covers the entire cluster. Redundant regions are removed by a repeated procedure.

Jain et al (1999) has suggested that cluster analysis as an exploratory discovery process. It can be used to discover structures in data without providing an explanation/interpretation. Cluster analysis includes two major aspects: clustering and cluster validation. Clustering aims at partitioning objects into groups according to a certain criteria. To achieve different application purposes, a large number of clustering algorithms have been developed.

Mihael Ankerst (1999) has proposed a cluster analysis method, based on the OPTICS algorithm. OPTICS computes an augmented cluster ordering of the database objects. There are two techniques used for exploring the clustering structure, offering additional insights into the distribution and correlation of the data. This OPTICS algorithm will to automatically extract not only ‘traditional’ clustering information but also the intrinsic, hierarchical clustering structure.

Chun-Hung Cheng et al (1999) has proposed the ENCLUS subspace clustering method based heavily on the CLIQUE algorithm. However, ENCLUS does not measure density or coverage directly, but instead measures entropy. The algorithm is based on the observation that a subspace with clusters typically has lower entropy than a subspace without clusters. Cluster ability of a subspace is defined using three criteria: coverage, density and correlation.
Jae-Woo Chang (2000) the Cell-based clustering method used for fast clustering the high dimensional data. For performance analysis, they compare cell-based clustering method with the CLIQUE method, which is one of the most efficient clustering methods for handling a large amount of high-dimensional data. The performance analysis results show that the cell-based clustering method shows slightly lower precision, but it achieves better performance on retrieval time as well as cluster construction time.

Harsha S et al (2000) has proposed MAFIA algorithm, which extends CLIQUE by using an adaptive grid, based on the distribution of data to improve efficiency and cluster quality. MAFIA also introduces parallelism to improve scalability. MAFIA initially creates a histogram to determine the minimum number of bins for a dimension. The algorithm then combines adjacent cells of similar density to form larger cells. In this manner, the dimension is partitioned based on the data distribution and the resulting boundaries of the cells capture the cluster perimeter more accurately than fixed sized grid cells. Once the bins have been defined, MAFIA precedes much like CLIQUE, using an APRIORI style algorithm to generate the list of clusters enabled subspaces by building up from one dimension. MAFIA also attempts to allow for parallelization of the clustering process.

Jiawei Han (2000) has proposed CLARANS (Clustering Large Applications based on RANdomized Search) as an improved k-medoid type method, based on randomized search. CLARANS is partly motivated by two well-known algorithms in cluster analysis, called PAM (Partitioning around Medoids) and CLARA (Clustering LARge Applications). Since CLARANS only checks a sample of the neighbor of a node and each sample is drawn dynamically, CLARANS is significantly more efficient than the PAM and CLARA. But CLARANS is still too inefficient to be applied to a large
database because it limits the size of the database by assuming that all objects are stored in main memory.

Kolatch (2001) has proposed the density-based groups of clustering algorithms which represent a data set in the same manner as partitioning methods; converting an instance to a point, using the data attributes of the source set. The plane contains clusters with high internal density and low external density in a similar manner to its partitioning ancestor. The process of adding points to a cluster is iterative, unlike partitioning methods. Nearest neighbors of each point can thus be investigated, arbitrary shapes formed and existing clusters merged, as the algorithm moves through all points.

Niu (2003) has suggested that Clustering techniques discover clusters of comparable characteristics of the given data. The growing availability of spatial information has resulted in efforts to develop efficient algorithms to modify the large dimensions of spatial databases and with a number of problems like noise, outliers, non-convex shapes etc. Completely different clustering algorithms exist for locating clusters showing spatial and temporal patterns in spatial databases and in temporal databases respectively.

Haining et al (2003) has proposed that crime analysts frequently assume that crime distributions are clustered and whether clusters exist or not, however some can be identified from random crime distributions. Testing for clustering is the first step in revealing whether data have hot spots of crime. Several approaches can be applied to test for clustering in crime distributions. Most methods incorporate the basic principles of testing hypothesis and classical statistics, in which the initial assumption is that the crime distribution is one of complete spatial randomness.

Lance Parsons et al (2004) has proposed that the region growing density based approach for generating clusters, allows CLIQUE to find
clusters of arbitrary shape, in any number of dimensions. Clusters may be found in the same, overlapping, or disjoint subspaces. The DNF expressions used to represent clusters are often very interpretable and can describe overlapping clusters, meaning that instances can belong to more than one cluster. This is often advantageous in subspace clustering since the clusters often exist in different subspaces and thus represent different relationships.

Pilevar (2005) has proposed a new clustering algorithm GCHL (a Grid-Clustering algorithm for High-dimensional very Large spatial databases), capable of efficiently and effectively clustering large high dimensional data sets. It relies on a novel active sampling approach and uses a grid axis-parallel partitioning scheme to identify the dense region in the input data space. GCHL has good accuracy and scalability. It is robust to noise, automatically detects the number of clusters in the data and can successfully operate with limited memory resources. We show the generality of our approach, i.e. we are able to simulate a density-based and grid-based clustering. The advantage of GCHL algorithm in using AED technique and discovering concave/deeper and convex/higher regions makes it applicable in medical and geographical applications.

Chainey et al (2005) have proposed temporal forecasting, expounded with temporal distribution of crime. There is not any model specified for temporal forecasting in literature. Temporal distribution and time series graphs of crime are the tools for predicting the probable time of the event.

Nancy Lin et al (2007) has proposed a new grid-based clustering algorithm, called the Deflected Grid-based (DGD) algorithm, which has the obvious wider ranges of size of the cell and the threshold of density. And the experimental results show that the effect of DGD algorithm is less influenced
by the size of the cells than other grid-based ones. At the same time, the DGD algorithm still inherits the advantage with a low time complexity.

Borah et al (2008) has presented DDSC algorithm that provides a solution for the problem of finding clusters with varying sizes, shapes and densities. Clusters extracted from a data set are non-overlapped spatial regions, having different densities. Adjacent clusters may be very close, without being separated by any sparse regions. The algorithm is an extension of DBSCAN and uses a homogeneity test to detect density variations. A cluster is reasonably homogeneous locally.

Braga et al (2010) has proposed that places are important in crime prevention and crime control. Crime does not spread evenly across city landscapes; rather, crime clusters in very small places that generate a disproportionate amount of criminal events. Compared to people who commit crime, places where crime occurs are far more predictable. So, if crime control policies respond to places with concentrated crime, crime can be reduced more effectively.

Anant Ram et al (2010) has proposed that DVBSCAN is an enhancement of DBSCAN algorithm. The proposed clustering algorithm can find clusters that represent relatively uniform regions without being separated by sparse regions. The parameters $\alpha$ and $\beta$ are used to limit the amount of allowed local density variations within the cluster. The future works can be focused on to reduce the time complexity of algorithms and to determine the value of parameters automatically for better clustering, for any given data set.
3.7 GEOGRAPHIC SIMULATION AND ANALYSIS

Box and Jenkins (1970) have proposed two of the additional widely known univariate ways are the Box-Jenkins models. Box-Jenkins models are appealing theoretically. However they are difficult to use and customarily are not the foremost correct forecasting ways. Exponential smoothing ways are widely employed in observing. They are easy to know and use and have consistently yielded good results, if not the simplest forecast with accuracy.

Philip Canter (1990) has proposed that Crime analysts and researchers are creating substantial progress in identifying high crime activity areas and predicting future target locations. It is evident that descriptive and analytic mapping is going to be a very important part of strategic and tactical crime analysis efforts. The micro simulation model could be used to conduct policy simulations and forecasting. The simulation would entail conducting a baseline simulation using a given initial population sample and later changing either the sample characteristics or parameters within the model in order to gauge the effects of policy or structural changes.

Gilbert (1999) has proposed that research using simulation begins with the representation of a theory in the form of a model. To do this, reality must be simplified and only the most important aspects represented in the model have to be included and all extraneous details are dropped. In contrast to statistical models that are implemented as equations and solved within a software program, a simulation model is itself a software program.

Simulation may be a broad field of science that encompasses a variety of approaches that share a group of characteristics. The sphere of the simulation involves the creation of models that distill a phenomenon into its most significant components. Interactions at intervals in simulation models
are very advanced and therefore simplicity at the beginning is important to understanding.

Berk et al (2000) has proposed simulation Models, developed from theory and specified in a computer program where they were able to accommodate dynamic, non-linear interactions that play out over time. A formalized computer code provides concrete documentation for the assumptions of the model and allows transparency within the analytic enterprise that is necessary for replication and testing of results. These attributes are particularly necessary when one is attempting to find the mechanisms through the observed macro-level patterns that are formed. In short, simulated experiments involve identifying and codifying a variety of advanced relationships and then making a model of the globe because it exists consistent with the theory or before an intervention.

WAN Hongtao et al (2001) have proposed that the automated generation algorithm of the 2-D flood simulation model based mostly on the GIS technology is mentioned intimately. The detailed flood model construction method involves these stages (1) delineating the computational domain. The computational domain is obtained from an existing file of domain boundary or digitized from the desktop or screen; (2) generating the computational mesh. The computational domain is discrete by regular or irregular computational mesh; (3) generating the mesh topology within the irregular and unstructured mesh, the topology ought to be designed, like the topology between mesh polygon and part interface, the topology between part interface with the mesh polygon and nodes; (4) extracting the model parameters the derivations of boundary conditions, initial conditions, physical parameters and parameters for monitoring and adjusting the simulation method.
Bonabeau (2002) has proposed Agent-Based Modeling (ABM), which is another type of computer simulation created through development of object-oriented programming languages and aided by advancements in information and computing power. Agent-based models rely on a bottom-up approach to computer simulation, where a couple of, simple, theory based rules are developed for the individual agents. The interactions of people within the model, manufacture macro-level patterns that emerge from the simulation.

Gorr et al (2003) has shown that crime forecast error measures vary inversely with increasing incidence count utilized in estimating time series forecast models. Average crime counts per unit time amount and geographic space of a minimum of 25–35 are required before three key stages. The primary forecast errors become acceptable.

Hawkins et al (2003) has pointed the WeBCAT system was implemented using Microsoft Visual Interdev as the development platform. Active server page (ASP) and VBScript were used to build the web pages, ESRI ArcIMS to build the GIS tools and XML to store, transfer and access the crime data sets.

Wilpen et al (2004) has proposed early crime warning systems and introduced an application of tracking signals for detecting experienced time series pattern changes in crime maps. The basis of the tracking signal is, information obtained from the counter-factual forecasts, for each point examined. These are forecasts providing business-as-usual estimates for a point in time, as if no changes to pattern existed. The tracking signal automates detection of pattern changes, by matching the decisions of crime analysts, as to what data points constitute the start of a new time series pattern.
Yifei Xue et al (2006) has proposed two models of spatial alternative. The primary model is a uniform spatial alternative model, which makes a powerful assumption on criminal preferences. Criminals have a similar alternative setting and preferences. The second model is a distinct spatial alternative model, which relaxes this assumption. Along the models show how the preferences of criminals will be modeled to perceive the spatial patterns of crime. When evaluated with actual breaking and coming into information, these strategies increased the accuracy of the prediction of future criminal locations considerably. The variations of the paired predictions between the new spot approach and also the new models are visualized by bar plots. The testing results and also the visualization how that the new models perform far better than the new spot strategies using density estimation. Additionally, the tactic provides some way to interpret the link between criminal call creating and spatial attributes.

Sergio Rey (2006) has proposed that STARS (Space–Time Analysis of Regional Systems) could be a powerful atmosphere for exploring data that have each temporal and spatial dimensions, additionally some basic styles of spatial clustering are identified by a variety of graphs and maps created within the current version of STARS. It is comfortable with writing straightforward macro-type scripts (in Python) to use STARS for simulation experiments also as for linkages with different model systems and statistical packages.

Daniel Neill et al (2007) has suggested that the violent crime and leading indicator data demonstrates that expectation-based scan statistics can efficiently and accurately detect significant spatial clusters of crime, at a higher spatial and temporal resolution than previously proposed crime detection techniques. And they demonstrated that detected clusters of leading indicator crimes can be used to predict significant clusters of violent crime 1-
3 weeks in advance, allowing police departments to dynamically allocate patrols to these areas and carry out other interventions to prevent crime.

Mark Mitchell et al (2007) has proposed development of a crime forecasting tool that integrates with the WebCAT system. The major steps in this project involved the selection of a suitable algorithm, as well as appropriate statistical and GIS software packages. Once these choices were made, it was necessary to develop the different components of the system and then integrate them with WebCAT. Testing was conducted to verify whether the algorithm used to perform acceptably, relative to other algorithms, while also being fast enough to remain usable.

Kelvin Leong (2008) has proposed that a new model known as Space-Time-Event Model (STEM) has been introduced. The model incorporates space, time and event into consideration, for the purpose of crime pattern discovery purpose.

Elizabeth Groff (2008) has proposed that technology, requiring programming knowledge, is a substantial barrier to develop simulation models for practitioners. Therefore, a bridge between the domain experts with the most to gain and those with the technical know how to design and develop models must be built. Secondly, work must be done to address the issues associated with validating simulation output against recorded crime data.

Malleson and Brantingham (2009) have proposed an agent based model to forecast burglary. Their experiments are fashioned around “potential burglar” agents who would like wealth to support them. The agents will either work or burgle to induce wealth, however not all agents will work enough to
induce the wealth they require. Every agent decides whether or not to burgle and what to burgle, primarily based on predefined rules.

Johnson et al (2009) has proposed that another benchmark model is that the KDE, introduced by (Gatrell 1995). This model is incredibly the same as the random walk, since it is also a retrospective technique, which uses knowledge from a precise time interval within the past. It is not clear what quantity of information and what time interval, is best to use.

Chung-Hsien Yu et al (2011) has suggested the preliminary results of a crime forecasting model developed in collaboration with the police department of a United States city in the Northeast. At first, we discuss our approach to architecting datasets from original criminal records. The datasets contain aggregated counts of crime and crime-related events categorized by the police department. The location and time of these events are embedded in the data. Additional spatial and temporal features are harvested from the raw data set. Second, an ensemble of data mining classification techniques is employed to perform the crime forecasting. We analyze a variety of classification methods to determine which is best for predicting crime “hotspots”.

Tibor Bosse et al (2011) has used the model to investigate the effect of different class compositions on the development of the delinquency in the total group of pupils. The two different scenarios represented two extreme situations: all delinquent pupils put together, or all delinquent pupils distributed over all classes. Therefore, our tentative conclusion is that the composition of classes has not so much effect on the overall development of the delinquency of the pupils in a school. The result finds few remarks that can be made about our experiments First of all, the model is possibly not very
precise (see the relatively limited accuracy) because of the small size of the training set. A second remark concerns the size of the classes.

3.8 SCOPE OF THE THESIS

From the literature review, spatial temporal forecasting model, subspace spatial clustering and temporal clustering of crime attributes provided high prevention for forecasting crime. As hotspot detection, crime attribute classification and simulation are added in service and the result provides top quality.

Based on this observation, the proposed research work satisfied the above in the following implementation.

- To classify the crime data attributes and prepare KDE map to produce crime hotspot and cold spot.
- To detect crime hotspot and to identify hot spot based on high density and shape of the crime clusters.
- To apply discrete computer simulations for identifying crime hotspots and also to provide long and short time forecasting for preventing crime.

3.9 CONCLUSION

Crime is a very unequally distributed happening and cannot be understood outside its social context. The association of demographic and socioeconomic characteristics with the locations of crimes can provide a clearer picture of crime, such as demographic and socioeconomic characteristics, by using GIS together with simulation analysis. As are often
seen from the literature reviewed in this chapter, there is considerable proof that spatial, time characteristics are associated with crime, particularly the crime classification. Crime hotspot identification is that the major part during this analysis is reviewed in this chapter, spatial clustering in data mining give variety of techniques for identifying crime hotspots and cold spots. Arithmetic computer simulations are reviewed in this chapter; it gives evidence for executing a range of simulation tools for crime forecasting.