

## **ABSTRACT:**

Crime analysis and prevention is a systematic approach for identifying and analyzing patterns and trends in crime. Our system can predict the type of crime activity which have high probability for given location interms of latitude and longitude and date and also we can visualize crime prone areas. With the increasing advent of computerized systems, crime data analysts can help the Law enforcement officers to speed up the process of solving crimes. Using the concept of data mining we can extract previously unknown, useful information from an unstructured data. Here we have an approach between computer science and criminal justice to develop a data mining procedure that can help solve crimes faster. Instead of focusing on causes of crime occurrence like criminal background of offender, political enmity etc we are focusing mainly on crime factors of each day.

**KEYWORDS:** Clustering, k-means Algorithm, Decision Tree, Crime.

# CONTENTS

<b>TITLE</b>	<b>PAGE NO.</b>
Abstract	5
Keywords	5
List of Figures	vi
List of Tables	vii

<b>1. Introduction</b>	9-10
1.1 Introduction	9
1.2 Problem Statement	10
<b>2. Literature Survey</b>	11-12
2.1.Cluster Aanlysis For Aanmoly Detection In Accounting Data	11
2.2 Analysing Violent Criminal Behaviour By Simulation Model	11
2.3 An Intelligent Analysis Of A City Crime Data	12
<b>3. Methodology</b>	13-15
3.1	
3.1.1 Clustering	
3.1.2 Algorithm K-Means	13
3.1.3 Algorithm Illustration Process	13
3.2. Proposed System	14
3.2.1. Architecture	15
<b>4. Design</b>	16-22
4.1 UML Diagrams	
4.1.1 Use case Diagram	17
4.1.2 Class Diagram	18
4.1.3 Sequence Diagram	19
4.1.4 Collaboration Diagram	20
4.1.5 Activity Diagram	21
4.1.6 Deployment diagram	22

<b>5. Experimental analysis and results</b>	23-29
5.1 System configuration	23
5.1.1 Software requirements	23
5.1.2 Hardware requirements	23
5.2 Sample Code	24-29
5.3 Screen shots	30-38
5.4 Experimental Analysis/Testing	39-42
<b>6. Conclusion and Future work</b>	43-44
6.1 Conclusion	43
6.2 Future Work	44
<b>7. References</b>	45

# **INTRODUCTION**

## **1.1 INTRODUCCION:**

Day by day the crime rate is increasing considerably. Crime cannot be predicted since it is neither systematic nor random. Also the modern technologies and hi-tech methods help criminals in achieving their misdeeds. According to Crime Records Bureau crimes like burglary, arson etc have been decreased while crimes like murder have been increased. Even though we cannot predict who all may be the victims of crime but can predict the place that has probability for its occurrence. The predicted results cannot be assured of 100% accuracy but the results shows that our application helps in reducing crime rate to a certain extent by providing security in crime sensitive areas. So for building such a powerful crime analytics tool we have to collect crime records and evaluate it.

## **1.2 PROBLEM STATEMENT:**

Criminals are nuisance for the society in all corners of world for a long time now and measures are required to eradicate crimes from our world. Our mission is to offer crime prevention application to keep public safe. Current policing strategies work towards finding the criminals, basically after the crime has occurred. But, with the help of technological advancement, we can use historic crime data to recognize crime patterns and use these patterns to predict crimes beforehand. We are using clustering algorithms to predict crime prone areas.

## **2. LITERATURE SURVEY:**

There are various papers which contributed to the study of sentimental classification of citations. Based on the study of these papers, this project was proposed.

### **2.1 CLUSTER ANALYSIS FOR ANOMOLY DETECTION IN ACCOUNTING DATA**

Paper-1 Summary: Proposed by Sutapat Thirprungsri

The purpose of this study is to examine the possibility of using clustering technology for continuous auditing. Automating fraud filtering can be of great value to preventive continuous audits. In this paper, cluster-based outliers help auditors focus their efforts when evaluating group life insurance claims. Claims with similar characteristics have been grouped together and those clusters with small population have been flagged for further investigations. Some dominant characteristics of those clusters are, for example, having large beneficiary payment, having huge interest amount and having been submitted long time before getting paid. This study examines the application of cluster analysis in accounting domain. The results provide a guideline and evidence for the potential application of this technique in the field of audit.

### **2.2 ANALYZING VIOLENT CRIMINAL BEHAVIOUR BY SIMULATION MODEL**

Paper-2 Summary: Proposed by K. Zakhir Hussain

Crime analysis, a part of criminology, is a task that includes exploring and detecting crimes and their relationships with criminals. The high volume of crime datasets and also the complexity of relationships between these kinds of data have made criminology an appropriate field for applying data mining techniques. Identifying crime characteristics is the first step for developing further analysis. The knowledge that is gained from data mining approaches is a very useful tool which can help and support in identifying violent criminal behaviour. The idea here is to try to capture years of human experience into computer models via data mining and by designing a simulation model.

### 2.3 AN INTELLIGENT ANALYSIS OF A CITY CRIME DATA

Paper-3 Summary: There had been an enormous increase in the crime in the recent past. The concern about national security has increased significantly since the 26/11 attacks at Mumbai, India. However, information and technology overload hinders the effective analysis of criminal and terrorist activities. Crime deterrence has become an upheaval task. The cops in their role to catch criminals are required to remain convincingly ahead in the eternal race between law breakers and law enforcers. Data mining applied in the context of law enforcement and intelligence analysis holds the promise of alleviating such problem. In this paper we use a clustering/classify based model to anticipate crime trends. The data mining techniques are used to analyse the city crime data from Police Department.

### 2.4 DATA CRIME APPROACHES TO CRIMINAL CAREER ANALYSIS

Paper-4 Summary: Narrative reports and criminal records are stored digitally across individual police departments, enabling the collection of this data to compile a nation-wide database of criminals and the crimes they committed. The compilation of this data through the last years presents new possibilities of analysing criminal activity through time. Augmenting the traditional, more socially oriented, approach of behavioural study of these criminals and traditional statistics, data mining methods like clustering and prediction enable police forces to get a clearer picture of criminal careers.. Four important factors play a role in the analysis of criminal careers: crime nature, frequency, duration and severity. This method yields a visual clustering of these criminal careers and enables the identification of classes of criminals. The proposed method allows for several user-denied parameters.



## **3.METHODOLOGY:**

### **3.1.1. CLUSTERING:**

Clustering is an unsupervised task without having a priori knowledge by discovering groups of similar documents. There are two types of categories in clustering algorithms; they are the partitioned algorithm and the hierarchical algorithm. K-Means algorithm and the link clustering they come under these two categories. K-Means and hierarchical clustering have many comparisons. In hierarchical clustering the size of data increases as the computational expansive, K-Means is faster. It updates the centroid clusters with each iteration and reallocates each document by its nearest centroid by this we can say that it is an iterative algorithm

### **3.1.2 ALGORITHM K-MEANS:**

K-means clustering is one of the method of cluster analysis which aims to partition  $n$  observations into  $k$  clusters in which each observation belongs to the cluster with the nearest mean. K means algorithm complexity is  $O(tcn)$ , where  $n$  is instances,  $c$  is clusters, and  $t$  is iterations and relatively efficient . It often terminates at a local optimum. Its disadvantage is applicable only when mean is defined and need to specify  $c$ , the number of clusters, in advance. It unable to handle noisy data and outliers and not suitable to discover clusters with non-convex shapes.K-Means clustering investigation plans to partition  $n$  perceptions into  $k$  bunch during which each perception includes a place with the bunch with the nearest centroid. 19

### **3.1.3 Algorithm Illustration Process:**

1. Initially, the number of clusters must be known let it be  $k$
2. The initial step is to choose a set of  $K$  instances as centres of the clusters.
3. Next, the algorithm considers each instance and assigns it to the cluster which is closest.
4. The cluster centroids are recalculated either after whole cycle of re-assignment or each instance assignment.
5. This process is iterated.

### **3.2 PROPOSED SYSTEM:**

In the proposed system, we done crime data analysis of with many parameters and factors including daily arrests, monthly arrests, number of domestic violence, top 5 monthly, weekly and daily crime are visualized.

Using Decision Tree algorithm and K-means clustering algorithm, we are predicting the type of crime for the given latitude and longitude.