ABSTRACT

Crime analysis and prediction is a systematic approach for identifying the crime. This system can predict regions which have high probability for crime occurrences and visualize crime prone areas. Using the concept of data mining we can extract previously unknown, useful information from unstructured data. The extraction of new information is predicted using the existing datasets. Crimes are treacherous and common social problems faced worldwide. Crimes affect the quality of life, economic growth and reputation of a nation. With the aim of securing the society from crimes, there is a need for advanced systems and new approaches for improving the crime analytics for protecting their communities. Propose a system which can analyze, detect, and predict various crime probability in a given region.Explains various types of criminal analysis and crime prediction using several data mining techniques.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF ABBREVIATIONS

ABBREVIATION	EXPANSION
CSS	CASCADING STYLE SHEET
HTML	HYPER TEXT MARKUP LANGUAGE
HTTP	HYPERTEXT TRANSFER PROTOCOL
KNN	K-NEAREST NEIGHBORS
URL	UNIFORM RESOURCE LOCATOR
WWW	WORLD WIDE WEB

CHAPTER 1

INTRODUCTION

1.1 OUTLINE OF THE PROJECT

Day by day crime data rate is increasing because the modern technologies and hi-tech methods are helps the criminals to achieving the illegal activities .according to Crime Record Bureau crimes like burglary, arson etc have been increased while crimes like murder, sex, abuse, gang rap etc have been increased.crime data will be collected from various blogs, news and websites. The huge data is used as a record for creating a crime report database. The knowledge which is acquired from the data mining techniques will help in reducing crimes as it helps in finding the culprits faster and also the areas that are most affected by crime .

1.2 OBJECTIVE

This system gives the most trending technology-based skills used at the present. To help police to detect the crime type based on location. Provides the user with the technology he is saving a life and saves a lot of time. Html (Hypertext Markup Language) and CSS (Cascading Style Sheet) are two of the core technologies for building Web pages.

HTML provides the structure of the page, CSS the (virtual and aural) layout for a variety of devices. Along with graphics and scripting HTML and CSS are the basis of building Web pages and Web Applications. HTML gives authors the means to : Publish online documents with headings, text, tables, lists, photos. Retrieve online information via hypertext links, at the click of a button. Design forms for conducting transactions with remote services, for use in searching for information etc. Include spreadsheets, video clips, Sound clips and other applications directly in their documents. Flask is the most

popular framework of python for web development. It is free, open source and server-side(the code is executed on the server).

Machine learning is a process that is widely used for prediction. N number of algorithms are available in various libraries which can be used for prediction. In this article, we are going to build a prediction model on historic data using different machine learning algorithms and classifiers, plot the results and calculate the accuracy of the model on the testing data. Building/Training a model using various algorithms on a large dataset is one part of the data. But using these models within different applications is the second part of deploying machine learning in the real world.To put it to use in order to predict the new data, we have to deploy it over the internet so that the outside world can use it. In this article, we will talk about how we have trained a machine learning model, created a web application on it using Flask.

CHAPTER 2
LITERATURE SURVEY

Literature survey is the most important step in the software development process. Before developing the tool it is necessary to determine the time factor, economy and company strength. Once these things are satisfied, then the next step is to determine which operating system and language can be used for developing the tool. Once the programmers start building the tool the programmers need a lot of external support. This support can be obtained from senior programmers, from books or from websites. Before building the system the above considerations are taken into account for developing the proposed system. The major part of the project development sector considers and fully surveys all the required needs for developing the project. For every project Literature survey is the most important sector in the software development process. Before developing the tools and the associated designing it is necessary to determine and survey the time factor, resource requirement, man power, economy, and company strength. Once these things are satisfied and fully surveyed, then the next step is to determine about the software specifications in the respective system such as what type of operating system the project would require, and what all the necessary software is needed to proceed with the next step such as developing the tools, and the associated operations.

CHAPTER 3

AIM AND SCOPE OF THE PROJECT

3.1 EXISTING SYSTEM

Data mining in the study and analysis of criminology can be categorized into main areas, crime control and crime suppression. De Bruin et. al. introduced a framework for crime trends using a new distance measure for comparing all individuals based on their profiles and then clustering them accordingly. Manish Gupta et. al. highlights the existing systems used by Indian police as e-governance initiatives and also proposes an interactive query based interface as crime analysis tool to assist police in their activities. He proposed an interface which is used to extract useful information from the vast crime database maintained by National Crime Record Bureau (NCRB) and find crime hot spots using crime data mining techniques such as clustering etc. The effectiveness of the proposed interface has been illustrated on Indian crime records. Sutapat Thiprungsri examines the application of cluster analysis in the accounting domain, particularly discrepancy detection in audit. The purpose of his study is to examine the use of clustering technology to automate fraud filtering during an audit. He used cluster analysis to help auditors focus their efforts when evaluating group life insurance claims.

3.2 PROPOSED SYSTEM

In this project, we will be using the technique of machine learning and data science for crime prediction of crime data sets. The crime data is extracted from the official portal of police. It consists of crime information like location description, type of crime, date, time, latitude, longitude. Before training the model data preprocessing will be done following this feature selection and scaling will be done so that the accuracy obtained will be high. The K-Nearest Neighbor (KNN) classification and various other algorithms (Decision

Tree and Random Forest) will be tested for crime and propose one with better

CHAPTER 4

METHODOLOGY

4.1 INTRODUCTION TO MACHINE LEARNING

A comparative study was carried out between violent crime patterns from the Communities and Crime Unnormalized Dataset versus actual crime statistical data using the open source data mining software Waikato Environment for Knowledge Analysis (WEKA). Three algorithms, namely, linear regression, additive regression, and decision stump, were implemented using the same finite set of features on communities and actual crime datasets. Test samples were randomly selected. The linear regression algorithm could handle randomness to a certain extent in the test samples and thus proved to be the best among all three selected algorithms. The scope of the project was to prove the efficiency and accuracy of ML algorithms in predicting violent crime patterns and other applications, such as determining criminal hotspots, creating criminal profiles, and learning criminal trends.

When considering WEKA , the integration of a new graphical interface called Knowledge Flow is possible, which can be used as a substitute for Internet Explorer. IT provides a more concentrated view of data mining in association with the process orientation, in which individual learning components (represented by java beans) are used graphically to show a certain flow of information. The authors then describe another graphical interface called an experimenter, which as the name suggests, is designed to compare the performance of multiple learning schemes on multiple data sets.

4.2 TRAINING THE DATA

There are basically two widely-used types of training that can be done to create a model: